

LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV Band 5 · Nr. 2/3 · März 1988

Editorial

LDV-FORUM 5(1987/88)2/3

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

Herausgeber

Gesellschaft für Linguistische Datenverarbeitung (GLDV)

Redaktion

Prof. Dr. G. Knorz, FH Darmstadt, Fachbereich LuD, Schöffersstraße 8, D-6100 Darmstadt

Tel.: (06151) 317881

Netz: xid2knor@ddathd21.bitnet

unter Mithilfe von Athena Gavrides und unter Beteiligung einer Projektgruppe "elektronisches Publizieren" am Fachbereich LuD der FH Darmstadt.

Wissenschaftlicher Beirat des LDV-Forum:

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

Wissenschaftliche Zusammenarbeit

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

Erscheinungsweise:

halbjährlich, bis Bd. 5, Heft 1 einschließlich: zum 30. Juni und 30. Dezember – ab Bd. 5, Heft 2/3: zum 31. März und 30. September

Bezugsbedingungen:

Der Bezug des LDV-Forum ist für Mitglieder der GLDV im Jahresbeitrag eingeschlossen. Jahresabonnements können zum Preis von DM 30,00 (incl. Versand), Einzel Exemplare zum Preis von DM 15,00 (zuzügl. Versandkosten) bei der Redaktion bestellt werden. Back-up-Exemplare bis einschließlich Ausgabe 2/85 sind zum Preis von DM 10,00, ab Ausgabe 2/86 zum regulären

EDITORIAL

Das LDV-Forum ändert mit der vorliegenden Ausgabe sein Erscheinungsdatum: nicht mehr im Dezember und Juni, sondern im März und September soll das jeweils neueste Heft ausgeliefert werden. Für dieses neue Zeitraster sprechen mindestens zwei Gründe: zunächst ist es realitätsnäher, denn bisher war bis in den Juni hinein die aktuelle LDV-Forum-Ausgabe die Nummer 2 der Vorjahres gewesen. Zum zweiten ist es redaktionsfreundlicher, denn die "heiße" Phase der redaktionellen Arbeit kann nunmehr – zumindest zum Teil – in die Semesterferien verlagert werden.

Seitens der Leserschaft und der GLDV wäre etwas mehr Transparenz, konkret eine entsprechende Ankündigung des Umstellungszeitpunktes, sicherlich wünschenswert gewesen. Ich muß zugeben, daß ich selbst auf die Vielzahl von Gründen für eine sofortige Umstellung eigentlich nur reagiert habe und mit der Zusammenlegung der Dezember-Ausgabe 1987 und der ersten Ausgabe im neuen Zeitraster, der März-Ausgabe, eine Reihe dringender Probleme lösen konnte: Die Fertigstellung von 2/87 zum angestrebten Termin war aus arbeitstechnischen Gründen in keiner Weise zu schaffen, und der Umfang des Heftes hat die angestrebten 72 – 80 Seiten wieder deutlich überschritten. Die gefundene Regelung nimmt den Zeitdruck aus der weiteren redaktionellen Planung und Arbeit, entschärft das Problem der Herstellungskosten und schafft insgesamt eine bessere Ausgangsbasis für die Zukunft (s. o.).

Zum heutigen Zeitpunkt gehe ich nach Rücksprache mit dem Hoppenstedt-Verlag davon aus, daß das LDV-Forum noch rechtzeitig zur GLDV-Jahrestagung in der zweiten März-Woche fertiggestellt sein kann (realistischer Optimismus).

Eine Entwicklung, die für die Zukunft der Computerlinguistik von ganz besonderer Relevanz ist, sind die gegenwärtigen Bemühungen um die Einrichtung eines DFG-Programms mit dem Arbeitstitel *Verarbeitung sprachlich repräsentierter Information*. Zwar kann dieser Schwerpunkt nicht in Anspruch nehmen, die gesamte Bandbreite der Computerlinguistik abzudecken, es wird aber langfristig auch für die nicht direkt angesprochenen Bereiche der CL von entscheidender Bedeutung sein, daß Computerlinguistik in der DFG nicht weiterhin grundsätzlich und strukturbedingt zwischen die Stühle Informatik und Linguistik fällt.

Bestimmend für die spezielle Ausrichtung des angestrebten überregionalen Schwerpunktes war die Tatsache gewesen, daß das BMFT bei der Formulierung seines neuen Fachinformationsprogrammes die Förderung informationswissenschaftlicher (–linguistischer) Grundlagenforschung unter Hinweis auf die DFG ausgeklammert hatte. Diese Sicht ist nicht zu kritisieren, wenn der Bedarf an Grundlagenforschung für sprachlich fundierte Transferprozesse vom Wissensproduzenten zum Nutzer bei der DFG auch tatsächlich gesehen wird und dort dieser Bereich durch kompetente Gutachter abgedeckt ist.



EDITORIAL

Als Ergebnis einer ersten breiteren Diskussion am 5. Juni 87 wurden im Rahmen des geplanten Schwerpunktes 4 Themenbereiche definiert, für die sich einzelne Arbeitsgruppen zusammenfanden: *Wissensproduktion* (Koordination: Endres-Niggemeyer/Rothkegel), *Repräsentation* (Koordination: Lenders), *Mensch-Maschine-Kommunikation* (Koordination: Krause) und *Informationsprodukte* (Koordination: Kuhlen). Auf dem GLDV-Workshop der GAL-Tagung (3. September 87) wurden die Bemühungen um die Definition und Beantragung eines DFG-Schwerpunktes in einem Paper von Kuhlen begründet und der Stand der Arbeiten vorgestellt. Nach einem zweiten Zusammentreffen aller Interessierten am 15. und 16. Januar sind wir insofern einen wichtigen Schritt weiter, als daß einstimmig die Beantragung des Schwerpunktes im Herbst 1988 beschlossen wurde. Dieses Beschlußergebnis ist durchaus nicht selbstverständlich, denn die Barrieren für einen DFG-Schwerpunkt sind nicht niedrig angesetzt – und leichtfertig ein Vorhaben in Gang zu setzen, hinter dem nicht genug Substanz (auch im Sinne von Arbeitskapazität) steht, wäre kaum zu verantworten.

Ziel der weiteren Arbeit muß es sein, noch vor der Sommerpause eine endgültige Antragsfassung fertigzustellen. Die nach den Themenbereichen gegliederten und im Januar durchdiskutierten Rohfassungen befinden sich noch in recht unterschiedlichem Zustand. In jedem Falle bleibt den Koordinatoren (und hoffentlich nicht nur diesen allein) noch sehr viel zu tun. Mit Sicherheit wird auch auf der GLDV-Jahrestagung in Saarbrücken über den aktuellen Stand diskutiert werden.

Ich gehe davon aus, daß bereits die nächste Ausgabe, Bd. 5, Nr. 4, im Herbst dieses Jahres im fachlichen Teil von den wissenschaftlichen Vorbereitungen der Beantragung profitiert. Ich hoffe sehr, daß es sich – organisatorisch betrachtet – dann wieder um eine Routineausgabe handelt. In Hinblick auf die formale und inhaltliche Gestaltung sollte man Routineausgaben wohl besser nicht anstreben.

g.k.

P.S.: Der elektronische Newsletter *gldv-nl* funktioniert!!! Nachrichtenwert und Attraktivität hängen allerdings, nicht anders als beim LDV-Forum, von seinen LeserInnen ab!

P.P.S.: In Zusammenarbeit mit SCOPE (Osprey, FL: Paradigm) bietet die Redaktion eine neue Version von disc SCOPE an: Referenzen auf ausgewählte Software und Courseware, speziell für den geisteswissenschaftlichen Bereich, mit komfortablem menu-gesteuertem Zugriff auf einer 5¹/₄-Zoll Floppy für IBM (-Kompatible). Formlos anfordern und den Selbstkostenpreis von DM 9,00 für Floppy Disk und Kopieren auf das LDV-Forum-Konto (s. Impressum) überweisen!

Studienführer LDV

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

Arbeitskreise:

Ausbildung und Berufsperspektiven; LDV und Nachbarn; Maschinelle Lexikographie und Lexikologie; Maschinelle Übersetzung; Spracherkennung, Sprachgenerierung und phonetische Datenbanken; Textanalyse

Preis von DM 15,00 (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 März-Heft: 15. Jan., für das September-Heft: 15. Juli

Herstellung:

Verlagsdruckerei Hoppenstedt, Havelstraße 9, D-6100 Darmstadt

Auflage:

550 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

Vorstand:

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

INDIKATOREN DER KOMPETENZ IM SPRACHLICHEN AUSDRUCK - EIN BEITRAG ZUR PARTNERMODELLIERUNG IN NATÜRLICHSPRACHLICHEN SYSTEMEN

J. Krens und M. Schießl

Universität Regensburg, Institut für
Psychologie Universitätsstr. 31a, 8400
Regensburg

Ergebnisse der Kommunikationspsychologie, der Rhetorik und der Soziolinguistik werden für die Diskussion von Benutzermodellen in natürlichsprachlichen Systemen herangezogen. Der Beitrag geht davon aus, daß solche Systeme ihre Effektivität und ihre Kooperationsfähigkeit durch eine zutreffende Einschätzung der Fachkompetenz des Benutzers erhöhen können. Indikatoren für Kompetenz, soweit sie als Merkmale des Sprachverhaltens von Benutzern erkannt werden können, füllen die Slots eines "Kompetenz-Frames".

This study discusses user modeling in NL-systems taking into account findings of communication psychology, rhetorics and sociolinguistics. It is assumed that an adequate estimate of the partner's competence in the area discussed increases the systems's effectivity and ability for optimal cooperation. Indicators for competence as a trait that can be derived from the user's language behavior are gathered and used to fill the slots of a competence frame.

Daß natürlichsprachliche Systeme (NL-Systeme) ob als Auskunfts-, Experten- oder Instruktionssysteme - ihre Kooperationsfähigkeit mit dem Benutzer durch die Berücksichtigung partnerbezogenen Wissens steigern können, gilt mittlerweile als weithin akzeptierte Tatsache. In der automatischen Sprachverarbeitung werden deshalb Annahmen über Ansichten, Ziele und Eigenschaften des jeweiligen Gesprächspartners in expliziten Partnermodellen abgebildet (MORIK & ROLLINGER, 1983; KOBASA, 1985; RICH, 1983; SLEEMAN et al., 1985) und sowohl in der Bedeutungsanalyse wie in der Antwortgenerierung eingesetzt. Die vorliegende Studie soll dem umfassenden Nachweis, daß Partnermodelle unverzichtbare Bestandteile von Dialogsystemen sind, keine weiteren Argumente hinzufügen. Im Mittelpunkt steht als Detailausschnitt vielmehr die Partnereigenschaft Kompetenz, wie sie als Merkmal dem jeweiligen Gegenüber im Dialog zugesprochen wird. Neben den Absichten und Zielen von Dialogpartnern, deren Identifikation aus dem Gesprächsverhalten insbesondere Thema computerlinguistischer Analysen ist, wird die Persönlichkeitsdimension Kompetenz neuerdings von der Expertensystemforschung (vgl. SLATTER, 1987) aufgegriffen. Es soll gezeigt werden, daß zum Zustandekommen von kooperativen oder 'normalen' Dialogen die Berücksichtigung der Sachkenntnisse des Partners in der Strukturierung des Sprachverhaltens unumgänglich ist. Einzelne, von der Sprachpsychologie bislang identifizierte Indikatoren der Kompetenz - Charakteristika des Sprachverhaltens, die Rückschlüsse auf den Kompetenzgrad des Sprechers erlauben - sollen exemplarisch für die Anwendung in einem NL-System diskutiert werden.

1. Kompetenz und Dialogeffektivität

"Kompetenz", mit dem Gegensatzbegriff "Inkompetenz", wird hier in der konventionellen Bedeutung der Alltagssprache verwendet und bezeichnet sachbezogene Fertigkeiten und Kenntnisse. Der Begriff wird sowohl für die Benennung einer Dimension wie für die Bezeichnung eines Extremwertes auf dieser Dimension gebraucht. Kompetenzgrade werden einer Person aufgrund des Verhältnisses ihrer individuellen Fähigkeitsausprägung zu den erreichbaren Extremwerten, von der Sache und/oder einer Normpopulation her definiert, zugesprochen. Daraus folgt auch, daß Kompetenz einer Person nicht schlechthin, sondern nur im Hinblick auf einen einzelnen Sachverhalt unterstellt wird. Die Effektivität eines Dialogsystems bemißt sich daran, in wie kurzer Zeit vom Benutzer wie viele seiner jeweiligen Kommunikationsziele erreicht werden können. Damit sind Systemantworten zu vermeiden, die

- a) sich mit bereits vorhandenem Wissen stark überschneiden, da sie für den Partner keine neuen Gesichtspunkte enthalten. Andererseits muß der Benutzer in der Lage sein, die Antwort des Systems zu verstehen, egal ob vom eingesetzten Wortschatz oder vom benutzten Weltwissen her. Daher sind auch Antworten zu vermeiden, die
- b) beim Partner nicht bereits vorhandenes, aber zur adäquaten Äußerungsinterpretation notwendiges Wissen voraussetzen. Die unter a) und b) angeführten Kriterien der Antwortgestaltung eines Systems erfordern ein Bild der Sachkenntnisse und Fertigkeiten des Partners. Wird das Kompetenzniveau nicht beachtet, ist ein weniger reibungsloses Funktio

nieren der Kommunikation eher wahrscheinlich, da im Regelfall

1. der Benutzer gezwungen ist, zusätzliche Nachfragen zu unternehmen (bei "Unterbeantwortung", s. Fall b, wenn nicht vorhandenes Wissen vorausgesetzt wird;
2. der Benutzer gelangweilt ist durch Antworten, die ihn mit bereits bekanntem Wissen versorgen. Dies gilt für die einfache Beantwortung direkter Fragen, wie auch für das sog. 'overanswering' (WAHLSTER et al., 1983, S. 643 ff) (s. Fall a»;
3. der Benutzer einem ihm unverständlichen Wortschatz oder Sprachgebrauch gegenübersteht (" mißverständliche Signalebene" , (s. Fall b»;
4. der Benutzer verblüfft ist durch die Korrektur von Fehlannahmen, wenn er die Sache kennt und sich nur ungenau ausdrückte bzw., da er fehlgeleitet wird, wenn er nicht kompetent ist und eine Richtigstellung unterbleibt. (vgl. "misconceptions" , WEBBER & MAZS, 1983, S.650ff, (s. Fall b»;
5. der Benutzer verwirrt ist durch den Differenzierungsgrad der Systemantwort, die er aufgrund seiner mangelnden Fachkenntnis nicht integrieren kann (s. Fall b);
6. die eigentlichen Absichten und Ziele des Partners verkannt werden, wenn seine Anfrage nur innerhalb subjektiver Bezugssysteme (z.B. Unschärfe von natürlichsprachlichen Mengenangaben) angemessen zu verstehen ist.

Damit sind nur einfache Konfliktstellen genannt, deren Auftreten durch die Mißachtung des Kompetenzniveaus eher wahrscheinlich wird. Letztlich sind von einem System, das die Kompetenz des Partners in seinen Antworten berücksichtigt, zwei Effekte beim Benutzer eher zu erzielen:

- emotionale Sicherheit (d.h. Freude am Dialog, keine Langeweile, kein Ärger, Gefühl des Verstandenwerdens) und
- rationale Sicherheit (d.h. geistiges, inhaltliches Erfassen der gegebenen Informationen und Integration mit dem bereits vorhandenen Wissensfundus).

2. Zur Repräsentation von Partnerbildern

2.1 Frames als Repräsentationsprinzip

In Anlehnung an sozialpsychologische Theorien zur interpersonellen Wahrnehmung (vgl. CANTOR & MISCHEL, 1979) kann subjektives Wissen über Kompetenzunterschiede von Partnern durch Prototypen, die für unterschiedliche Kompetenzgrade (der 'typische' Experte, Fachmann, Laie usw.) stehen, modelliert werden. Sie sind eine Variante sog. impliziter Persönlichkeitstheorien. Der Begriff bezeichnet "... die Auffassungen einer Person (als Beurteiler) über das Vorkommen, die Häufigkeits- und

Ausprägungsverteilung von (Persönlichkeits-) Merkmalen, über deren Bedingungen und deren Auswirkungen, über deren Zusammenhänge (Korrelationen) untereinander und mit weiteren Faktoren, über die Möglichkeiten zu deren Beurteilung" (PREISER, 1979, 5.83). Es bietet sich an, auch Dialogsysteme mit Datenstrukturen auszustatten, in denen Wissen über 'typische' Partnereigenschaften abgebildet ist. Als Repräsentationsformat eignen sich Frames nicht nur wegen ihrer Nähe zu den sozialpsychologischen Konzepten, sondern auch wegen ihrer Fähigkeit zur Darstellung erwartungsgeleiteter Informationsverarbeitung (sie liegt dem Prozeß der Verifikation von Partnerbildern zugrunde) und wegen der Möglichkeit, prozedurales Wissen leicht integrieren zu können. Das System ist dann mit einzelnen 'Kompetenzframes' auszustatten, in denen vorgegeben ist, wie ein Fachmann, ein Experte, ein Laie egal auf welchem Gebiet - zu indentifizieren ist.

2.2. Sprachliche Indikatoren der Kompetenz

Festzuhalten ist, daß für die Identifikation der Partnerkompetenz kein 'Modell' eines Experten oder Laien im Sinne einer isomorphen Repräsentation der kennzeichnenden Eigenschaften notwendig ist. Es ist keine Theorie 'kompetenter' versus 'inkompetenter' Informationsverarbeitung erforderlich, die als Komponente eines Dialogsystems darzustellen wäre. Entscheidend ist im vorliegenden Zusammenhang ausschließlich, wie zu erkennen ist, ob der Partner Fachmann oder Laie ist. In der Definition des Fachmannframes sind also vorrangig die gesprächsdiagnostisch relevanten Gesichtspunkte - Kovariationen von Merkmalen des Sprachverhaltens mit individuellen Kompetenzgraden - aufzunehmen. Die Bestimmung der Bestandteile eines Fachmannframes reduziert sich dann auf die Suche nach beobachtbaren Textmerkmalen, die als leidlich verlässliche Indikatoren mit dem Kompetenzgrad korrelieren und eine Aussage zur unterstellten Persönlichkeitseigenschaft zulassen.

Zur Einschätzung des Kompetenzgrades stehen im Dialog wenigstens drei Informationsquellen zur Verfügung:

a) Personencharakteristika: Hierzu zählen Partnereigenschaften, die eine Ableitung des Kenntnisstandes aufgrund seiner Biographie, Zugehörigkeit zu bestimmten sozialen Gruppen usw. erlauben. Beispielsweise kann aus dem Ausbildungsgang, Studienabschluß oder der ausgeübten Tätigkeit der ungefähre Kompetenzgrad auf einzelnen Leistungsgebieten abgeschätzt werden. Darauf wird hier nur beiläufig verwiesen, da einerseits relevante Informationen auf den jeweiligen Anwendungsbereich zugeschnitten werden müssen und andererseits die benötigten Daten leicht innerhalb des Dialogs abgefragt werden können. Anzumerken ist aber, daß durch explizites Fragen oder Nachfragen der reibungslose Verlauf des Dialogs eventuell aufgehalten wird.

b) **Dialogthemen:** Auf das dem Partner verfügbare Wissen kann auch aus den von ihm angeführten Themen, Gesichtspunkten etc. geschlossen werden. Es ist dabei zu klären, welche Informationen ihm 'vermutlich' aufgrund des von ihm erfragten, angesprochenen Sachverhalts präsent sind. Auch hierfür wären Inferenzregeln anzugeben, die aufgrund der vom Partner angeführten Gesichtspunkte sein Sachwissen, evtl. auch seine Fehlannahmen, hypothetisch ableiten. Diese Datenquelle soll in der vorliegenden Studie ebenfalls nur am Rande erwähnt sein. Im Mittelpunkt stehen hingegen

c) **Formulierungscharakteristika.** Sie umfassen das 'wie' der Partneräußerung - die Formeigenschaften - und sollen aus der Art und Weise, in der formuliert wird, Rückschlüsse auf Personeneigenschaften erlauben. Gesucht sind Ausdrucksindikatoren der Kompetenz, die den Kompetenzgrad des Benutzers aufgrund typischer Merkmale der Textgestalt erschließen lassen. Es sind extralinguistische Variationen (vgl. MAHL & SCHULZE, 1964) der geschriebenen Sprache zu identifizieren, die als aussagekräftige Hinweissignale zum Aufbau eines Partnerbildes aufgegriffen werden können. In der automatischen Sprachanalyse werden stilistische Informationsquellen bislang kaum berücksichtigt. KOBASA (1985, 116ff.) untersucht den 'Indizienwert' der syntaktischen Form von Dialogbeiträgen und der Verwendung bestimmter Sprachpartikel für die Identifikation von Benutzerüberzeugungen. Aber auch von der Sprachpsychologie wurde nur in wenigen Arbeiten (z.B. ISAACS & CLARC, 1987) der Aussagewert von Formulierungsvarianten im Hinblick auf Persönlichkeitseigenschaften experimentell untersucht. Empirische Analysen stammen überwiegend aus der Soziolinguistik, Stilforschung und kommunikationspsychologie, aus Gebieten, die in der Konstruktion von NI-Systemen bislang kaum oder nur am Rande Beachtung fanden.

a) **Ergebnisse soziolinguistischer Untersuchungen:**

Im Unterschied zur Analyse struktureller Merkmale der gesprochenen Sprache liegen für die Stilforschung vergleichsweise wenige Arbeiten vor, in denen die Attribution von Kompetenz aufgrund schriftlich fixierter Texte untersucht wird. In mehreren Untersuchungen (vgl. BRADAC et al. 1976) konnte gezeigt werden, daß 'linguistic diversity' operationalisiert durch den Umfang des eingesetzten Vokabulars, die Kompliziertheit grammatikalischer Konstruktionen, die verwendeten Satzbaupläne u.ä., zu einer eher positiven Einschätzung des Sprechers führt. Er wird u.a. in seinen intellektuellen Fähigkeiten höher eingestuft. BRADAC et al. (1980) wiesen nach, daß auch "beliebdiscrepant messages", also unerwartete Nachrichten, um so eher als "kompetent und glaubwürdig" eingestuft werden, je mehr sie sich durch Wortreichtum und lexikalische Redundanz

auszeichnen. Ein differenzierter Wortschatz, eine facettenreiche Begriffswahl, variantenreiche syntaktische Konstruktionen lassen den Hörer eher an einen kompetenten Sprecher glauben.

Der Zusammenhang von Persönlichkeitseigenschaften und sprachlichem Ausdruck wurde für die vokale Kommunikation intensiv beleuchtet (vgl. SCHERER, 1982). Für die 'personality markers in speech' (SCHERER & GILES, 1979) müßte im Detail untersucht werden, wie weit sie auch unter den typischen Variablen, die den Mensch-Maschine-Dialog als ggf. eigenständige Kommunikationssituation abgrenzen, wirksam sind. Beispielsweise wäre zu klären, ob relativ unspezifische Charakteristika wie Äußerungslänge (vgl. BROWN, STRONG & RENCHER, 1975), orthographische Fehler usw. auch in der schriftlichen Kommunikation mit dem vermuteten Kompetenzgrad kovariieren. Extralinguistische Merkmale der Benutzerbeiträge in der Kommunikation mit einem NI-System, wie Länge, lexikalische Differenziertheit u.a. könnten dann wenigstens als erste Hypothese über die dem Partner zu unterstellenden Sachkenntnisse herangezogen werden.

b) **Ergebnisse kommunikationspsychologischer Forschung:** Eine weitere Quelle zur Identifikation kompetenzindizierender Sprachmerkmale sind Arbeiten der Kommunikationspsychologie, wie sie insbesondere von VUKOVICH (1977) vorgestellt wurden. VUKOVICH systematisiert Argumentationsfiguren, die als 'building blocks' des verbalen Informationsaustausches aufzufassen sind und neben binnenstrukturellen Eigenheiten in erster Linie aufgrund der gezeigten Wirkung beim Hörer/leser abgegrenzt werden. "Rhetorische Figuren sind Informationsschemata mit spezifischen Wirkungsaspekten" (VUKOVICH, 1977, S.160). Für die vorliegende Fragestellung sind Figuren von Interesse, deren Effekt sich insbesondere in der Vermittlung eines kompetenten Eindrucks durch den Sprecher beim Hörer niederschlägt. Dafür kommen Figuren sie können im einzelnen hier nicht dargestellt werden (vgl. VUKOVICH, 1987) - in Frage, die den Sprecher kompetent erscheinen lassen, weil

- die Figur andeutet, daß der Sprecher um den wahren Sachverhalt weiß, z.B. zunächst bloß Schwierigkeiten hat, ihn verbal auszudrücken, es ihm aber letztlich doch gelingt (z.B. Wortsuche, Sachbezogene Correctio);
- die Figur zeigt, daß der Sprecher einen Kompetenzgrad beansprucht, der den des Adressaten übersteigt, z.B. dem Adressaten abspricht, im Gegensatz zu ihm die wesentlichen Aspekte des Gesagten zu erkennen und richtig zu verstehen (z.B. Sie sind nicht auf dem laufenden, Der Kenner der Materie);
- die Figur die Kompetenz des Sprechers durch 'gestohlene Kompetenz' erweitert, z.B. er sich auf Autoritäten zu berufen versteht, denen der Adressat nichts entgegenhalten kann und

die der Adressat auch wohl oder übel anerkennen muß (z.B. Die allgemein anerkannte Autorität);

- die Figur zeigt, daß der Sprecher in der Lage ist, den angesprochenen Sachverhalt von unterschiedlichen Seiten zu beleuchten und zu beschreiben (z.B. Bedeutungshäufung, Bedeutungsdifferenzierung).

c) Weitere mögliche sprachliche Indikatoren der Kompetenz:

ISAACS & CLARK (1987) untersuchen Dialogstrategien von Laien und Experten in der Alltagskonversation. Als Indikatoren von Kompetenz identifizieren sie u.a. die Verwendung von Eigennamen im Gegensatz zu Umschreibungen und den vergleichsweise häufigeren Gebrauch von definiten Verweisen. Weitere Hinweise auf Kompetenzindikatoren sind auch Lehrbüchern der Stilkunde zu entnehmen. Nach BUSEMANN (1948, S.96f) beispielsweise ist der sachliche Stil, der auf einen kompetenten Sprecher schließen läßt, durch die Verwendung von Fachausdrücken, Angabe von Anzahl und Art von Gegenständen, insgesamt durch einen 'nominalen' Charakter gekennzeichnet. Auch die detaillierte Analyse 'persuasiver' Texte sollte einen Fundus stilistischer Merkmale zur Beanspruchung und Demonstration von Sachkompetenz eröffnen. In komplexen, schwer durchschaubarer Themengebieten, in denen vom Rezipienten die Sachgemäßheit einer Aussage nur in Annäherung zu beurteilen ist, sind Plausibilität, Eindruck und Überzeugungskraft der Formulierungen, die der Hörer unter Maßgabe seines eigenen mehr oder weniger großen Sachverständnisses empfindet, entscheidend. So ist es eine erfolgsversprechende Strategie, in der Identifikation von Indikatoren Texte zu analysieren, in denen aufgrund der Komplexität der Sachverhalte insbesondere mittels Formulierungskünsten ein Kompetenzeindruck erweckt werden soll. Bereits eine oberflächliche Analyse von Beispieltexten zeigt, daß die Verfasser vorzugsweise durch die Anhäufung von FACHAUSDRÜCKEN, ORTSANGABEN, EIGENNAMEN, JAHRESZAHLEN, DATUMS-, PROZENT-, GRÖßENANGABEN aller Art Kompetenz bezüglich des in Frage stehenden Themas beanspruchen (vgl. SCHIEBL, 1983).

3. Kompetenzidentifikation in einem NL-System

Vorschlag dieser Studie ist es, NL-Systeme mit einer expliziten Wissenskomponente auszustatten, die in Partnermodellen Informationen über unterschiedlich kompetente Dialogteilnehmer enthält und in der Diagnose des Wissensstandes des jeweils aktuellen Partners eingesetzt werden kann. Es soll nun abschließend in Grundzügen eine Konkretisierung der bislang vorgestellten Überlegungen zur Diagnostik von Partnereigenschaften in Dialogen vorgestellt werden, die gegenwärtig im Rahmen eines NL

Forschungssystems entwickelt wird. Da allerdings im vorliegenden Forschungsrahmen Implementierungen in erster Linie als Formalisierungsverfahren und als Medium zur Durchführung von Experimenten dienen, deren Ergebnisse ihrerseits Anknüpfungspunkte für empirische, sprachpsychologische Untersuchungen darstellen, bleiben Performanz- und Anwendungsgesichtspunkte weitgehend außer Betracht. Deshalb wird an manchen Stellen ein eher exemplarischer, gelegentlich auch skizzenhafter Charakter in Kauf genommen. Beansprucht wird allerdings, die Verwertbarkeit sprachpsychologischer Erkenntnisse soweit zu präzisieren und dem Entwicklungsstand computerlinguistischer Techniken hinreichend anzunähern, daß an einem Prototyp wenigstens Konkretisierungslinien demonstriert werden können. Anzuführen ist auch, daß es sich hier nur um einen von sicherlich mehreren Realisierungsvorschlägen handelt, der zudem auf einer eher traditionellen Systemarchitektur beruht.

3.1 Kurze Darstellung des Dialogsystems

Das gegenwärtig entwickelte System besteht aus drei Analysekomponenten: morphologische, syntaktische und 'kommunikationsdiagnostische' Analyse. Für die morphologische und syntaktische Zerlegung werden Komponenten von PLIDIS (KOLVENBACH, LOTSCHER, LUTZ, 1979) adaptiert. Es handelt sich im wesentlichen um die Bausteine PASSO und PASSI, die hier nicht näher erläutert werden.

Die kommunikationsdiagnostische Analysekomponente umfaßt zwei Bestandteile:

- (a) Dialogpartnermodelle
- (b) Figurenparser

(a) Dialogpartnermodelle

Dialogpartnermodelle sind in unserem System statische Datenstrukturen, die zur permanenten Wissensbasis des NL-Systems gehören. In ihnen ist das stereotypisierte Partnerwissen im bislang beschriebenen Sinne repräsentiert. Vorläufig besitzt diese Komponente nur einen 'Kompetenzframe', der die diagnostisch relevanten Merkmale des Sprachverhaltens zur Identifikation des Partnerwissens enthält. In der aktuellen Ausbaustufe besteht dieser Frame aus 8 Slots, die zwei Klassen zuzuordnen sind:

- (1) Kontextabhängige Merkmalsdimensionen. Dazu gehören Merkmalsgrößen, deren Werte nur innerhalb einzelner Anwendungsgebiete bedeutsam sind. Beispielsweise sind die Werte eines möglichen BERUF-Slots, die eine hohe Kompetenz erscheinen lassen, für jedes Sachgebiet spezifisch festzulegen. Dies gilt nicht für die hier diskutierten.
- (2) extralinguistischen Sprachmerkmale, für die überwiegend unterstellt werden kann, daß ihr Informationswert sachgebietsübergreifend gegeben ist. Sie sind als inhaltsinvariante Formulierungsmerkmale unabhängig vom je

(FACHMANN	(A-KIND-OF (BERUF	(VALUE PERSON-FRAME)) (VALUE beruf! beruf2 . . .) (IF-NEEDED ASK))
	(TÄTIGKEIT	(VALUE tätigkeit1 tätigkeit2 " .) (IF-NEEDED ASK)) ..
	(ARG-FIGUR	(VALUE BEDEUTUNGSHAUFUNG CORRECTIO ...) (IF-NEEDED IDENTIFY))
	(FACHAUSDRUCK	(VALUE ausdruck1 ausdruck2 ...) (IF-NEEDED IDENTIFY))

Abb. 1: Ausschnitt aus einem Fachmann-Frame

weiligen Umfeld und umfassen die in 2.2 genannten kompetenzindizierenden Sprachgemeinschaften.

In Abb. 1 ist ein Ausschnitt aus diesem Fachmannframe wiedergegeben. Es sind u.a. die Slots - BERUF, TÄTIGKEIT - angeführt, die zu den kontextabhängigen Merkmalsdimensionen zu rechnen sind. Ihnen sind jeweils zwei 'facets' beigestellt: eine Werteliste, die als Elemente jene Berufsbezeichnungen enthalten soll, die in einem ausgegrenzten Anwendungsgebiet eine hohe Kompetenz des Dialogpartners vermuten lassen. Unter dem Indikator IF-NEEDED (vgl. WINSTON, 1984, S.253fF) ist eine Prozedur ASK vorgesehen, die hier ein hypothetisches Verfahren bezeichnet, das die Nachfrage beim Gesprächspartner organisieren soll. Aus der Abbildung wird auch deutlich, daß in der syntaktischen Definition von 'Frames' eine gängige Grammatik (vgl. WINSTON, 1984, S.253fF.; WINSTON & HORN, 1984, 311fF.) verwendet wird. Auch die stilistischen Indikatoren der Kompetenz sind nach dem gleichen Prinzip als Slots und Slotwerte definiert. ARGFIGUR beispielsweise enthält die Namen der in diesem Zusammenhang informationsträchtigen Kommunikationsfiguren. Der Kompetenzframe enthält damit 'Suchmuster', deren Vorliegen in Texten für einen eher hohen Kompetenzgrad des jeweiligen Autors spricht.

(b) Figurenparser

Die Identifikation von Eigennamen, fachspezifischen Abkürzungen, Maßangaben usw. ist vergleichsweise einfach: Es ist zu überprüfen, ob Elemente der den Slots vorweg zugeordneten Wertelisten in den Partnerbeiträgen aufzufinden sind. Ein wesentlich schwierigeres Problem hingegen ist die Angabe von Prozeduren, die ermitteln, ob eine der relevanten Gesprächsfiguren in einer Äußerung vorliegt. In einem ersten Ansatz entwickeln VUKOVICH, KREMS & BOCK (1985) eine ATN- Grammatik für einfache Kommunikationsfiguren und setzen sie zur Identifikation von stilistischen Mustern in Texten ein. Diese Grammatik wird im Rahmen der vorliegenden Überlegungen erweitert. Der bislang ent-

wickelte Übergangsgraph umfaßt ca. 60 Knoten, etwas mehr als 100 Kanten und enthält strukturelles Wissen zur Kovariationsargumenten, Bedeutungshäufung und Bedeutungs differenzierung (vgl. 2.2.b (d)). Die bisher abgebildeten Figuren beruhen allerdings weitgehend auf lexikalisch (WRD-Kanten) oder grammatikalisch bestimmbar en Eigenschaften. Um auch eine einfachere semantische Analyse durchführen zu können, wurde eine Teilmengen des Lexikons aufgrund semantischer Gesichtspunkte in Klassen (z.B. 'Steigerungsverb') eingeteilt und entsprechende Kantenbedingungen (CAT STVERB) definiert. Vorläufig ist folgender Verarbeitungsverlauf konzipiert: Wie bereits erwähnt, wird in der morphologischen und syntaktischen Analyse, auch in der Identifikation von Eigennamen, Größenangaben usw. auf Bausteine von PLIDIS zurückgegriffen. PASSO und PASSI werden ohne Modifikation eingesetzt. Die Bestimmung der kompetenzindizierenden Sprachmerkmale geschieht im Anschluß an die morpho-syntaktische Zerlegung. Ihr Ergebnis ist Eigabematerial für die 'kommunikationsdiagnostische Analyse', die in einem eigenen 'Parsing- Schritt' durchgeführt wird. Die dabei anfallenden Ergebnisse werden in einem Frame vermerkt, der die aktuellen, partnerbezogenen Dialogdaten repräsentiert. Er enthält als Instanz des TypeFrames zu Beginn der Bearbeitung nur dessen Slots mit Default-Werten, die sukzessive durch die Analyseergebnisse ersetzt bzw. ergänzt werden. Als Werte werden die im Gesprächsverlauf ermittelten, kompetenzspezifischen Daten eingetragen, etwa der Beruf des Partners, die Fachbegriffe, die er verwendet, die Argumentationsfiguren usw. Dies geschieht mittels spezieller BUILD-Aktionen im Anschluß an die Analyse der Gesprächsbeiträge. So jedenfalls lautet die Zielvorgabe! Es ist einschränkend zu betonen, daß der skizzierte Ablauf bei sowohl reduziertem Lexikon wie bei einer nur rudimentär implementierten Figurengrammatik vorläufig ausschließlich an Beispielsätzen zu demonstrieren ist.

3.2 Implementation

Erste Bestandteile des kurz charakterisierten Forschungssystems wurden zunächst in INTERLISP

D auf SIEMENS-7536 implementiert, wobei die erwähnten Bausteine aus PLIDIS vom Institut für Deutsche Sprache, Mannheim, dankenswerterweise übernommen werden konnten. Gegenwärtig wird an einer Portierung auf SIEMENS-5800 (XEROX1132), vorläufig erneut unter INTERLISP-D, gearbeitet.

Es ist überflüssig, hervorzuheben, daß es in der Implementation unproblematisch ist, die vorgeschlagenen Slots und Slotwerte in statische Datenstrukturen abzubilden. Die hauptsächliche, bereits erwähnte Schwierigkeit ist die Entwicklung von Verfahren, die die Identifikation so komplexer sprachlicher Größen leisten, wie sie Argumentationsfiguren darstellen. Erste Versuche an sehr einfachen Formulierungen zeigen, daß die exemplarische Implementation unter Verwendung von reduzierten Lexika und Eingabeformaten schnell Grenzen der Darstellungsbreite erreicht, da rhetorische Figuren ihre Wirkung insgesamt nur unter natürlichsprachlichen Bedingungen entfalten. Es ist damit vorläufig unklar, ob das Erkennen von Kommunikationsfiguren in Texten durch die sukzessive Erweiterung einer Figurengrammatik und ihre Anwendung auf morpho-syntaktische Analyseergebnisse in einem hinzugefügten Parsingschritt letztlich zu befriedigenden Ergebnissen führen wird. Folglich ist auch für die Ermittlung der Partnerkompetenz in weiteren Tests erst zu überprüfen, ob sich die Abbildung von kompetenzindizierenden Formulierungsmustern in ATN-Grammatiken und ihre Verwendung innerhalb von sequentiellen Analysestrategien für das Auffinden von Daten, die als Werte eines Kompetenzframes in Betracht kommen, eignet. .

Davon unberührt ist der eigentliche Vorschlag, in der Identifikation von Personeneigenschaften nicht nur die Ziele und Pläne (vgl. CARBONELL, 1980) des Gesprächspartners, sondern auch die extralinguistischen Merkmale seines Sprachverhaltens als ohnehin verfügbare Datenquelle zu nutzen. Zur Einschätzung des Kompetenzgrades jedenfalls können Muster des sprachlichen Verhaltens angegeben werden, die anzeigen und annehmen lassen, daß der Partner - wahrscheinlich - kompetent ist. Befunde der sprach- und insbesondere kommunikationspsychologischen Forschung können in dieser Weise in der Partnermodellierung und damit der Optimierung der Benutzerfreundlichkeit von NL-Systemen aufgegriffen werden. In ergänzenden sprach psychologischen Untersuchungen ist zu klären, welche Stilmerkmale über die bislang abgegrenzten sprachlichen Indikatoren hinaus von Lesern/Hörern als Anzeichen für die Sachkunde eines Autors aufgegriffen werden. Auch deren Interaktion mit sonstigen Informationen zum Gesprächspartner (Beruf, Alter, Ausbildung etc.), mit Charakteristika seiner Dialoggestaltung, Merkmalen seiner Frageziele usw. ist zu analysieren.

Literatur

BRADAC, J. J., COURTRIGHT, J. A., SCHMIDT, G. " DAVIES, R. A. (1976):

The effect of perceived status and linguistic diversity upon judgements of speakers attributes and message effectiveness. *J. of Psychology*, 93, 213-220.

BRADAC, J.J., COURTRIGHT, J.A. " BOWERS, J.W. (1980):

Effects of Intensity, Immediacy and Diversity Upon Receiver Attitudes Towards a Belief-Discrepant Message and Its Source. In: H. GILES, W.P. ROBINSON & P.M. SMITH (eds.), *Language*. Oxford: Pergamon Press, 217-221.

BROWN, B.L., STRONG, W.J., RENCHER, A.C. (1975):

Acoustic determinants of perceptions of personality from speech. *Int. J. of the Society of language*, 6, 11-32.

BUSEMANN, A. (1948):

Stil und Charakter. Untersuchungen zur Psychologie der individuellen Redeform. Meisenheim/Glan: Westkultur-Verlag.

CANTOR, N., MISCHEL, W. (1979):

Prototypicality and personality: Effects on free recall and personality impressions. *J. of Research in Personality*, 13, 187-205.

CARBONELL, J.G. (1980):

Towards a Process Model of Human Personality Traits. *Artificial Intelligence*, 15,49-74.

ISAACS, E.A., CLARK, H.H. (1987):

References in Conversation between Experts and Novices. *J. of Exp. Psychology: General*, 116,26-37.

KOBSA, A. (1985):

Benutzermodellierung in Dialogsystemen, Berlin: Springer.

KOLVENBACH, M., LÖTSCHER, A. " LUTZ, H.D. (1982):

Künstliche Intelligenz und natürliche Sprache. Tübingen: Narr-Verlag.

MAHL, G.F. " SCHULZE, G. (1964):

Psychological research in the extralinguistic area. In: T.A. SEBOECK, A.S. HAYES & M.C. BATESON (eds.), *Approaches to semiotics*. Den Haag: Mouton.

MORIK, K. " ROLLINGER, C.-R. (1983):

Partnermodellierung im Evidenzraum. In: B. Neumann (Hrsg.), *GWAI-83*. Berlin: Springer, 158-168.

PREISER, S. (1979):

Personenwahrnehmung und Beurteilung. Darmstadt: Wissenschaftliche Buchgesellschaft.

RICH, E. (1983):

Users are individuals: individualizing user models. *Journal of Man-Machine Studies*, 18, 199-214.

SCHERER, K.R. (ED.) (1982):

Vokale Kommunikation. Weinheim: Beltz.

SCHERER, K.R. & GILES, H. (eds.) (1979):
Sodal Markers in Speech. Cambridge: University Press.

SCHIEBL, M. (1983):

Aufbau und Veränderung von Partnerbildern Aspekte eines Simulationsmodells. Universität Regensburg. Diplomarbeit.

SLATTER, P.E. (1987):

Building Expert Systems: Cognitive Emulation. Chichester: Horwood.

SLEEMAN, D., APPELLT, D., KONOLIGE, K., RICH, E., SRIDHARAN, N. & SWARTOUT, B. (1985):

User modelling panel. Proc. of IJCAI-9. 1298-1302.

VUKOVICH, A. (1977):

Der rhetorische Forschungsansatz in der Kommunikations-Psychologie. In: W. H. TACK (Hrsg.), Bericht über den 30. Kongreß für Psychologie in Regensburg 1976, Göttingen: Hogrefe, 157-166.

VUKOVICH, A. (1987):

Kommunikationsfiguren. in Vorbereitung. VUKOVICH, A., KREMS, J. & BOCK, H. (1985): Formalisierungsvarianten umgangssprachlicher Kovariantargumente und ihrer kognitiven Voraussetzungen. In: D. Albrecht (Hrsg.), Bericht über den 34. Kongreß für Psychologie in Wien 1984. Göttingen: Hogrefe. 453-455.

WAHLSTER, W., MARBURGER, H., JAMESON, A. & BUSEMANN, S. (1983):

Over-Answering Yes-No Questions: Extended Responses in a NL-Interface to a Vision System. Proc. of IJCAI-9, 643-646.

WEBBER, B.L. & MAYS, E. (1983):

Varieties of User Misconceptions: Detection and Correction. Proc. of IJCAI-8, 650-652. WINSTON, P.H. (1984):

Artificial Intelligence. Reading: Addison-Wesley.

COMPUTERLINGUISTISCHE PROBLEME BEI DER BEARBEITUNG GROSSER TEXTMENGEN

Christoph Schwarz

ZT ZTI IINF 323, Siemens AG München

Anhand von Beispielen wird aufgezeigt, in welchem hohem Maß die computerlinguistische Bearbeitung von großen Mengen von Textdaten Techniken und Erkenntnisse der Empirischen Sprachwissenschaft erfordern. Da die sich an der Künstlichen Intelligenz orientierende Computerlinguistik derzeit die Massentextdatenverarbeitung nicht eigentlich als ihren Gegenstand betrachtet, können möglicherweise Grenzen dieses Ansatzes hier deutlich werden.

1. Einleitung

Die folgenden Ausführungen stehen im Zusammenhang mit Forschungs- und Entwicklungsarbeiten für Textdatenbanken unter Industriebedingungen. Freitext-Recherchehilfen, Auto-Indexing, Auto-Abstracting für große Textmengen, Hilfen für die Erstellung von anwendungsrelevanten Thesauri, Volltext-Klassifikation und Inhaltserschließung - all das sind Aufgabenstellungen, denen sich die Computerlinguistik (CL) nur sehr zögernd stellt.

In den USA gibt es inzwischen eigene Firmen, die sich ausschließlich diesen Fragen widmen. Von Forschungs- und Entwicklungsarbeiten vor dem Hintergrund der genannten Aufgabenstellungen soll im folgenden aus der Sicht der Linguistik berichtet werden. Schwerpunkt der folgenden Ausführungen sind methodische Ansätze, die sehr stark empirisch-linguistisch ausgerichtet sind.

In den USA gibt es inzwischen eigene Firmen, die sich ausschließlich diesen Fragen widmen. Von Forschungs- und Entwicklungsarbeiten vor dem Hintergrund der genannten Aufgabenstellungen soll im folgenden aus der Sicht der Linguistik berichtet werden. Schwerpunkt der folgenden Ausführungen sind methodische Ansätze, die sehr stark empirisch-linguistisch ausgerichtet sind.

Daß inzwischen Silbentrennungsprogramme als "Expertensysteme" und damit als "KI" verkauft werden, spricht weniger gegen den seriös arbeitenden Forscher auf dem Gebiet des anwendungsorientierten "natural language processing" als vielmehr gegen einen Trend, derzeit jede Entwicklung in der CL mit diesen Etiketten zu versehen - perspektivisch möglicherweise zum Schaden derjenigen Entwickler, die unter Industriebedingungen produktnah und anwendungsorientiert konkrete Benutzeranforderungen befriedigen müssen.

Computerlinguisten arbeiten seit über 30 Jahren auf dem Gebiet der Massentextdatenverarbeitung. Saltons (SAL66) nicht sehr glücklicher Versuch, mit dem Kuno-Oettinger-Parser aus dokumentationstechnischer Sicht ziemlich tief in syntaktische Strukturen zu gehen und insbesondere sein Ansatz, Normalisierungen vorzunehmen, die sich allenfalls an den Möglichkeiten der damaligen Transformationsgrammatik orientierten und kaum am Stand der Forschung der Dokumentationswissenschaft, folgten bald andere Ansätze. LEADER (HIK69) schien erfolgversprechend mit einer stark oberflächensyntaktisch-orientierten Analyse.

Als neuere Systeme seien kurz erwähnt: FASIT (01083), CITE (00582) oder SIRE (KOL84), das zunächst als experimentelles System entwickelt wurde und inzwischen kommerziell angeboten wird.

CONTEXT (CON87) als System mit "natural language" Komponenten hat inzwischen zahlreiche Installationen in Israel. Auf der letzten ACM-Tagung für Information Retrieval (ACM86) wurden allein fünf Beiträge zum Thema "natural language processing" angeboten. Einen kritischen Bericht unter Berücksichtigung gegenwärtiger Tendenzen in der KI gibt 00586. Eine Literaturliste zu Verfahren findet sich in SCT86.

Die Tatsache, daß auch nach 30 Jahren Forschung und Entwicklung auf den genannten Gebieten dieser industriell hochrelevante Bereich nur an der Peripherie der CL angesiedelt ist, hat Gründe:

- Verfahren zur Bearbeitung von Massendaten sollen insbesondere für den industriellen Einsatz möglichst anwendungsunabhängig arbeiten. Die Mehrzahl der CL-Verfahren arbeitet anwendungsabhängig.
- Sie müssen für den uneingeschränkten Gebrauch von geschriebener natürlicher Sprache einsetzbar sein. Entsprechenden Grammatiken dürften sehr aufwendig werden und damit kaum für Massentextdaten einsetzbar sein.
- Linguistische Regelmäßigkeiten in inhaltlich unrestringierten, großen Textmengen zu entdecken und zu formalisieren, erfordert enormen empirischen Aufwand: Es gibt in der CL keine "Theorie der Massentextdatenverarbeitung". Im Unterschied zu den Gewohnheiten in der empirischen Sprachforschung muß man gleichsam "alle" Phänomene in geschriebenem Text irgendwie in den Griff bekommen und einheitlich behandeln.
- Algorithmen auf dem Gebiet der Massentextdatenverarbeitung sind vergleichsweise einfach und brauchen möglicherweise nicht den Komfort von KI-artigen Sprachen.
- Die Methoden der linguistischen Analyse großer Textmengen können mit einer geringen Fehlerrate arbeiten. Viele linguistische Phänomene brauchen z.B. im Information-Retrieval-Kontext (Information Retrieval: im folgenden IR) nicht berücksichtigt werden.
 - Oberflächenorientiertes Parsing (die Terminologie ist unscharf: Man spricht gelegent-

lich von "Partiellem Parsing") oder auch eine morphologische Analyse großer Textmengen erfordern u.a. so "altmodische" Dinge wie umfassende Kenntnisse von Grammatikregeln (und eine intensive Beschäftigung mit empirischem Textmaterial) oder auch so "unprogressives" Wissen wie das der Dokumentationswissenschaft.

2. Beispiele

Da es nun einmal keine Theorie der Massentextdatenverarbeitung gibt, möchte ich im folgenden an vier verschiedenen linguistischen Problembereichen aus dem Englischen anhand von Beispielen aufzeigen, wie man an Lösungen herangehen kann. Die folgenden Ausführungen sollen auch ein Bild vermitteln, wie sehr die Kreativität des empirisch arbeitenden Sprachwissenschaftlers gefordert ist.

Ich werde vier verschiedene Ansätze vorstellen, die wir bei der morphologischen und syntaktischen Analyse englischer Texte im Bereich der Ernährungswissenschaft, der Elektrotechnik und des Patentwesens hatten. Im einzelnen handelt es sich um eine lexikonfreie Erkennung englischer Adjektive, um Probleme der Nominalisierung im Englischen, um englische Verlaufsformen und um Skopusphänomene bei Präpositionen. Die Beispiele sind ziemlich wahllos und unabhängig voneinander gewählt. Die angesprochenen Problemkreise stellen nur einen ganz kleinen Ausschnitt aus unserem Entwicklungsgebiet dar.

Was sie verbindet, ist ihre Integration in eines der sowohl von der Anwendungsbreite als auch vom funktionalen Leistungsumfang vermutlich größten freitextanalytischen Verfahren mit computerlinguistischen Komponenten, die in den letzten 30 Jahren weltweit entwickelt wurden.

2.1 Lexikonfreie Adjektive-Erkennung

Ein anwendungsunabhängiges Lexikon englischer Adjektive wird wohl mehrere zehntausend Einträge aufweisen müssen. Möglicherweise sind es wesentlich mehr, wenn man daran denkt, daß zu vielen Adjektiven über ein Negationspräfix ein negiertes Adjektiv gebildet werden kann. Gerade im wissenschaftlich-technischen Bereich sind bestimmte Adjektiv-Suffixe sehr produktiv. Exhaustivität ist also gerade hier prinzipiell nicht zu erreichen. Man denke z.B. an den "ic"-Auslaut im Englischen bzw. in den Römischen Sprachen an die entsprechenden Entwicklungen aus lat. "icus".

Für die anwendungsunabhängige Verarbeitung großer Textmengen sollte ein Lexikon vollständig sein. Umfangreiche Lexika sind verbunden mit erheblichen Erstellungs- und Wartungsproblemen. Um einerseits möglichst exhaustiv arbeiten zu können und andererseits die Nachteile großer Lexika zu vermeiden, sind wir wie folgt vorgegangen:

Wir verwendeten verschiedene kleine sog. "Negativlisten" und eine umfangreichere sog. "Positivliste". Die Listen sind als Bäume organisiert. Die Negativlisten enthalten zusammen etwa 850 Einträge, die Positivliste enthält etwa 900 Einträge. Letztere führt alle Adjektive auf, die nicht über Auslautregelhaftigkeiten erkannt werden können, die also nicht in systematisierbarer Weise auslauten (z.B. "good", "distinct", "ultimate"). Eine andere Systematik ergibt sich im Gegensatz dazu im Hinblick auf typische Adjektiv-Endungen. Hier kommen die Negativlisten zum Einsatz. Wörter mit den entsprechenden Endungen werden gegen diese Listen geprüft. Sind sie in den Listen enthalten, handelt es sich bei dem betreffenden Wort um kein Adjektiv (deshalb "Negativliste").

Ein Beispiel vorweg: Endet ein Wort auf "ble", handelt es sich in den meisten Fällen um ein Adjektiv, d.h. wir gehen zunächst von dieser Hypothese aus. Nach dem Auslaut-Check für "ble" geht man wie folgt vor: Falls das auf "ble" endende Wort nicht in der "ble"-Negativliste steht, die z.B. "vegetable" oder "marble" enthält, ist es ein Adjektiv. Dabei wird nicht eine große Liste von Adjektiven durchsucht, sondern nur ein Check der Endungsgrapheme vorgenommen und eine sehr kleine Liste von "Gegenbeispielen" überprüft.

Wie das Verfahren im einzelnen abläuft, sei an einem Auszug aus der umgangssprachlichen Beschreibung des Ablaufplans aufgezeigt:

Das zu klassifizierende Wort muß mindestens 3 Buchstaben haben (Wörter mit 1 oder 2 Buchstaben sind keine Adjektive).

Anschließend werden adjektivtypische Auslaute auf 4 Buchstaben geprüft:

-less
-some
-tive

"Nevertheless", "chromosome", "motive" sind in der Negativliste derjenigen Wörter enthalten, die mit diesem Auslaut versehen und keine Adjektive sind. Adjektive gemäß der Regel sind dann z.B. "hopeless", "handsome", "attractive". Die Negativliste "less/some/tive" enthält etwa 20 Einträge.

Es geht weiter mit adjektivischen Auslauten auf 3 Buchstaben:

-ble

Adjektive in dieser sehr umfangreichen und obendrein produktiven Klasse sind z.B. "fashionable" oder "comfortable", aber nicht z.B. das in der Negativliste enthaltene "vegetable". Die Klassifizierung der tatsächlichen "ble"-Adjektive wird über die Überprüfung eines Auslautes, nämlich "ble", erreicht und das Absuchen eines Baumes mit nur etwa 30 Einträgen.

Vergleichbares kann man mit den Auslauten "ary", "ory" und "ant" machen. Beim Überprüfen von Adjektiv-Kandidaten auf "ent" kann man sich

eine große Klasse von Nomina sparen, wenn man, unter der Voraussetzung, daß das Wort auf "ment" endet, erst einmal prüft, ob es nicht identisch ist mit "vehement" oder "clement". In diesen beiden Fällen ist das Wort eine Adjektiv, sonst nicht. Man umgeht mit diesem einfachen Trick vor allem die lateinischen Nominalbildungen auf "ment".

Das geht so weiter für andere Auslaute auf 3 Buchstaben und anschließend für solche auf 2 Buchstaben.

Vorausgesetzt, bis jetzt wurde noch nichts gefunden, geht man, nach Tilgen des Negationspräfixes "un", "in" oder "non" in die Adjektivpositivliste. Dort findet man dann "good", "distinct" oder "ultimate" bzw. nach Tilgen von "un" bei "unhealthy" "healthy". Verschiedene Feinheiten müssen noch berücksichtigt werden, auf die aber hier nicht eingegangen werden kann.

Die genannten Effizienzgründe (kleine Listen, schnelle Abarbeitung, weitgehende Exhaustivität) werden durch folgende linguistischen Überlegungen abgestützt: In den Negativlisten handelt es sich oft um Wörter, die keine echten Suffixe haben (z.B. "seal" in der "al"-Liste) oder um Suffixe, die heute nicht mehr als solche empfunden werden (z.B. "boundary" in der "ary"-Liste). Massenhafte Neubildungen im nominalen und verbalen Bereich sind hier nicht zu erwarten.

Die produktiven und bezüglich der Auslaute offenen Klassen der entsprechend suffigierten Adjektive erfaßt man über den einfachen Trick, der lautet "ist nicht Element der betreffenden Negativliste" in sehr umfassendem Maß. Und das gilt auch für entsprechende Neubildungen, die wiederum weitgehend auf die Auslaute beschränkt sind. Die Adjektiv-Positivliste ist somit auch diachron eine weitgehend geschlossene Liste. Sollten in ihr Einträge fehlen, dann nicht, weil es sich um eine Neubildung handelt, sondern weil die Ersteller der Liste nicht sauber gearbeitet haben.

2.2 Probleme mit Nominalisierungen

Im folgenden soll ein Problem aus der englischen Nominalphrasen (NP)-Syntax vorgestellt werden, das sich speziell im IR-Kontext ergab. Wieder ging es darum, ein schnelles Verfahren zu entwickeln, das möglichst viele Fälle abdeckt. Daß das Verfahren völlig fehlerfrei arbeitet, war keine Anforderung.

Zunächst ein paar Worte zu einer grundlegenden strukturellen Regelmäßigkeit, die wir für die NP-Normalisierung verwendeten. Ein immer wieder genutztes, einfaches Prinzip bei der syntaktischen Analyse von NP ist das Prinzip der Spezifikation bzw. die Head-Modifier-Relation. Wir haben diese Relation bei der Darstellung in Dependenzbäumen für NP genutzt, und zwar, sehr grob skizziert, in folgender Weise:

Folgen von bedeutungstragenden Wörtern in NP's spezifizieren i.a. von links nach rechts, Folgen von durch Präpositionen getrennten bedeutungstragenden Wörtern spezifizieren von rechts nach links. Dieses Grundprinzip erlaubt in erster Näherung eine sehr effiziente, einfache und damit schnelle Strukturierung von NP: "storage of milk" entspricht "milk storage", wenn man eine gemeinsame Struktur "storage < milk" ("milk spezifiziert, ist Modifier zu "storage")

Nach dem Prinzip gilt die Gleichwertigkeit zwischen folgenden NP:

- transport problem of Emmental cheese
- problems of Emmental cheese transport
- Emmental cheese transport problems
- problems of transport of cheese from Emmental
- usw.

Die vereinfachte, die reine Abfolge berücksichtigende Dependenzstruktur lautet:

problems < transport < cheese < Emmental

Von Ausnahmen zu dieser Regel wird noch die Rede sein. Zuerst soll der linguistisch motivierte Ansatz behandelt werden, der uns half, IR-spezifisch motivierte Fehler zu reduzieren. Es handelte sich um Fehler, die entstanden, obwohl linguistisch korrekte Dependenzstrukturen erzeugt worden waren.

Vom Standpunkt des IR sind die Sachverhalte "soluble protein" und "protein solubility" gleichwertig: Wer nach dem einen sucht, interessiert sich auch für das andere – zumindest in einer Datenbank mit mehreren Hunderttausenden von Objekten. Linguistisch gesehen, sind die NP, gleiche Wortstämme vorausgesetzt, unterschiedlich: Im einen Fall hat man "protein < solub", im anderen "solub < protein". Verwendet man im Suchauftrag die eine Formulierung und trifft im Text auf die andere Formulierung, passen die strukturellen Patterns nicht zueinander. Das Dokument wird nicht gefunden, obwohl bei der linguistischen Strukturierung nichts falsch gemacht wurde.

Über verschiedene Elementartransformationen ist es möglich, die Strukturen so zu serialisieren, daß das gewünschte Matching erfolgreich ist. "Water solubility of cheese" ist dann identifizierbar mit "water soluble cheese", zwei von der Dependenzstruktur her gesehen völlig verschiedene Phrasen. Werden sie vorher entsprechend transformiert, sind sie nach Anwendung der Normalisierungsalgorithmen äquivalent. Wie das im einzelnen geschieht soll hier nicht erörtert werden. Eine linguistisch interessante Frage unter Effizienzgesichtspunkten ist die, aufgrund welcher Kriterien transformiert werden soll.

Vom Adjektiv zum Nomen führt keine Regelmäßigkeit, zumindest keine, die mit einfachen Mitteln und schnell in den Griff zu bekommen wäre. Man kann nicht pauschal jedes Adjektiv nehmen und es mit "ity" suffigieren. Die Liste der Ausnahmen ist schon aus praktischen Gründen nicht möglich, denn

dazu benötigte man eine vollständige Liste aller Adjektive. Die müßte dann überprüft und reduziert werden auf die Einträge, die mit "ity"-Ableitung existieren.

Man kommt aber weiter mit folgendem Trick: Man sucht mit dem Stämmen derjenigen Nomina auf "ity", denen Adjektive zugrundelegen, kommt also auf diese Weise von "solubility" zu "soluble". Das Problem scheint im ersten Moment nur verschoben, denn welche sind diejenigen Nomina auf "ity", die auf Adjektive zurückgehen? Eines kann durch einen Blick in ein rückläufiges Wörterbuch sehr schnell festgestellt werden: Diese Klasse wird auf jeden Fall kleiner sein also die Klasse der Adjektive, mit der wir es zuerst versuchen wollten.

Die "ity"-Ableitungen bilden wieder eine produktive Klasse und diese Überlegung ließ uns folgendermaßen vorgehen: Wir erstellten die wesentlich kleinere Klasse derjenigen "ity"-Ausdrücke, denen keine Adjektiv zugrundeliegt bzw. wo sich das Adjektiv so weit vom Nomen entfernt hat, daß eine inhaltliche Gleichheit nicht mehr besteht ("quality", "university" etc.). Diese Klasse ist geschlossen, Neubildungen sind hier kaum zu erwarten. Alles, was auf "ity" endet und nicht in dieser Negativliste ist, geht auf ein Adjektiv zurück. Geringe Fehlerraten sind einkalkuliert.

Auf diese Weise gelang es, linguistisch korrekte Ergebnisse, die zu Fehlern im IR-Kontext geführt hatten, über linguistische Zusatzregeln zu korrigieren. Man könnte nun zufrieden sein, wenn da nicht ganz analoge Fälle wären, die im Grunde auch behandelt werden müßten. Es gibt selbstverständlich noch andere Zusammenhänge zwischen Adjektiven und ihren nominalen Entsprechungen. Im Fall von "ness"-Ableitungen ("hardness", "hard") waren sie noch systematisierbar, andere Fälle konnten angesichts der Anforderung, die die enormen Textmengen stellten, nicht gelöst werden.

Aus der Sicht der Empirischen Linguistik ist es nun interessant, dieses nicht ganz zufriedenstellende computerlinguistische Herangehen gleichsam ex post zu rechtfertigen. Insbesondere die sprachgeschichtlich älteren Bildungen ("death", "dead") entziehen sich fast vollständig einer string-orientierten Systematik wie in den angegebenen Fällen. Diese Bildungen sind es aber häufig, die sich im Lauf der Zeit zum Teil so weit auseinander entwickelt haben, daß ihre Rückführung aufeinander oft problematisch ist. Das germanische "ness" läßt sich gerade noch systematisieren (unter Verwendung einer Negativliste). "ity", der französische Auslaut als historisch jüngste Bildung, ist weitgehend systematisierbar, bringt weit mehr Bildungen als "ness" und ist als Suffix auch heute noch produktiv.

Verschiedene andere Auslaute wurden von uns durchgetestet, ohne Erfolg. Auch die gewonnenen Ergebnisse bezüglich "ity" und "ness" müßten im IR-Kontext noch extensiv untersucht werden: Nicht

so sehr, um sie linguistisch zu verfeinern (das kann man auf jeden Fall), sondern um zu prüfen, ob der versierte Datenbankrechercheur überhaupt, und wenn ja, in welchem Maß, mit der Problematik Nomen-Ableitung/Adjektiv in Schwierigkeiten gerät.

Wenn man als Linguist einem Datenbankbenutzer mehr anbietet, muß man davon ausgehen, daß der Benutzer von allen sprachlichen Möglichkeiten Gebrauch macht. Und dann kann man nicht sagen, bei diesen grammatischen Phänomenen hat man vorsichtig zu sein, bei jenen nicht. Unter Umständen kann es durchaus sinnvoll sein, aufgrund von Anwender-Tests eine noch so ausgefeilte linguistische Funktion wieder zurückzunehmen.

2.3 ING-Formen: verbal oder nicht-verbal?

Von einem anderen Problem, das mit syntaktischen Mitteln unlösbar schien, soll nun die Rede sein. Da es sich um ein sehr häufig auftretendes Phänomen handelte (etwa 30 % aller Nominalphrasen enthielten die entsprechenden Formen), mußten wir angesichts hoher Fehlerraten zu einer Lösung kommen. Im Hinblick auf die Leistungsfähigkeit des Gesamtsystems hätte man mit den Fehlern leben können: Die Mehrzahl der 30 % von Fällen war korrekt. Langjährige Erfahrung mit "unlösbaren" Problemen im Bereich der Massentextdatenverarbeitung, eine daraus resultierende Skrupellosigkeit im Umgang mit sprachlichen Phänomenen und linguistische Neugier ließen uns dann doch zu brauchbaren Ergebnissen kommen.

Das Problem waren Formen auf "ing", genauer: jene "ing"-Formen, die nicht lexikalisiert sind. Letztere wurden über eine Negativliste ausgefiltert (sie enthält z.B. "spring", "swing"). Die Frage war, welche Abhängigkeitsstruktur für komplexe Nominalphrasen anzusetzen ist, die eine nicht-lexikalisierte "ing"-Form enthalten.

"Singing boys" hat die erwähnte Head-Modifier-Struktur: "boys < singing", entspricht also der grob skizzierten Links-Rechts-Serialisierung. Es reiht sich in Fälle wie "nice boys", "school boys" usw. Immer bleibt "boys" der Head. Das Prinzip ist nicht generell übertragbar: "singing songs" spezifiziert von rechts nach links. Die Head-Modifier-Struktur ist "singing < songs", paraphrasiert "the singing of songs".

Syntaktisch scheint das Problem zunächst schwer lösbar, semantisch ist es vor dem Hintergrund der Anforderungen an die Erschließung sehr großer Textmengen nicht zu bewältigen. Die erste Überlegung war nun die, ob es eine bestimmte Serialisierungsrichtung bei diesem Konstruktionstyp gibt. Dazu selektierten wir auf der Grundlage von etwa 240.000 Phrasen per Programm eine repräsentative Anzahl von NP, die einen nicht-lexikalischen "ing"-Ausdruck enthielten.

Die intellektuelle Auswertung von etwa 1.000 Phrasen dieses Typs ergab jedoch keine signifikante Verteilung. Hätten z.B. 80 % der Konstruktionen

von rechts nach links spezifiziert, hätte man für diesen Konstruktionstyp eine spezielle Serialisationsregel aufgestellt und in weiteren Testläufen die restlichen 20 % auf eventuell vorhandene strukturelle Eigenheiten untersucht. Das hätte möglicherweise wieder zu Zusatzregeln geführt.

Ein anderes Vorgehen war also angesagt. Wir untersuchten zunächst etwas wahllos Konstruktionen mit nicht-lexikalisierten "ing"-Formen und versuchten Regelmäßigkeiten herauszufinden. Anhand entsprechender Testläufe und deren Auswertungen zeigte es sich, daß unter bestimmten kategorialen Kontextbedingungen eine ganz bestimmte Serialisierung sehr wahrscheinlich war. Handelt es sich z.B. bei dem Wort, das der "ing"-Form folgt, um ein Adjektiv oder ein Partizip-Perfekt-Passiv, besteht eine hohe Wahrscheinlichkeit für eine verbale Lesart und damit für die Serialisierung von rechts nach links ("driving big cars").

Ohne irgendwelche semantische Information liefert der kategoriale Hinweis "Adjektiv" bei "big" im letztgenannten Beispiel die Information, daß die erwähnte Serialisierung mit hoher Wahrscheinlichkeit durchbrochen ist. Die mit außerordentlich hohem Textaufwand entwickelten Regeln gelten auch für erweiterte Kontexte des genannten Typs. So ist z.B. die Wahrscheinlichkeit, daß die "ing"-Form verbalen Charakter hat, auch dann hoch, wenn der "ing"-Form und dem nachfolgenden Bedeutungswort ein Plural vorausgeht ("boys bringing newspapers"). Das setzt eine Pluralerkennung voraus, die listen- und stringorientiert implementiert wurde.

Die Problematik "singing boys" versus "singing songs" hat mit deverbalen Ableitungen bei transitiven Verben zu tun und deren Agens/Patiens-Struktur. Schwierigkeiten im Zusammenhang mit Ableitungen aus transitiven Verben in kompositionellen Bildungen sind auch aus dem Deutschen bekannt. Auch hier gibt es den unangenehmen Fall, daß man nicht weiß, was Agens und was Patiens ist: "Wasserverschmutzung" ("Wasser" als Agens oder Patiens) versus "Benzinverschmutzung" ("Benzin" als Agens oder Patiens). Ähnlich liegt der Fall beim eingangs geschilderten Problem mit "singing boys" und "singing songs": je nach Agens- oder Patiens-Interpretationen ergeben sich unterschiedliche Dependenzstrukturen.

Das Problem war im Englischen zumindest teilweise lösbar vor dem Hintergrund der Forderung nach großer Abdeckungsbreite der Regeln und unter Berücksichtigung der Effizienzanforderung im IR. Bestimmte Kontexte legen die eine Lesart näher, bestimmte eine andere. Die Kontexte zu eruieren, war sehr testaufwendig und erforderte nicht zuletzt enorme Geduld bei den Werkstudenten, die mit der Auswertung endloser Testläufe für immer wieder neue Regelhypothesen betraut waren...

Die Ergebnisse in der Form von Regeln waren sehr einfach, die Strukturierungsalgorithmen

sind entsprechend schnell und robust. Mit den Regeln für "ing"-Formen konnte die Fehlerrate von 28 % auf 7 % reduziert werden. Geht man davon aus, daß gut 20 % der Fälle nicht regelhaft lösbar sind, hat man die Gesamtfehlerrate um etwa die Hälfte reduziert. Jede weitere Verbesserung ist mit Sicherheit mit einem enormen Zuwachs an Lexikonaufwand verbunden, bringt eventuell neue Fehler und macht das System langsamer. Eine geringe Fehlerrate wird auch dann bleiben.

2.4 Skopusprobleme bei Präpositionen

Ein weiteres Beispiel soll aus der Syntax von Präpositionalphrasen vorgestellt werden. Hat man nur syntaktische Mittel zur Dependenzstrukturierung zur Verfügung, gerät man bei der Normalisierung von NP mit mehreren Präpositionalphrasen in Schwierigkeiten.

Bestimmte Phänomene sind grundsätzlich nicht zu strukturieren, wenn man an semantische Ambiguitäten denkt. "The man with the car from Munich" ist allenfalls kontextabhängig zu lösen. Mit semantischen Mitteln lösbar sind Fälle wie "the man with the car without spark plugs". Auch hier kann es sich um den Mann ohne Zündkerzen handeln, wahrscheinlicher ist die Lesart "der Mann mit dem Auto, das keine Zündkerzen hat".

Versucht man die Skopusproblematik, d.h. hier die Problematik der Klammerungsstärke von Präpositionen, ohne Semantik in den Griff zu bekommen, sieht es wieder so aus, als käme man zu keiner Lösung. Eine allgemein anwendbare, semantisch motivierte Skopustheorie für Freitext-Datenbanken mit mehreren Millionen von NP aus beliebigen Anwendungsbereichen ist nicht nur aus Effizienzgründen schwer vorstellbar.

Um die Problematik zumindest zu entschärfen, sind wir wie folgt vorgegangen. Aus den uns zur Verfügung stehenden maschinenlesbaren Texten haben wir mehrere Tausend Phrasen per Programm selektiert, die dem Kriterium "Bedeutungstragendes Wort + Präpositionalphrase + Präpositionalphrase + ..." genügten. Eine Menge von 2.500 NP dieser Art wurden dann intellektuell untersucht, indem Klammern eingefügt wurden, die die Skopusverhältnisse von Präpositionen explizierten. Mit den in dieser Weise aufbereiteten Daten arbeiteten dann wieder Programme, die den Skopus im Hinblick auf die verwendeten Präpositionen auswerteten.

Unsere Arbeitshypothese war die, daß möglicherweise bestimmte Präpositionen stärker klammern als andere, bzw. auch in der Kombination mit anderen Präpositionen stärker klammern als bestimmte Präpositionen. Eine weitere Überlegung betraf die Reihenfolge der Präpositionalphrasen: Sollte die Klammerung auch davon abhängig gemacht werden können, wäre dies ein einfaches formales Mittel gewesen, um zumindest ein wenig Struktur in die Dependenzverhältnisse von Präpositionalphrasenfol-

gen zu bringen. Andere Überlegungen betrafen bestimmte Präpositionen wie "by" zusammen mit in bestimmter Weise suffigierten Nomina (z.B. "ion"): Kann man davon ausgehen, daß in einem solchen, formal sauber eingrenzenden Fall die "by"-Präpositionalphrase nur den Head der gesamten NP spezifiziert und nicht den unmittelbaren Vorgänger?

Das Ergebnis der Untersuchung war ernüchternd. Das einzige Ergebnis, das allgemein verwertbar war, war eine Erkenntnis im Hinblick auf die häufigste Präposition, nämlich "of". "of" klammert in den meisten Fällen relativ stark. Mit dieser Erkenntnis bekommt man zumindest ein wenig Struktur in die Abfolge der Präpositionalphrasen. Ein Beispiel wäre "the effect of acid on the solution" (die Wirkung auf die Lösung) im Gegensatz zu "the effect on acid in the solution" (hier ist nicht mehr die Rede von einer Wirkung auf die Lösung, sondern von einer Säure in einer Lösung). In jedem Fall geht es um einen Effekt im Hinblick auf eine Säure.

Berücksichtigt man, daß Gefüge mit mehr als zwei Präpositionalphrasen seltener sind, ist dieses Ergebnis angesichts der sehr häufigen Präposition "of" dennoch ermutigend.

Das Ergebnis war für uns mit einer nicht uninteressanten linguistischen Überlegung verbunden. "Of" ist die häufigste und, von ihrem semantischen Gehalt her gesehen, allgemeinste Präposition. Die nachträgliche linguistische Rechtfertigung unseres Ergebnisses war nun die, daß wir davon ausgingen, daß in diesem Fall irgendeine "strukturelle Markiertheit" vorliegt. Vergleichbares kennt man aus der Theorie des Sprachwandels: Sehr vereinfacht formuliert haben morphologisch unmarkierte Einheiten eine eher feste Satzstellung, die morphologisch markierten sind diesbezüglich freier.

Unsere Überlegung war nun bezüglich "of" die, daß durch die speziellere Semantik anderer Präpositionen strukturelle Bezüge weitaus klarer sind als bei "of". Die Markiertheit, das was "of" gleichsam "fehlt" im Hinblick auf andere Präpositionen, ist seine eindeutige Klammerung. Diese Erkenntnis, so sie richtig ist, ist sicher nicht weltbewegend und, so sie richtig ist, möglicherweise bekannt. Immerhin war es nützlich, Erkenntnisse der empirischen Linguistik in einem uns bisher nicht geläufigen Kontext zu überdenken, um rein statisch gewonnene Ergebnisse intuitiv plausibel abzustützen.

3. Anwendungen

Auf der Grundlage eines Kategoriensystems, in das die Ausführungen zur Adjektiverkennung einen kleinen Einblick gaben, wurde ein Partielles Parsing implementiert, das zu folgenden Anwendungen geführt hat:

1) Wir entwickelten ein experimentelles IR-System über folgende Datenbanken:

- 270.000 Titel aus FSTA (Ernährungswissenschaft)

- 200.000 Index Phrasen aus INSPEC (Elektrotechnik)
- 20.000 Abstracts aus FSTA (Ernährungswissenschaft)

Das Verfahren erlaubt die natürlichsprachliche Eingabe eines Suchauftrags in beliebiger Formulierung über Terminal, macht eine syntaktische Analyse, erzeugt Abhängigkeitsstrukturen der erhaltenen NPs des Suchauftrags und vergleicht ihn mit den Dokumenten, die ebenfalls syntaktisch analysiert worden sind. Die Dokumente werden gemäß ihrer Entsprechung zum Suchauftrag in verschiedenen Güteklassen ausgegeben (Ranking). Hierbei werden über komplizierte Matching-Verfahren für NP die dependentiellen Bezüge innerhalb dieser NP berücksichtigt.

Das System läuft auf einer Siemens BS2000 Main-Frame-Maschine und ist in Pascal geschrieben. Z.Z. analysiert und normalisiert die syntaktische Erschließung Textsegmente von 200 Wörtern pro CPU-Sekunde. Die Ende 1987 abgeschlossene Reimplementierung wird etwa 5 Mal so schnell sein, so daß wir in einer CPU-Sekunde 10 Abstracts mit 100 Wörtern syntaktisch erschließen werden.

2) In der Entwicklung sind derzeit auf der Basis der Syntexanalyse und der NP-Normalisierung Verfahren zur Volltextklassifizierung. Die Volltexte werden syntaktisch analysiert, normalisiert und dann bezüglich ihrer Abhängigkeitsstrukturen im Hinblick auf Head- und Modifier-Häufigkeiten ausgewertet. Die Erschließungszeiten entsprechen denen der IR-Syntax. Die Verfahren werden auf PC's implementiert.

3) Als Teil größeren Systems zur Indexierung, Thesaurusbildung und zur Suchauftragsformulierung sind die syntaktischen Verfahren mit zwei weiteren Komponenten seit einiger Zeit im Großtest unter realen Anwenderbedingungen.

4. Interessenten

Die Syntaktische Erschließungssoftware zusammen mit einer automatischen Stammformenerzeugung und einer rein statistischen Komponente wird derzeit im Rahmen des schon erwähnten integrierten Projekts (REALIST: retrieval aids by linguistics and statistics) vom US-Handelsministerium für Patentanwendungen in allen amerikanischen Bundesstaaten getestet. Die Verfahren sind bei Datenbankproduzenten und bei Anwendern im Bereich des "electronic publishing" anlässlich von Vorführungen auf außerordentliches Interesse gestoßen.

Anmerkung:

Die geschilderten Verfahren werden z.Z. in den Forschungslabors bei Siemens entwickelt von Gerda Ruge, Alois Baumer, Christoph Schwarz.

Weitere Arbeiten sind beschrieben in Schwarz / Thurmair (Eds.), *Informationslinguistische Texterschließung*, Hildesheim, 1986. In diesem Sammelband finden sich neben Literaturangaben und den Beschreibungen weitere Verfahren auch mehr theoretische Ausführungen zu den Grundlagen der im vorliegenden Beitrag geschilderten Ansätze.

Literatur:

ACM86:

Proceedings ACM Conference on Research and Development in Information Retrieval, Pisa, 1986

CON87:

CONTEXT, Natural Language Full-Text Retrieval System. CONTAHAL LTD., 1987

DID83:

Dillon M.; McDonal L.K.: Fully Automatic Book Indexing. Journal of Documentation 39B, 1983, pp. 135-154

DOS82:

Doszkocs T.E.: From Research to Application: The CITE Natural Language Information Retrieval System. In: ACM Conference on Research and Development in Information Retrieval, Proceedings, Berlin, 1982, pp. 251-262

DOS 86:

Doszkocs, T.E.: Natural Language Processing in Information Retrieval. Journal of the American Society for Information Science 37/4, 1986, pp. 191-196

KOL84:

Koll, M.B. et al.: Enhanced Retrieval Techniques on a Microcomputer. Proceedings of Natural Online Meeting, 1984

SAL66:

Salton, G.: Automatic Phrase Matching, in: Hays (Ed.): Readings in Automatic Language Processing, New York, 1966, pp. 169-188

SCT86:

Schwarz, C.; Thurmaier, G.: Informationslinguistische Texterschließung, Hildesheim, 1986

METHODEN ZUR TEILAUTOMATISCHEN ERSTELLUNG VON TRANSFERWÖRTERSÜCHERN *

Ulrich Heid

Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, Computerlinguistik
Keplerstraße 17, D-7000 Stuttgart 1

Ursula Eckert

IAI-EUROTRA-D

Martin-Luther-Straße 14, D-6600 Saarbrücken 3

Die Erstellung von Transferwörterbüchern ist ein aufwendiger Arbeitsschritt bei der Entwicklung eines transferbasierten automatischen Übersetzungssystems. Dies gilt besonders für das europäische Übersetzungsprojekt EUROTRA, das alle EG-Sprachen berücksichtigen soll. In dem deutschen Teilprojekt EU ROTRA-D wurde nun vorgeschlagen, die Tatsache zu nutzen, daß Lexeme innerhalb der Einzelsprachen in syntaktischer und semantischer Hinsicht bereits sehr detailliert beschrieben werden müssen. Die Hypothese ist, daß Lexeme, die für zwei Sprachen analoge Beschreibungen haben, unter bestimmten Bedingungen als Äquivalente angesehen werden können. Im vorliegenden Bericht wird ein Experiment zur Überprüfung dieses Vorschlags von EUROTRA-D mit ersten Ergebnissen vorgestellt. Das Ziel, einen Teil der Transferwörterbücherstellung zu automatisieren um so den Arbeitsaufwand bei der Wörterbücherstellung zu reduzieren (und sich von der zweisprachigen Kompetenz des Lexikographen weitgehend unabhängig zu machen) scheint erreichbar.

O. Einleitung

0.1. In der transferbasierten maschinellen Übersetzung hat das Transferwörterbuch, ähnlich dem zweisprachigen Wörterbuch, die Aufgabe, quellsprachlichen Lesarten von Lexemen zielsprachliche Äquivalente zuzuordnen. Dazu ist eine übersetzungsbezogene Disambiguierung von Lexemen beider Sprachen notwendig.

Bei der Erstellung von Transferwörterbüchern wird meist von der zweisprachigen Kompetenz des Lexikographen und von direkt auf die Wörterbücherstellung bezogener kontrastiver Arbeit ausgegangen. Diese nahezu ausschließlich auf der lexikographischen Kompetenz beruhende Vorgehensweise ist sehr arbeitsaufwendig: In einem multilingualen Projekt wie EUROTRA ist der Aufwand besonders groß, da jede Sprachgruppe in der Lage sein müßte, Transferwörterbücher für die Übersetzung aus allen zu bearbeitenden Sprachen zu erstellen.

Ziel des Experiments, das im Folgenden beschrieben werden soll, ist es, den Vorschlag von EUROTRA-D zu überprüfen, einen Teil der Transferwörterbücherstellung zu automatisieren um so den Arbeitsaufwand bei der Wörterbücherstellung zu reduzieren und sich von der zweisprachigen Kompetenz des Lexikographen weitgehend unabhängig zu machen. Die Transferwörterbücherstellung soll nach diesem Vorschlag zu einem guten Teil ein Nebenprodukt der einzelsprachlichen Beschreibung von Lexemen sein. Diese Beschreibung ist ohnehin für Analyse- und Synthesephase des maschinellen Übersetzungsvorgangs notwendig, so daß die für diese Phasen benötigten Informationen auch für die Äquivalenzordnung ausgenutzt werden können.

0.2. Der Vorschlag zielt darauf ab, daß Lexeme innerhalb der Einzelsprachen in syntaktischer und

semantischer Hinsicht detailliert beschrieben werden. Von den Beschreibungen wird angenommen, daß sie für verschiedene europäische Sprachen gleichermaßen gelten, so daß Lexeme zweier Sprachen, die analoge Beschreibungen haben, unter bestimmten Bedingungen als Äquivalente angesehen werden können [1].

Arbeitstechnisch betrachtet, stellt man sich vor, daß die Lexeme der beteiligten Sprachen jeweils separat nach einem gemeinsamen System klassifiziert und beschrieben werden. Dieser Schritt erfolgt ohne Blick auf die Übersetzung, nur nach den Vorgaben des Beschreibungssystems. Zur Erstellung des Transferwörterbuchs werden auf der Basis von Äquivalenzangaben aus gängigen zweisprachigen Wörterbüchern alle Lesarten potentieller Äquivalente mit ihren Beschreibungen ermittelt. Ein Matching-Mechanismus überprüft dann die Beschreibungen und sucht die Paare von quell- und zielsprachlichem Lexem heraus, die dieselben (oder zusammenpassende) Beschreibungen haben.

0.3. Wir berichten im Folgenden über ein Experiment, in dem dieses Vorgehen anhand der deutsch-französischen Übersetzung von Verben angewandt wurde; dazu werden zunächst die Grundlagen der Beschreibung dargestellt, dann konkrete Ergebnisse am Beispiel eines Verbs aufgeführt. Schließlich wird eine vorläufige statistische Auswertung gegeben. Da das Experiment noch nicht abgeschlossen ist, werden in der Diskussion auch die geplanten weiteren Arbeiten angesprochen.

1. Grundlagen des Experiments

1.1. Die Beschreibung der Lesarten der deutschen und französischen Verben erfolgt in zwei Stufen.

Zunächst werden Lesarten auf der Basis einer syntaktisch-semanticen Beschreibung ermittelt. Diese Lesartenbeschreibung erfolgt parallel einerseits nach der Verbklassifikation, die STEINER 1986 [2] im Rahmen von EUROTRA-D vorgeschlagen hat, und die auf der Systemic Functional Grammar [3] beruht (im Folgenden gelegentlich als SFG-Ansatz bezeichnet). Andererseits nach Prädikat-Argument-Strukturen, wie sie in der Lexikalisch-Funktionalen Grammatik verwendet werden [4].

Die zweite Stufe stellt eine Beschreibung der semantischen Vereinbarkeitsbeziehungen zwischen Verben und ihren Valenzkomplementen dar. Hierbei wurde ein System semantischer Merkmale verwendet, das von ZELINSKY-WIBBELT 1986 [5] (ebenfalls im Rahmen von EUROTRA-D) zur Klassifikation von Nomina entworfen wurde. Mit dieser semantischen Beschreibung können Selektionsrestriktionen über Verbkomplemente angegeben werden. Experimente, die HIRSCHMAN/GRISHMAN/NHAN 1986 zum Parsing von fachsprachlichen Texten [6] angestellt haben, haben eine Filterwirkung semantischer Merkmale gezeigt: die Zahl falscher Analysen wurde stark reduziert, wenn Selektionsrestriktionen verwendet wurden. In analoger Weise soll im hier vorgestellten Experiment überprüft werden, ob mittels Selektionsbeschränkungen unerwünschten Übersetzungsvorschläge ausgeschieden werden können, dort wo neben dem korrekten auch falsche Äquivalentvorschläge vorliegen.

1.2. Syntaktisch-semantiche Verbklassen: Lesartentrennung und Äquivalenzbedingungen

Die Verbklassifikation für Deutsch und Französisch wurde parallel nach der Systemic Functional Grammar (Abkürzung SFG) und nach der Lexical Functional Grammar (Abkürzung LFG) durchgeführt. Bei beiden Beschreibungssystemen mußten zuerst Bedin-

gungen für das Vorliegen unterschiedlicher Lesarten formuliert werden.

1.2.1. Semantisch-syntaktische Verbklassen nach STEINER 1986

Für die SFG-Beschreibung gelten zwei Lesarten eines Verbs als verschieden, wenn

- sie unterschiedlichen Prozeßtypen (= Verbklassen) zugehören,
- sie dem selben Prozeßtyp zugehören, aber unterschiedlichen Untertypen
- die Anzahl der Rollen unterschiedlich ist (Beschreibungsgrundlage: "Maximaler Frame", d. h. Berücksichtigung sowohl obligatorischer als auch fakultativer Rollen ohne Unterschied)

Abbildung 1 stellt Beispiele für diese Bedingungen zusammen, in Abbildung 2 sind nochmals die Verbklassen und ihre Unterklassen mit je einem deutschen und französischen Beispiel zusammengestellt.

1.2.2. Prädikat-Argument-Strukturen im Format von LFG

In der LFG-Beschreibung werden für das Vorliegen zweier unterschiedlicher Lesarten folgende Bedingungen angenommen:

- die Anzahl der Argumente ist gleich (Beschreibungsgrundlage: obligatorische und fakultative Komplemente werden separat berücksichtigt)
- es liegen unterschiedliche Argumenttypen vor:
 - propositionale Argumente
 - prädikative Argumente (ACOMP, NCOMP)
 - Lokal- oder Direktionalangaben als Argumente
 - Adverb(ial) als Argument
- unterschiedliche Kontrollstrukturen [7]

Die folgende Abbildung 3 zeigt jeweils ein Beispiel für diese Bedingungen.

- PROZESSTYP :	beobachten, daß es regnet	[MENTAL PROC]
	der Arzt beobachtet den Kranken	[ACTION]
- UNTERTYP :	ich bekomme ein Buch	[RELATIONAL-ASSOCIATIVE]
	das Brot bekommt mir gut	[REL.-CLASSIFICATORY]
- ROLLENZAHL:	das Dokument stellt die Grundlage ... dar	[RELATIONAL-CLASSIFICATORY, 2 Rollen]
	der Redner stellt die Situation als schlecht dar.	[RELATIONAL-CLASSIFICATORY, 3 Rollen]

Abb. 1: Beispiele für unterschiedliche Lesarten nach SFG

RELATIONAL PROCESS	
- LOCATIONAL	er geht in die Stadt il va en ville
- ASSOCIATIVE	er gibt mir ein Buch il me donne un livre
- CLASSIFICATORY	ich nenne ihn den Chef je l'appelle le chef
- EXISTENCE	es gibt noch Kuchen il y a encore du gâteau
MENTAL PROCESS	
- PROCESSOR OR.	Ich sehe, daß er kommt Je vois qu'il vient
- PHENOMENON OR.	Daß du das sagst, gefällt mir Cela me plaît que tu dises cela
- TWO PHENOMENA	Daß er kommt, bedeutet, daß er sich freut Qu'il vienne, signifie qu'il est heureux
COMMUNICATION	
	Er sagt: "Es regnet" Il dit: "Il pleut"
ACTION	
	Er malt ein Bild Il peint un tableau

2 Deutsche und französische Beispiele für die Verbklassen (=Prozesstypen) der Systemic Functional Grammar

- ARGUMENTZAHL:	ich bemerke, daß ein Zug kommt [2 Arg.] ich bemerke am Geräusch, daß ein Zug kommt. [3 Arg.] (Drittes Argument fakultativ)
- ARGUMENTTYP:	propositional: Ich merke, daß es regnet prädikativ: Ich bezeichne ihn als fleißig Lokal-/Direkt.: Ich wohne in Saarbrücken Adverb als Arg.: Das Kleid steht ihr gut.
- KONTROLLE:	Ich bitte den Chef, zur Sitzung zu kommen Ich bitte den Chef, zur Sitzung gehen zu dürfen

Abb. 3: Beispiele für lesartentrennende syntaktische Merkmale in der LFG-Beschreibung

1.2.3. Äquivalenzbedingungen

Die Bedingungen für die Zuordnung von Übersetzungsäquivalenten ergeben sich aus den Annahmen über die Distinktheit von Lesarten. Abbildung 4 zeigt die Äquivalenzbedingungen für die beiden Ansätze.

Hat ein Verb ein Argument, das durch unterschiedliche syntaktische Konstruktionen realisiert werden

kann, zum Beispiel ein propositionales Argument (durch eine Nominalgruppe, einen Infinitiv, *daß*-, *ob*- oder *w*-Satz [8]), so müßte strenggenommen in der LFG-Beschreibung die Übersetzbarkeit jeder einzelnen Realisierungsvariante separat überprüft werden; dies ist geplant, aber noch nicht erfolgt: bisher wird, (wie in der SFG-Beschreibung) nur geprüft, welche Argumenttypen vorliegen.

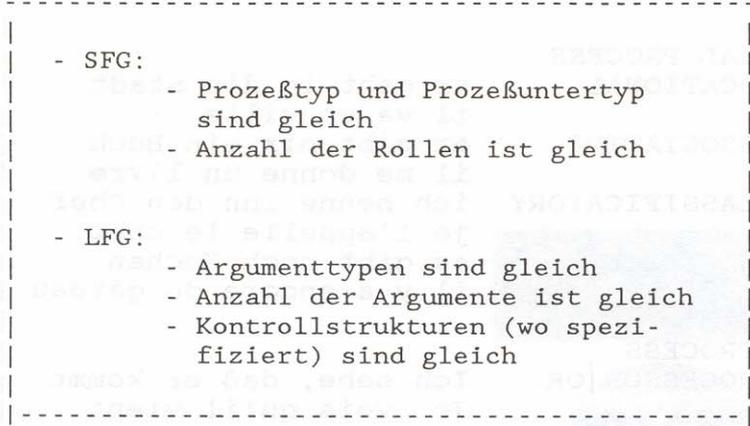


Abb. 4: Äquivalenzbedingungen in der SFG- und LFG-Beschreibung

1.3. Selektionsrestriktionen: Das System semantischer Merkmale und die Äquivalenzbedingungen

Die semantischen Merkmale für die Nomina (Igruppen) werden im Experiment nach dem von ZELINSKY-WIBBELT 1986 vorgeschlagenen System beschrieben. Die Merkmale sind auf mehreren Ebenen angeordnet und unterliegen bestimmten Kombinationsrestriktionen. Abbildung 5 stellt das Merkmalsystem im Überblick dar.

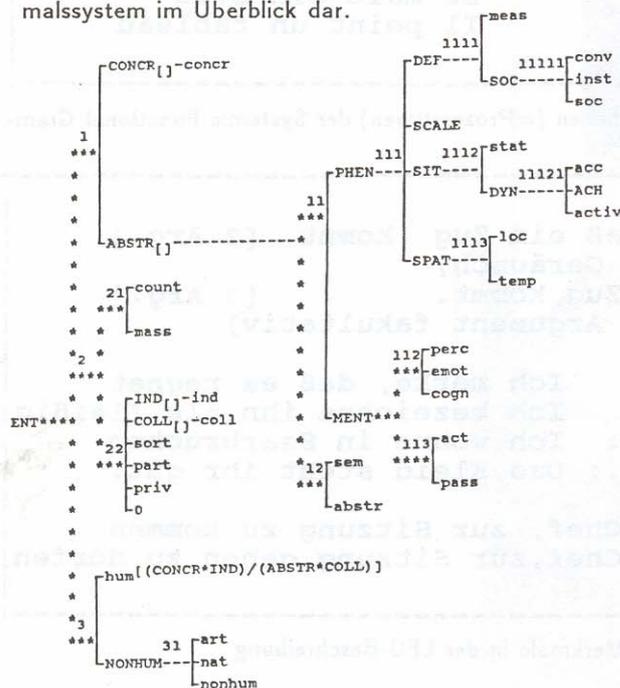


Abb. 5: Das System semantischer Merkmale nach ZELINSKY-WIBBELT (Sterne markieren *und*-Verbindungen, gerade Striche markieren *oder*-Verbindungen)

Manche Verben lassen zu, daß die Nominalgruppe, die eines ihrer Komplemente realisiert, wahlweise entweder das eine oder das andere semantische Merkmal trägt. Die Zuweisung beider Merkmale

ist jedoch in der derzeitigen Form den Merkmalsystems nicht möglich. Bei weiteren Experimenten sollen Alternativen bei der Merkmalszuweisung berücksichtigt werden; bislang wird in solchen Fällen jedoch darauf verzichtet, überhaupt eine Selektionsbeschränkung anzusetzen, was die Filterwirkung des Systems reduziert.

Als äquivalent werden Lesarten mit gleichartigen Selektionsrestriktionen aufgefaßt: zwei Lesarten aus verschiedenen Sprachen werden als potentielle Äquivalente angesehen, wenn die Selektionsrestriktionen gleich oder diejenigen der Zielsprache weniger restriktiv sind, als die Selektionsrestriktionen der Quellsprache.

2. Arbeitsweise und Ergebnisse

2.1. Ablauf des Experiments

Das Experiment ging aus von 49 deutschen Verben, die aus dem von EUROTRA und der EUROTRA-D-Begleitforschung benutzten ESPRIT-Korpus [9] unter dem Aspekt der Polysemie ausgewählt wurden. Sie lieferten insgesamt 174 Lesarten. Zu diesen Lesarten wurde je ein Beispielsatz gebildet.

Diese Beispielsätze wurden manuell ins Französische übersetzt [10]. Stilistisch markierte Übersetzungsvarianten blieben unberücksichtigt, jedoch wurden unmarkierte Synonyme [11] mit einbezogen.

Für die so erhaltenen französischen Verben wurden, wie bei den deutschen Verben auch, die wichtigsten Lesarten nach beiden Beschreibungssystemen beschrieben. 114 französische Verben ergaben 380 Lesarten.

Anschließend wurden jedem deutschen Verb sämtliche französischen Verben gegenübergestellt, die mögliche Äquivalente zu dem betreffenden Verb sind. Für das Verb *schließen* zeigt Abbildung 6 diese Zuordnung, die z. B. aus den üblichen zweisprachigen Wörterbüchern ableitbar ist [12]:

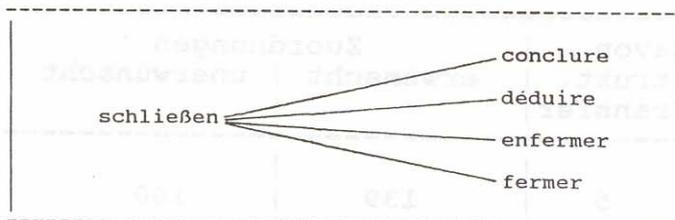


Abb. 6: Lexeme, die Äquivalente von *schließen* sein können.

Nun wurden durch ein Abbildungsprogramm [13] die Beschreibungen der einzelnen Lesarten verglichen, zunächst hinsichtlich der syntaktisch-semantischen Beschreibungen (Semantische Relationen nach SFG und Prädikat-Argument-Strukturen nach LFG), in einem zweiten Schritt innerhalb der auf dieser Basis möglichen Zuordnungen hinsichtlich der Selektionsrestriktionen (semantische Merkmale). In der SFG-Beschreibung, die in Abbildung 7 wiedergegeben ist, hat *schließen* sieben verschiedene Lesarten, zu deren Übersetzung vier verschiedene Verben mit insgesamt zwölf Lesarten zur Verfügung stehen:

Deutsche Lesarten	Lesarten der französischen Äquivalente
SCHLIESSEN 1 action, Agent - Affected "Hans schließt die Tür."	CONCLURE 1 action, Agent - Affected CONCLURE 2 action, Agent - Effected
SCHLIESSEN 2 action, Agent - Effected "Die Firma schließt einen Vertrag."	CONCLURE 3 mental process/2 phen. Processor - Phen1- Phen2
SCHLIESSEN 3 action/one role, Affected only "Das Unternehmen schließt."	CONCLURE 4 action, Agent - Range CONCLURE 5 communication, Sender - Message
SCHLIESSEN 4 action/one role, Attribuant only "Die Tür schließt."	DEDUIRE 1 mental process/2 phen. Phen1 - Phen2
SCHLIESSEN 5 mental process, Processor - Phen1 - Phen2 "Sie schließt aus dem Brief, daß Peter morgen kommt."	DEDUIRE 2 relational/locational 3rd - Attrib. - Loc.
SCHLIESSEN 6 mental process, Processor - Phen1 - Phen2 "Ich schließe von seinem Verhalten auf seinen Charakter."	ENFERMER 1 relational/locational 3rd - Attrib. - Loc. ENFERMER 2 action, Agent - Affected
SCHLIESSEN 7 relational/locational 3rd-p-ag - Attribuant - Location "Die Mutter schließt das Brot in den Schrank."	FERMER 1 action, Agent - Affected FERMER 2 action/one role, Affected only FERMER 3 action/one Attribuant only

Abb. 7: SFG-Beschreibungen der Lesarten von *schließen* und der möglichen Äquivalente

2.2. Ein Beispiel: *schließen*

Der automatische Vergleich der syntaktisch-semantischen Beschreibungen (Zugehörigkeit zu einer Verbkategorie) ergibt das in Abbildung 8 gezeigte Resultat: durch "===" sind erwünschte Zuordnungen bezeichnet, durch "=|=" unerwünschte Zuordnungen.

SCHLIESSEN 1	= = = = ===	CONCLURE 1 ENFERMER 2 FERMER 1
SCHLIESSEN 2	===	CONCLURE 2
SCHLIESSEN 3	===	FERMER 2
SCHLIESSEN 4	===	FERMER 3
SCHLIESSEN 5	=== ===	CONCLURE 3 DEDUIRE 1
SCHLIESSEN 6	=== ===	CONCLURE 3 DEDUIRE 1
SCHLIESSEN 7	= = ===	DEDUIRE 2 ENFERMER 1

Abb. 8: Äquivalentwahl auf der Basis der syntaktisch-semantischen Verbklassifikation (STEINER 1987)

Für die Lesarten *schließen*₁ und *schließen*₇ wurden zusätzlich zum richtigen Äquivalent zwei bzw. eine unerwünschte Lesart ausgewählt; hier sollen die Selektionsrestriktionen eingreifen, um im Idealfall alle unerwünschte Lesarten auszuschließen (ein Vergleich der Angaben zu den jeweiligen semantischen Merkmalen zeigt, daß für *schließen*₁ nur *fermer*₁ und für *schließen*₇ nur *enfermer*₁ als Äquivalente zugelassen werden (vgl. Abbildung 9)).

	FERMER 1	Hans ferme la porte. AGENT AFFECTED (_) (concr)
SCHLIESSEN 1	ENFERMER 2	Hans schließt die Tür. AGENT AFFECTED (_) (concr)
	CONCLURE 1	Une haie enferme le champs. AGENT AFFECTED (concr) (concr)
	CONCLURE 1	L'orateur conclut le congrès. AGENT AFFECTED (hum) (abstr)
SCHLIESSEN 7	ENFERMER 1	La mère enferme le pain 3RD-P-AGENT ATTRIBUANT (hum) (concr)
Die Mutter schließt		dans le placard. LOCATION (_) (concr)
das Brot in den Schrank.	DEDUIRE 2	Le client déduit ses frais 3RD-P-AGENT ATTRIBUANT (hum) (phen)
		de la facture. LOCATION (phen)

Abb. 9: Fälle mit mehreren Zuordnungen auf der Basis der Verbklassifikation: Filterwirkung der semantischen Merkmale

Beschreibungs- system	Ver- ben	Les- arten	davon strukt. Transfer	Zuordnungen	
				erwünscht	unerwünscht
EUROTRA-D	47	174	5	139	100
LFG	47	174	5	143	160

Abb. 10: Automatische Äquivalenzzuordnung auf der Basis der Verbklassifikation nach SFG (STEINER 1986) und nach syntaktischen Funktionen in der Notation von LFG

2.3. Statistische Auswertung

Geht man zunächst nur von der Verbklassifikation (nach dem Vorschlag von EUROTRA-D, bzw. nach der LFG-Beschreibung) aus, so lassen sich folgende statistische Aussagen machen:

Zu mehr als 80 % der deutschen Lesarten können automatisch richtige französische Äquivalente gefunden werden (vgl. Spalte 3 und 5 in Abbildung 10). Die beiden Lesartenbeschreibungssysteme (SFG und LFG) unterscheiden sich hinsichtlich der Zuordnung von richtigen Äquivalenten nahezu nicht [14]. Während nach beiden Systemen etwa gleichviel richtige Zuordnungen getroffen werden, unterscheiden sie sich hinsichtlich der Anzahl der falschen Zuordnungen, die zusätzlich zu der (oder den) richtigen geliefert werden. Nach STEINER 1986 erhält man zusätzlich zu 139 richtigen Zuordnungen 100 falsche, nach LFG zusätzlich zu 143 richtigen 160 falsche (d. h. auf je 100 Richtige nach SFG 70 falsch, nach LFG 118 falsche Zuordnungen; Spalte 5 und 6 in Abbildung 10).

Im Idealfall gibt es zu einer deutschen Lesart eine oder mehrere richtige Äquivalenzzuordnungen und nur diese. Es kommt aber auch vor, daß bei der automatischen Zuordnung außer der richtigen auch eine oder mehrere unerwünschte Übersetzungen ausgewählt werden. Im schlechtesten Fall erhält man nur falsche Ergebnisse oder keine (Null-Matching).

Unter dem Aspekt der Erstellung eines Transferwörterbuches für die Maschinelle Übersetzung sind jene Fälle problematisch, bei denen man die richtige Zuordnung nicht erhält. Nach dem von STEINER 1986 vorgeschlagenen System bekommt man für 21 Lesarten kein Ergebnis (Null-Matching), für 17 weitere nur falsche. Die LFG-Beschreibung liefert 13 Null-Matchings und für 28 Lesarten falsche Ergebnisse. Insgesamt betrachtet ist die Zahl der Fälle, in denen man die richtige Übersetzung nicht erhält, bei Verwendung beider Beschreibungssysteme ungefähr gleich (Addition von Spalte 3 und 4 in Abbildung 11).

Unterschiede zwischen der Kodierung nach STEINER 1986 und derjenigen nach syntaktischen Funktionen ergeben sich bei den Lesarten, denen *nur* die richtige Übersetzung zugeordnet wird. Bei der LFG-Beschreibung ist dies bei 40 Lesarten der Fall, nach STEINER 1986 bei 75.

Die ersten Experimente mit der Verwendung von semantischen Merkmalen zeigen, daß in einem Viertel der Fälle, in denen das syntaktisch-semantische "Matching" zwei oder drei Äquivalenzvorschläge liefert, von denen nur einer erwünscht ist, nach Anwendung des "semantischen Filters" nur noch die richtige Lösung übrigbleibt. Eine erwünschte und eine unerwünschte Lesart bleibt in rund der Hälfte der so gelagerten Fälle übrig.

B.-System	Anzahl von Zuordnungen		
	nur richtige	nur falsche	null
EUROTRA-D	75	17	21
LFG	40	28	13

Abb. 11: Richtige, falsche und Null-Zuordnungen, ausgehend von deutschen Lesarten

3. Diskussion

Eine Reihe von deutschen Lesarten kann mit Hilfe der vorgestellten Mechanismen nicht oder nicht richtig übersetzt werden. Dies sind zum einen Fälle, in denen Lesarten mit Objekt auf Lesarten mit Lokal- oder Präpositionalergänzung abgebildet werden müssen (Abbildung 12):

- | |
|--|
| <ul style="list-style-type: none"> - Er widmet sich seinen Studien.
<ACTION, Agent - Affected> - Il s'applique à ses études.
<RELATIONAL-LOCATIONAL> |
|--|

Abb. 12: Übersetzung eines Verbs mit indirektem Objekt durch ein Verb mit Präpositionalobjekt

In solchen Fällen wirkt sich die strenge Orientierung an syntaktischen (Oberflächen-)Kriterien negativ aus. Die Verben des in Abbildung 13 dargestellten Typs [15] würden z. B. zu systematischen Zuordnungsproblemen führen:

- | |
|---|
| <ul style="list-style-type: none"> - die Diskette mit Daten beschreiben - Daten auf die Diskette schreiben - écrire des données sur la disquette - das Brot mit Käse belegen - Käse auf das Brot legen - mettre du fromage sur la tartine |
|---|

Abb. 13: Deutsche Präfixverben mit SUBJ, OBJ, mit-OBJ und die Simplizia mit SUBJ, OBJ und Direktionalergänzung. Französische Übersetzung

Ebenfalls problematisch sind natürlich Fälle, die strukturellen Transfer verlangen, wie z. B. deutsche Verben, deren Komplement als indirekter Fragesatz realisiert werden kann, oder solche, die Subjektsätze subkategorisieren [16]. Sie bleiben hier unberücksichtigt.

Schließlich muß auch das Verbklassifikationssystem für das Französische angepaßt bzw. erweitert werden [17]. Beispielsweise kann die Definition der Rolle [PHENOMENON], für das Deutsche und das Französische nicht parallel formuliert werden: propositionale Argumente können im Französischen nicht immer durch *que*-Sätze (analog zu deutschen *daß*-Sätzen) ausgedrückt werden; vielmehr muß vorgesehen werden, daß auch Konstruktionen mit *le fait que* als Kriterium für die [PHENOMENON]-Rolle verwendet werden.

Die Experimente zur Verbesserung der Zuordnungsleistung durch die Benutzung von Selektionsrestriktionen sind noch nicht abgeschlossen. Vergleichswerte für die Kombination der syntaktischen Lesartenbeschreibung nach dem LFG-Format mit semantischen Merkmalen fehlen noch. Außerdem soll-

ten bei den Angaben von Selektionsrestriktionen Alternativen vorgesehen werden können (vgl. oben, z. B. : "ein Verb verlangt ein Objekt, das <concrete> oder <local> ist"). Wenn diese Möglichkeiten noch ausgeschöpft werden, so ist damit zu rechnen, daß die Zuordnungsleistung des Systems insofern verbessert werden kann, als der Prozentsatz der Fälle erhöht wird, in denen das erwünschte Äquivalent *und nur dieses* ausgewählt wird.

4. Ausblick

Das Experiment, über das in diesem Beitrag berichtet wurde, ist noch nicht abgeschlossen. Die bisher vorliegenden Ergebnisse zeigen aber, daß die Erstellung von Transferwörterbüchern für die maschinelle Übersetzung nicht rein manuell zu erfolgen braucht. Ein System wie das hier in Umrissen vorgeschlagene könnte dem Lexikographen Äquivalenzvorschläge machen. Die Daten, die das System dazu benötigt, müssen in der einen oder anderen Form, in der Notation des einen oder anderen Formalismus, ohnehin für jedes transferbasierte maschinelle Übersetzungssystem bereitgestellt werden. Die Daten, die das System liefert, entlasten diejenigen, der Transferwörterbücher erstellen soll, von einem Teil des Arbeitsaufwands, der für Routinefälle zu erwarten ist. Die "Fehler", die das System macht, weisen unter Umständen auf Probleme hin, für die strukturelle Transferregeln gesucht werden müssen. Auch in dieser Hinsicht könnte ein derartiges System ein Hilfsmittel bei der Transferwörterbüchererstellung sein.

Anmerkungen

* Überarbeitetes Manuskript eines Vortrags, der anlässlich der 18. Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL) in Heidelberg (1. bis 3. 10. 1987) gehalten wurde. Die Autoren danken B. Böhme, B. Roth und E. Steiner (IAI-EUROTRA-D, Saarbrücken) für ihre Unterstützung bei der Durchführung des Experiments. Besonderer Dank gilt Jörg Schütz, der die Zuordnungsprogramme implementiert hat.

- [1] Mit dieser Beschreibung sollen Generalisierungen formuliert werden, die für mehrere Sprachen gelten. Zielsetzung von EUROTRA ist es ja unter anderem, eine "euroversale" Interface-Struktur zu beschreiben; vgl. JOHNSON/KING/DES TOMBE 1985: 164 über die Interlingua-Ähnlichkeit der Interface-Struktur.
- [2] vgl. STEINER 1986 und STEINER 1987.
- [3] vgl. HALLIDAY 1985, HALLIDAY/FAWCETT 1987
- [4] vgl. BRESNAN/KAPLAN 1982 Die Beschreibung der syntaktischen Funktionen in der Notation von LFG soll hier im Wesentlichen als Vergleichsgröße verwendet werden,

- um zu prüfen, inwieweit sich eine Beschreibung der syntaktischen Funktionen (die – unabhängig vom benutzten Formalismus – jede Beschreibung von Verblesarten für die maschinelle Übersetzung zu leisten hat), für die teilautomatische Erstellung von Transferwörterbucheinträgen eignet.
- [5] vgl. ZELINSKY-WIBBELT 1986
- [6] vgl. HIRSCHMAN/GRISHMAN/NHAN 1986
- [7] Verben, deren Lesarten verschiedene Kontrollstruktur aufweisen, sind selten; im Falle von *bitten* spielt das Vorhandensein des Modalverbs *dürfen* eine Rolle für den Wechsel der Kontrolle. Ebenso ist der Fall *Le chef occupe (son temps/ses collaborateurs) à écrire des articles* marginal.
- [8] *ob*-Satz = indirekte Frage, mit *ob* eingeleitet
(*w*)-Satz = indirekte Frage, mit einem Konstituentenfrage wort eingeleitet
- [9] veröffentlicht in: Amtsblatt der Europäischen Gemeinschaften, C 321 vom 26. 11. 1983 (Luxemburg: EG)
- [10] Die Tatsache, daß nur ein Beispielsatz pro Lesart verwendet wurde, schließt die Behandlung von Kollokationen aus, in einem maschinellen Übersetzungssystem sollten sie aber separat beschrieben werden, da sie für sich äquivalenzstiftend sind und nicht durch Übersetzungsregeln für die als Kollokatoren (im Sinne von HAUSMANN 1985) fungierenden Verben erfaßt werden können.
- [11] "Markiert" im Sinne diasystematischer Angaben, speziell der "Niveaux de langue".
- [12] Auf diese Weise kann das Wissen über potentielle Äquivalente, das in den üblichen zweisprachigen Wörterbüchern enthalten ist, genutzt werden. Ein direkter Einbezug der Informationen über Äquivalenzbedingungen, die in zweisprachigen Wörterbüchern enthalten sind, ist wegen des Mangels an Systematizität in diesen Angaben derzeit sowieso nahezu nicht möglich: im Artikel *abhängen* von WEIS/MATTUTAT 1978 finden sich nebeneinander fünf unterschiedliche Informationskategorien in den Äquivalenzbedingungen: neben der syntaktischen Konstruktionsangabe auch Synonymangaben, Markierungen, Fachbereichsangaben und Lexeme, die als Objekt in Frage kommen.
- [13] Das Programm soll so erweitert werden, daß es vom Beschreibungsansatz völlig unabhängig ist und parallel Beschreibungen nach verschiedenen Ansätzen vergleichen kann.
- [14] Das läßt die Annahme zu, daß für das Sprachenpaar Deutsch-Französisch ein Vergleich der Argumenttypen und -anzahl (unabhängig vom gewählten Beschreibungsansatz) für die Äquivalenzzuweisung bei Verben verwendbar ist. Fälle mit unterschiedlicher Argumentzahl müssen ohnedies durch spezielle Regeln beschrieben werden (z. B. *die Krankenschwester nimmt dem Patienten dem Verband ab* (3-stellig) vs. *l'infirmière débande le patient* (2-stellig).
- [15] Vgl. WUNDERLICH 1987 und auch SEELBACH 1986.
- [16] Im Gegensatz zum Deutschen können indirekte Fragesätze im Französischen nicht Subjekt und fast nie präpositionales Objekt sein (vgl. HUOT 1981: 128-132, 152); für solche Fälle ist es notwendig, strukturelle Transferregeln zu formulieren. Vgl. dazu HEID in Vorbereitung.
- [17] Erste Vorschläge zur Anwendung des Klassifikationssystems der SFG aufs Französische in HEID 1987.

Literatur

BERGENHOLTZ/MUGDAN (Hg.) 1985

Henning Bergenholtz, Joachim Mugdan (Hg.): Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch, 28. – 30. 6. 1984, (Tübingen: Niemeyer) 1985, [= Lexicographica, Series Maior, 3]

BRESNAN (Hg.) 1982

Joan Bresnan (Hg.): The mental Representation of grammatical Relations, (Cambridge, Mass. London: MIT Press) 1982: 173 – 281

BRESNAN/KAPLAN 1982

Joan Bresnan, Ronald Kaplan: "Lexical-Functional Grammar: A formal System for Grammatical Representation", in: BRESNAN (Hg.) 1982: 173-281

GRISHMAN/HIRSCHMAN/NHAN 1986

Ralph Grishman, Lynette Hirschman, Ngo Thanh Nhan: "Discovery procedures for Sublanguage selectional Patterns: Initial Experiments", In: Computational Linguistics 1986, 205 – 215.

HALLIDAY 1985

M. A. K. Halliday: Introduction to functional Grammar, (London: Arnold) 1985.

HALLIDAY/FAWCETT (Hg.) erscheint

M. A. K. Halliday, Robin P. Fawcett (Hg.): New Developments in Systemic Linguistics, (London: Frances Pinter), erscheint

HAUSMANN 1985

Franz Josef Hausmann: "Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels", in: BERGENHOLTZ/MUGDAN (Hg.) 1985: 118 – 129

HEID 87

Ulrich Heid: "Relations sémantiques et fonctions syntaxiques en français - quelques notions sur la classification verbale de la Systemic Functional Grammar", (Stuttgart: EUROTRA-D-Begleitforschung) internes Papier, masch. verf. , 36 SS.

HEID in Vorbereitung

Ulrich Heid: "L'interrogation indirecte en tant que complément verbal – pour une description contrastive"

HUOT 1981

Hélène HUOT: Constructions infinitives du français, (Genève: Droz) 1981

JOHNSON/KING/DES TOMBE 1985

Rod Johnson, Maghi King, Louis des Tombe: "EUROTRA - A multilingual System under development", in: Computational Linguistics 11-2/11-3 (1985): 155 – 169

KLENK/SCHERBER/THALLER (Hg.) 1987

Ursula Klenk, Peter Scherber, Manfred Thaller (Hg.): Computerlinguistik und Philologische Datenverarbeitung, (Hildesheim: Olms) 1987

SEELBACH 1986

Dieter Seelbach: "Eine LFG mit semantisch und syntaktisch homogenen Prädikatsklassen", LDV-Forum 4(1986)

STEINER 1986

Erich Steiner: "Generating semantic structures in EUROTRA-D", in: Proceedings of COLING 1986, (Bonn: IKP) 1986: 304 – 306.

STEINER 1987

Erich Steiner: "Zur Zuweisung satzsemantischer Rollen im Rahmen des MÜ-Projektes EUROTRA-D", in: KLENK/SCHERBER/THALLER (Hg.) 1986

WEIS/MATTUTAT 1978

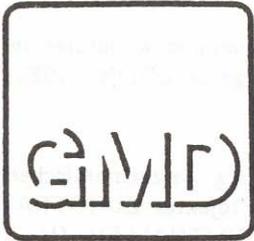
Erich Weis, Heinrich Mattutat: Pons-Globalwörterbuch Deutsch-Französisch, (Stuttgart: Klett) 1978

WUNDERLICH 1987

Dieter Wunderlich: "PASS", Vortrag beim Workshop der EUROTRA-D-Begleitforschung, Dudweiler 18. - 20. Feb. 87. Manuskript.

ZELINSKY-WIBBELT 1986

Cornelia Zelinsky-Wibbelt: "Toward an empirically based system of semantic features", in: Proceedings of COLING 1986.



Das in Darmstadt gegründete

**Forschungsinstitut für
Integrierte Publikations- und Informationssysteme
der Gesellschaft für Mathematik und Datenverarbeitung**

befindet sich im Aufbau und erweitert seinen Mitarbeiterstab.

Die Aufgaben des Institutes umfassen alle Aspekte des elektronischen Publizierens und der Informationssysteme. Spezielle Schwerpunkte bilden objektorientierte, multimediale Wissensbanken, verteilte Systeme, Wissensrepräsentationsmethoden, integrierte Publikationssysteme und intelligente Mensch-Maschine-Schnittstellen einschließlich natürlichsprachlicher Ansätze.

Das Forschungsinstitut sucht

Wissenschaftliche Mitarbeiter,

die mithelfen, dem Institut schnell internationale Geltung zu verschaffen. Promotion ist erwünscht. Jüngeren Mitarbeitern wird die Gelegenheit zur Promotion geboten.

Informatiker(innen)

mit vertieften Kenntnissen in zumindest einem der Gebiete: Datenbanken, Informationssysteme, verteilte Systeme, Wissensrepräsentationen, Compilerbau, Mensch-Maschine-Interaktion.

Computerlinguisten(innen)

mit vertieften Kenntnissen in mindestens einem der Gebiete: wissensbasierte Verarbeitung von Texten, semantische Repräsentationen, Unifikationsgrammatiken, linguistische Pragmatik und Dialogsysteme

Die Bezahlung erfolgt nach BAT (Bundesangestellten-Tarif). Daneben werden die üblichen Sozialleistungen des öffentlichen Dienstes gewährt.

Bewerbungen mit Lebenslauf und Schriftenverzeichnis sind zu richten an:

**G M D
Bereich Darmstadt
Postfach 10 43 26
Dolivostraße 15**

6100 Darmstadt

Phonetische Beiträge zur [REDACTED] maschinellen Spracherkennung

REALFORMANTEN IN DER AUTOMATISCHEN SPRACHERKENNUNG ERSTE ERGEBNISSE DER EXTREMWERTANALYSE IM ZEITBEREICH

Patrick Schweisthal, Thomas Schweisthal, Walter Kopetzky

Arbeitskreis Spracherkennung Sprachgenerierung und Phonetische Datenbanken
der Gesellschaft für linguistische Datenverarbeitung (GLDV)
Institut für Phonetik und sprachliche Kommunikation
Universität München, Schellingstr. 3, D-8000 München 40

Im LDV-Forum 5(1987)1 vom Juni 1987 wurde die Extremwertanalyse im Zeitbereich als ein Verfahren der akustischen Sprachschallanalyse vorgestellt. Darauf aufbauend werden nun die ersten Schritte in Richtung auf eine Transkriptionsmaschine als phonetischer Beitrag zur Spracherkennung dokumentiert. Den Autoren, Mitgliedern des GLDV-Arbeitskreises "Spracherkennung Sprachgenerierung und Phonetische Datenbanken" ist es gelungen, analysierte Realformanten im Bild des Intervallogramms so darzustellen, daß sie mit entsprechenden herkömmlichen Spektralformanten des Sonagramms verglichen werden können. Hierbei wird deutlich sichtbar: Die Strukturen der Realformanten sind im Nasalbereich genau so klar erkennbar wie im Vokalbereich. Dies gilt auch für die Analyse von Frauenstimmen.

In unserem letzten Beitrag postulierten wir analog zu den Formanten des Frequenzspektrums die sogenannten Realformanten (RF). Als Realformant bezeichnen wir die Konzentration zeitlicher Extremwertabstände vergleichbarer Dauern. Im Gegensatz zur herkömmlichen linearen Filterung beläßt ein symmetrisches Glättungsverfahren die Extremwerte in ihrer ursprünglichen Lage. Die zeitlichen Dauerhältnisse bleiben dadurch erhalten. Dies ist Voraussetzung für die Bestimmung der RF. Wird das digitalisierte Signal weniger stark geglättet, so ergeben sich Konzentrationen kurzer Zeitdauern der Extremwerte: die sogenannten RF 2.

Wird das digitalisierte Signal stark geglättet, so ergeben sich Konzentrationen langer Zeitdauern der Extremwerte: die sogenannten RF 1.

Die folgende Darstellung bezeichnet man als Intervallogramm. Vertikal ist der Kehrwert der zeitlichen Dauerhältnisse angetragen (in Hz bei logarithmischem Maßstab), horizontal die zeitliche Dauer der Äußerung (linearer Maßstab, Gesamtdauer einer Darstellung 1,2 s).

Die Graphik enthält den ersten Realformantenormanten (dargestellt von 200 bis 1400 Hz Extremwertfrequenz), sowie den zweiten Realformanten (von 1400 bis 20.000 Hz). Oberhalb des Intervallogramms befindet sich jeweils das Oszillogramm der Äußerung.

Die drei Intervallogramme stellen die Äußerung: "Unser Programm" dreier unterschiedlicher Versuchspersonen (A, B und C) dar, darunter eine weibliche (A) und zwei männliche (B und C).

Zum Vergleich ist neben jedem Intervallogramm das entsprechende Sonagramm derselben Äußerung abgebildet (allerdings bei linearem Frequenzmaßstab, von 0 bis 6000 Hz auf der Vertikalachse).

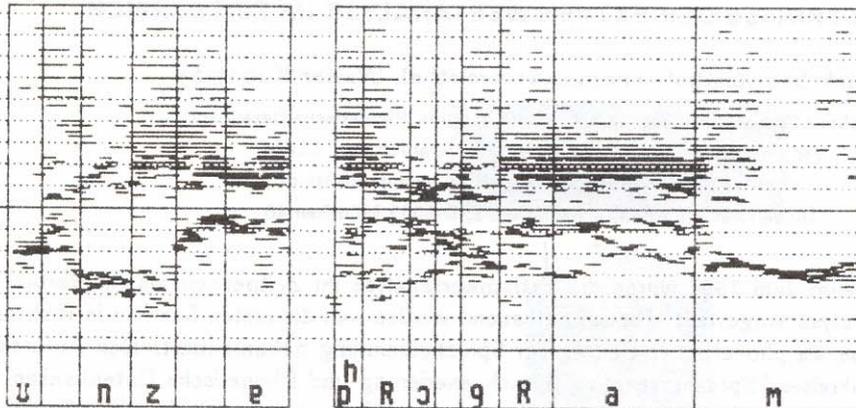
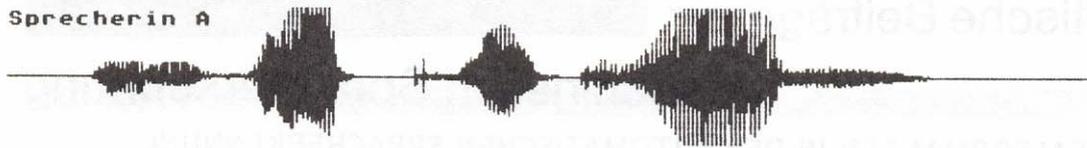
Beim Vergleich ist zu beachten, daß die Frequenzangaben in Sonagramm und Intervallogramm unterschiedliche Bedeutung haben: Das Sonagramm markiert die Schwingungsanteile (der einzelnen Frequenzen) des Sprachsignals, während das Intervallogramm die zeitliche Extremwertdichte ("Extremwertfrequenz") der Signalkurve bezeichnet.

Die Intervallogramme lassen Realformantverläufe erkennen, die sich mit Spektralformantverläufen im Sonagramm vergleichen lassen. Ausgehend von den Realformantverläufen haben wir die Intervallogramme segmentiert und den einzelnen Segmenten phonetische Transkriptionszeichen zugeordnet.

Inwiefern sich Hinweise auf eine hinreichende Definition einzelner Laute aus diesem Verfahren gewinnen lassen, muß sich bei weiteren Untersuchungen erweisen.

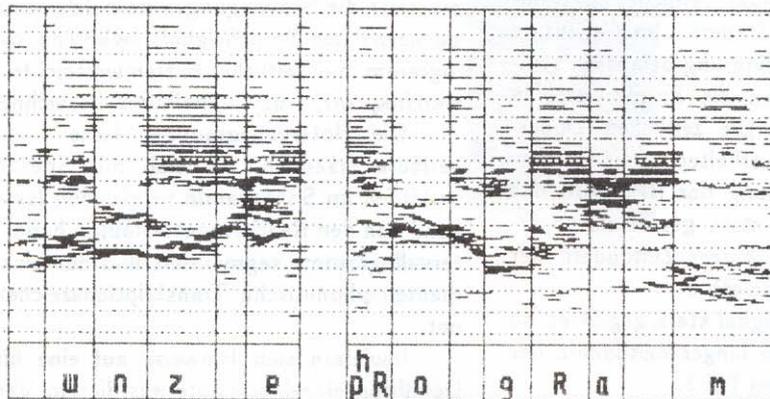
Zumindest eine objektive Transkription zeitlicher Segmente sollte bei geeigneten statistischen Verfahren möglich sein. Dies gilt auch für den stimmlosen Bereich (z.B. Frikative, Flüstersprache).

Sprecherin A



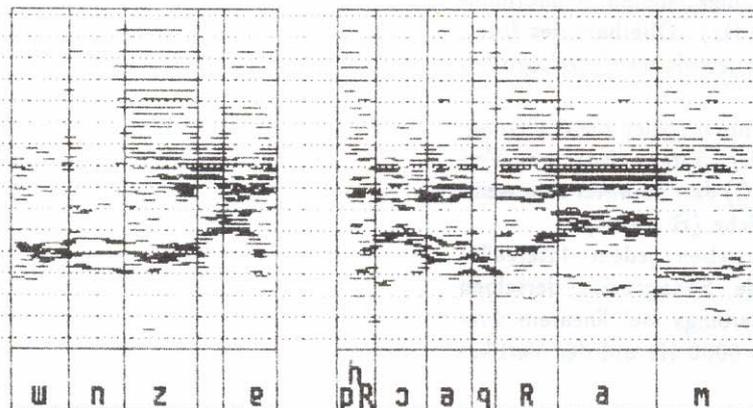
- 21434 Hz
- 15752 Hz
- 11576 Hz
- 8507 Hz
- 6252 Hz
- 4594 Hz
- 3376 Hz
- 2481 Hz
- 1823 Hz
- 1340 Hz
- 925 Hz
- 724 Hz
- 532 Hz
- 391 Hz
- 287 Hz
- 211 Hz

Sprecher B

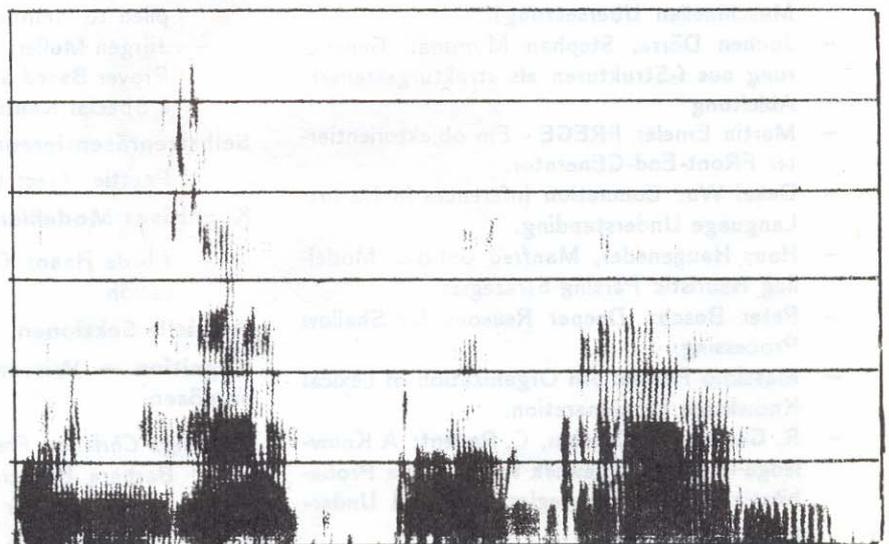
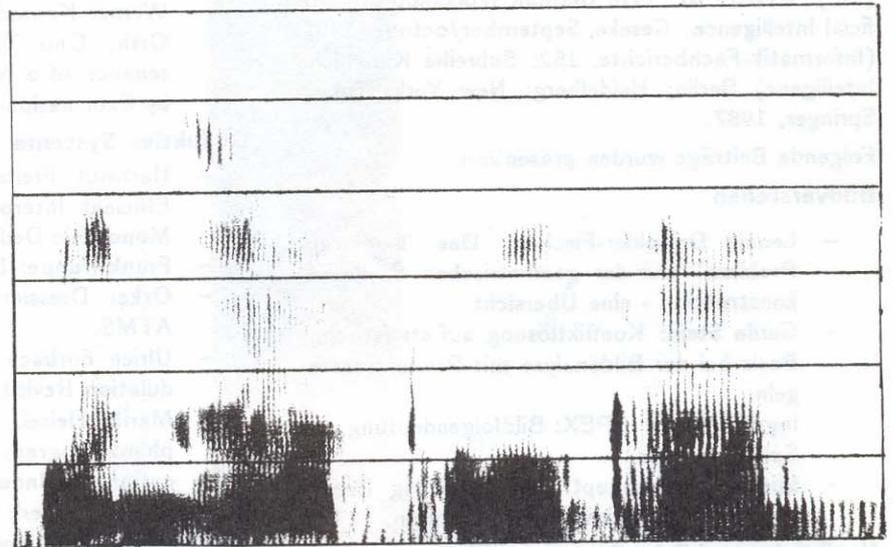
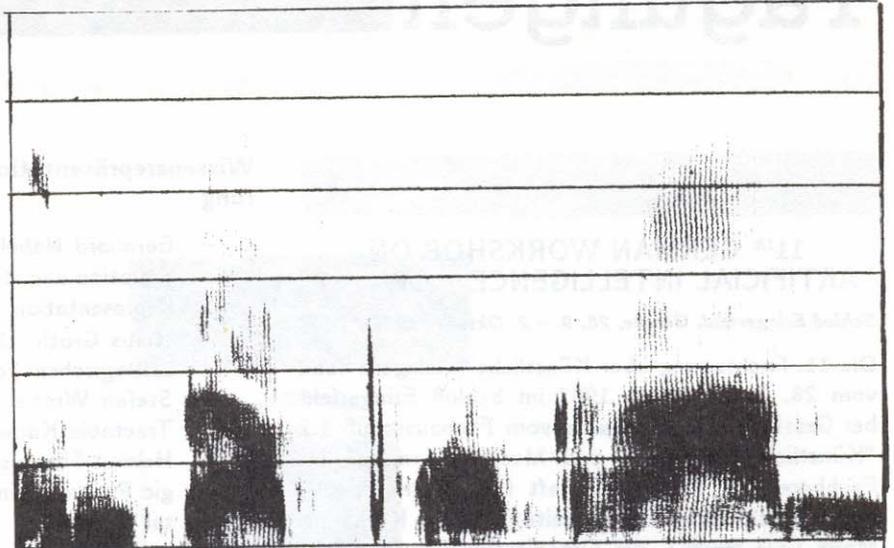


- 21434 Hz
- 15752 Hz
- 11576 Hz
- 8507 Hz
- 6252 Hz
- 4594 Hz
- 3376 Hz
- 2481 Hz
- 1823 Hz
- 1340 Hz
- 925 Hz
- 724 Hz
- 532 Hz
- 391 Hz
- 287 Hz
- 211 Hz

Sprecher C



- 21434 Hz
- 15752 Hz
- 11576 Hz
- 8507 Hz
- 6252 Hz
- 4594 Hz
- 3376 Hz
- 2481 Hz
- 1823 Hz
- 1340 Hz
- 925 Hz
- 724 Hz
- 532 Hz
- 391 Hz
- 287 Hz
- 211 Hz



Tagungen

11th GERMAN WORKSHOP ON ARTIFICIAL INTELLIGENCE – GWAI'87

Schloß Eringerfeld, Geseke, 28. 9. – 2. Oktober 1987

Die 11. Fachtagung über Künstliche Intelligenz fand vom 28. 9. bis 2. 10. 1987 im Schloß Eringerfeld bei Geseke statt. Sie wurde vom Fachausschuß 1.2 "Künstliche Intelligenz und Mustererkennung" im Fachbereich 1 der Gesellschaft für Informatik e.V. veranstaltet. Die Tagungsleitung hatte Katharina Morik (TU Berlin), die Organisationsleitung Dimitris Karagiannis (TU Berlin). Der Tagungsband ist seit Tagungsbeginn erhältlich: Katharina Morik (Ed.): GWAI - 87. 11th German Workshop on Artificial Intelligence. Geseke, September/october 1987. (Informatik-Fachberichte, 152: Subreihe Künstliche Intelligenz) Berlin; Heidelberg; New York; Tokyo: Springer, 1987.

Folgende Beiträge wurden präsentiert:

Bildverstehen

- Leonie Dreschler-Fischer: Das "Bootstrap-Problem" bei der geometrischen Szenenrekonstruktion - eine Übersicht.
- Gerda Stein: Konfliktlösung auf statistischer Basis bei der Bildanalyse mit Produktionsregeln.
- Ingrid Walter: EPEX: Bildfolgendeutung auf Episodenebene.
- Michael Mohnhaupt: On Modelling Events with an "Analogical" Representation.

Natürlichsprachliche Systeme und die Verarbeitung gesprochener Sprache

- Christa Hauenschild: KI-Methoden in der Maschinellen Übersetzung?
- Jochen Dörre, Stephan Momma: Generierung aus f-Strukturen als strukturgesteuerte Ableitung
- Martin Emele: FREGE - Ein objektorientierter FRont-End-GENERator.
- Dekai Wu: Concretion Inferences in Natural Language Understanding.
- Hans Haugeneder, Manfred Gehrke: Modeling Heuristic Parsing Strategies.
- Peter Bosch: Deeper Reasons for Shallow Processing.
- Massimo Poesio: An Organization of Lexical Knowledge for Generation.
- R. Gemello, E. Giachin, C. Rullent: A Knowledge-Based Framework for Effective Probabilistic Control Strategies in Signal Understanding.

Wissensrepräsentation und KI - Programmierung

- Bernhard Nebel, Kai von Luck: Issues of Integration and Balancing in Hybrid Knowledge Representation Systems
- Klaus Groth: Der Aspekt der Zeitstruktur in zeitlogischen Formalisierungen der KI.
- Stefan Wrobel: Higher-order Concepts in a Tractable Knowledge Representation.
- Helmut Simonis, Mehmet Dinbas: Using Logic Programming for Fault Diagnosis in Digital Circuits.

Expertensysteme

- Michael Beetz: Specifying Meta-Level Architectures for Rule-Based Systems.
- Werner Konrad, Andreas Jaeschke, Helmuth Orth, Ono Tjandra: Guiding the Maintenance of a Model-Based Advisory System by Explanation-Based Learning.

Deduktive Systeme

- Hartmut Freitag, Michael Reinfrank: An Efficient Interpreter for a Rule-Based Non-Monotonic-Deduction System
- Frank Puppe: Belief Revision in Diagnosis.
- Oskar Dressler: Erweiterungen des Basic ATMS.
- Ulrich Furbach: Oldy but Goody - Paramodulation Revisited.
- Marita Heisel, Wolfgang Reif, Werner Stephan: Program Verification by Symbolic Execution and Induction.
- R. Ulf Schmerl: Resolution on Formula Trees.
- Rainer Manthey, Francois Bry: A Hyperresolution-based Proof Procedure and its Implementation in PROLOG.
- Alexander Herold: Narrowing Techniques Applied to Idempotent Unification.
- Jürgen Müller: THEOPOGLES - A Theorem Prover Based on First-order Polynomials and a Special Knuth-Bendix Procedure.

Selbstrepräsentierende Systeme

- Paettie Maes: Computational Reflection.

Kognitives Modellieren

- Ab de Haan: Cognitive Modelling and Education.

Spezielle Sektionen:

Kognition – Wissensstrukturen beim Aufgabenlösen

(Leitung: Christian Freksa)

- Barbara Becker: Wissen und Können: Anmerkungen zur Wissensrepräsentation beim Aufgabenlösen.

- Gerhard Strube: Repräsentationsformen beim menschlichen Problemlösen.
- Klaus Rehkämper: Mentale Bilder und Wegbedeutungen.

Generierung in natürlichsprachlichen Systemen (Leitung: Wolfgang Hoepfner)

- Dietmar Rösner: Generation of Content vs. Generation of Form: A Review of Recent Work in the SEMSYN Project.
- Norbert Reithinger: Ein erster Blick auf POPPEL: Wie wird was gesagt?
- Helmut Horacek: How to say WHAT - IT or SOMETHING.
- Elisabeth Andre1, Thomas Rist, Gerd Herzog: Generierung natürlichsprachlicher Äußerungen zur simultanen Beschreibung von zeitveränderlichen Szenen.

Repräsentationssysteme für Grammatik und Lexikon

(Leitung: Günther Görz)

- Dietrich Paulus: Endliche Automaten zur Verbflexion und ein spezielles deutsches Verblexikon.
- Ulrich Heid: Zur lexikalischen Wissensquelle des Generierungssystems SEMSYN.
- Stephan Busemann: Generierung mit GPSG.

Künstliche Intelligenz und Datenbanksysteme – Systemarchitektur und konzeptuelle Modellierung

(Leitung: Rudi Studer)

- Gio Wiederhold, Surajit Chaudhuri, Waqar Hasan, Michael G.Walker, Marianne Wisnlett: Architectural Concepts for Large Knowledge Bases.
- Heinrich Jasper: Interfacing PROLOG and External Data Management Systems: A Model.
- Theo Härder, Nelson Mattos, Bernhard Mitschnang: Abbildung von Frames auf neuere Datenmodelle.

Bericht über die spezielle Sektion I

Generierung in natürlichsprachlichen Systemen: Beziehungen zwischen der Erzeugung von Inhalt und Form.

Leitung: Wolfgang Hoepfner

Die Sektion war von Wolfgang Hoepfner als Workshop organisiert worden. Thema war die traditionelle Einteilung der Generierung in die zwei Teilaufgaben der Bestimmung des Inhalts von Sprachäußerungen und die Formulierung dieses Inhalts in natürlicher Sprache.

Die Beziehungen dieser Teilaufgaben sollten aus theoretischer und praktischer Sicht erörtert und in einem größeren Rahmen diskutiert werden. Die bei-

den Extremstandpunkte skizzierte Wolfgang Hoepfner folgendermaßen:

- "die Bestimmung des Inhalts von Sprachäußerungen und die Produktion der sprachlichen Form sind völlig getrennte Prozesse, zwischen denen eine Schnittstelle definiert werden kann."
- "Inhalt und Form von Sprachäußerungen bedingen sich gegenseitig. Der Erzeugungsprozeß für die Inhaltsauswahl antizipiert die sprachliche Realisierbarkeit, und die sprachliche Formulierung kann Modifikationen bzw. Ergänzungen des Inhalts notwendig machen."

Zur Diskussion stand, ob die Unterscheidung dieser beiden Teilaufgaben noch ausreicht, um Sprachgenerierung zu charakterisieren. Dabei wurde erörtert, welche Rolle Modularisierungsgesichtspunkte der Architektur natürlichsprachlicher Systeme in diesem Zusammenhang spielen, welche Interaktionen zwischen den Teilprozessen zu finden sind, welche weiteren Aspekte der Generierung zu unterscheiden und zu berücksichtigen sind, welche anderen Generierungsmodelle angewendet werden und welche Rolle etwa pragmatische Informationen spielen. (vgl. Wolfgang Hoepfner, S.306)

In vier Beiträgen wurde zu diesen Fragen Stellung genommen:

Helmut Horacek (Universität Hamburg) erinnerte daran, daß in bestehenden Systemen bereits sehr viele größere Teilaufgaben ausgeführt werden. Er zeigte,

- daß in bestehenden Systemen und deren Komponenten keine klaren Grenzen zwischen der Verarbeitung konzeptueller und sprachlicher Einheiten zu finden sind,
- daß in unterschiedlichen Phasen und Komponenten (auch ad hoc) mit einzelsprachspezifischen Informationen gearbeitet wird,
- daß Bedingungen der Trennung nicht untersucht sind.

Er stellte Bedingungen und Möglichkeiten einer besseren Trennung vor und machte Vorschläge, wie die architektonisch attraktive Grenze weiter spezifiziert werden könnte.

Helmut Horacek: *How to say WHAT - IT or SOMETHING? S.320-329*

Norbert Reithinger (Universität Saarbrücken) unterschied nach dem Gesichtspunkt des Datenflusses drei Klassen von Architekturen von Generierungssystemen:

- sequentielle Modelle (unidirektionaler Datenfluß "what" - > "how")
- integrierte Modelle (keine Trennung des "what" und "how")
- Modelle mit Interaktion (Rückwirkung des "how" auf "what", bidirektionale Kommunikation über eine begrenzte, fest definierte Schnittstelle).

Er erörterte Funktion, Vorteile und Nachteile dieser Unterscheidung.

Er stellte das System POPEL vor, das als Grundlage eines Generierungssystems eine Architektur für ein Modell mit Rückwirkungen bereitstellt. Grundlage ist ein simulierter Parallelprozessor, die Generierung soll als parallele Kaskade ablaufen.

(Eine Implementation der parallelen Kaskade wurde vorgeführt).

Norbert Reithinger: Ein erster Blick auf POPEL - Wie wird was gesagt? S.315-319

Dietmar Rösner (Universität Stuttgart / GMD F4 Darmstadt) erinnerte daran, daß es auch wichtig ist, zu entscheiden, was nicht gesagt werden soll und zeigte, wie die Frage des Themas in SEMSYN behandelt wurde. Er wies darauf hin, daß der Forschungsstand bestehender Systeme noch nicht weit genug entwickelt ist, um die Diskussion als abgeschlossen betrachten zu können, daß vielmehr weitere Ergebnisse erwartet werden können, wenn

- alle Stufen des Generierungsprozesses gleichmäßig entwickelt und in Verbindung gebracht werden,
- die Beschränkung auf stark strukturierte Diskursbereiche überwunden werden kann,
- multilinguale Systeme konzipiert werden.

Er gab auch zu bedenken, daß ein einfacher Austausch sprachlicher Module nicht hinreichend ist, weil nicht nur sprachliche, sondern auch kulturelle Konventionen die Textorganisation bestimmen.

(Vorgeführt wurden aus dem Projekt SEMSYN zwei Anwendungen des Systems SEMTEX: eine Version einer multilingualen Generierung (deutsch und englisch) von Zeitungstexten (SEMTEX) und eine Version einer multilingualen (deutsch und englisch) und multimedialen (Text und Graphik) Generierung aus einer Wissensbasis (GEOTEX)).

Dietmar Rösner: Generation of Content vs. Generation of Form: A review of Recent Work in the SEMSYN Project. S.307-314

Elisabeth Andre1 (Universität Saarbrücken) behandelte Fragestellungen, die in bisherigen Generierungssystemen nicht auftraten. Sie erläuterte, welche Anforderungen die Aufgabe der simultanen Beschreibung zeitveränderlicher Szenen stellt. In der Situation der Simultanbeschreibung muß nach bisher unberücksichtigten Kriterien entschieden werden, welche Information über die Szene in welcher Form mitgeteilt werden soll.

Sie stellte das System SOCCER vor, das eine simultane Beschreibung eines Fußballspiels erzeugt. Es leistet eine Integration der Erkennung zeitveränderlicher Szenen und der Sprachproduktion. Der Schwerpunkt der Entwicklung lag bisher in der Integration der Prozesse.

(Das System wurde vorgeführt. Die Bildererkennungskomponente wird von der FhG Karlsruhe bearbeitet. Die Visualisierung der Szenen geschieht in SOCCER parallel zur Beschreibung).

Elisabeth Andre1, Thomas Rist, Gerd Herzog: Generierung natürlichsprachlicher Äußerungen zur simultanen Beschreibung zeitveränderlicher Szenen. S. 330-338

Ein Ergebnis der Diskussion war:

- Die Unterscheidung des "what" und "how" hat Auswirkungen auf alle bestehenden Systeme.
- Sie bietet Software-architektonisch eine attraktive Schnittstelle.
- Weiter zu definieren sind: einzelsprach- und domänenunabhängige Komponenten, das Format von Übergangsregeln, Übergangsregeln, ..
- Die Frage wird bekanntlich längst nicht mehr als einzige diskutiert und sie wird in anderen Kontexten gesehen.
- Neue Aspekte ergeben sich vor allem durch die Anforderungen multilingualer und multimedialer Generierung.
- Das Problem ist völlig ungeklärt.
- Offen ist, ob
 - der Traum des einzelsprachunabhängigen "Textinhalts" weiter für Architekturkonzepte fruchtbar gemacht werden kann,
 - ein neuer Lösungsansatz entwickelt werden kann.

Abschließend bleibt zu bemerken, daß ich die Sektion für gelungen halte.

Sie ist von Wolfgang Hoepfner sehr gut vorbereitet und durchgeführt worden. Die Beiträge waren sehr wertvoll und aufschlußreich. Wer nicht dabei sein konnte, sollte mehr als nur einen Blick in den Tagungsband werfen.

Übrigens ist auch für die nächste GWAI wieder eine spezielle Sektion unter einer aktuellen Themstellung in Planung.

Katharina Morik und Dimitris Karagiannis bleibt für eine inhaltlich und organisatorisch gelungene Tagung zu danken.

Dr. Karin Haenelt, Institut für integrierte Publikations- und Informationssysteme, GMD F4, Darmstadt

Ausschnitte aus einem randvollen Programm:

50. JAHRESTAGUNG DER AMERICAN SOCIETY FOR INFORMATION SCIENCE (ASIS)

Boston, Mass, 4. - 8. Oktober 1987

0. Der Rahmen

Ein halbes Jahrhundert ihrer Existenz konnte die ASIS anlässlich der diesjährigen Jahrestagung feiern. Das Programmkomitee hatte dem U.S.-amerikanischen Präsidenten sogar eine Grußbotschaft abringen können. Ein, dem Anlaß angemess-

senes, weitgespanntes Tagungsthema gab es auch, betitelt:

Information: The Transformation of Society.

Unter diesem breiten Thema (wohl eher als Motto zu verstehen) wurde ein prall gefülltes Angebot von Veranstaltungen und Komitee-Sitzungen präsentiert, welches die Spannweite der "Informationsarbeit" wohl deutlicher beschreibt, als alle Versuche volun-tarischer Begriffsdefinitionen:

Von gesellschaftlichen Erörterungen des Ein-satzes von Informations- und Kommunikations-technologien, über das Methodenspektrum des Information-Retrievals und seine Bewertungsmaß-stäbe, von linguistischen und empirischen Ansätzen von Benutzer-System-Schnittstellen bis zur Situa-tionsschilderung in einzelnen Fachinformationsbe-reichen (z.B. Chemie oder Biologie), von elektro-nischen Publizieren und Filing/Retrieval-Aspekten in Bürokommunikationssystemen über den Komplex Wirtschaftsinformation bis hin zu informationsethi-schen und -rechtlichen Erwägungen.

Insgesamt wurden 75 Einzelveranstaltungen in diesen knapp 4 Tagen angeboten, auf denen schätzungsweise 300 Redner (bzw. Moderatoren oder Diskutanten) auftraten. Die 900 - 1.000 an-wesenden Konferenz-Teilnehmer waren aufgefordert, unter jeweils 4 - 5 Parallelveranstaltungen eine - bisweilen qualvolle - Wahl zu treffen. So auch der Berichterstatter, der sein Interesse auf neuere Ent-wicklungen bzw. Arbeiten aus dem Hochschulbereich konzentrierte.

1. Beiträge aus den Hochschulen

Abschlussarbeiten aus den ca. 40 Hochschulen der U.S.A., die im Großbereich Kommunikations-, Informations- und Bibliotheksausbildungsgebiet tätig sind, stellen die Substanz dieser Veranstaltungsrei-hen.

In der Sitzung "Doctoral Research Award", für welche das ASIS Research Committee die Schirm-herrschaft hatte, werden 5 Beiträge geboten, die aus-nahmelos aus den Hochschulen stammten.

Barbara Wildemuth, PhD Drexel-University, Philadelphia stellte mit ihrem Beitrag "End-User Computing: A Study of the Adoption of an Intellec-tual Technology in Corporate Settings" die Ergeb-nisse ihrer PhD-Arbeit vor. Interessant an dieser empirisch- abgesicherten Studie ist der Beitrag zu betrieblichen organisationellen Erfordernissen, von denen die Einpassung einer "intellektuellen" Tech-nologie stark abhängt. Erfordernisse des Endbenut-zers unter starker Berücksichtigung der Organisati-onsstruktur eines gegebenen Arbeitsplatzes stellen somit die wirklichen Orientierungspunkte, auf die hin Benutzer- Modellierung von End-Benutzer-Systemen ausgerichtet sein müssen. Dieser, in Tenor und Ab-sicht wohl nicht neue Ansatz bestach jedoch durch die präzisen und materialreichen Belege, die für ein solches Vorgehen angeführt wurden. Die Arbeit

wurde durch die ASIS ausgezeichnet. Sie stellt in gewisser Weise einen empirisch- konzeptuellen Bau-stein für Informations-Management dar.

Paul Nicholls, University of Western Ontario, stellte seine Arbeit "The Lotka Hypothesis and Bi-bliometric Methodology" die die Beziehungen von mathematisch-statistischen Modellen und den realen Zitierungsmustern in den verschiedensten Fachge-meinschaften heraus.

Über den Wert solcher nun schon recht betag-ten Untersuchungen läßt sich wohl trefflich strei-ten, jedoch bestechen - so wie auch in dieser Ar-beit - die gründlichen statistischen Kenntnisse und die Differenziertheit statistischer Untersuchungsver-fahren. Als dritter Beitrag im "Doctoral-Forum" bleibt die Arbeit von Mark E. Rorvig, University of California, Berkeley, zu erwähnen, der über ein Experiment über die Auswahl von Dokumenten in einer simulierten Informationsumgebung berichtete. "An Experiment in Human Preferences for Doku-ments in a Simulated Information System". Be-merkenswert hierbei war die Anordnung des Expe-rimentes in der sorgsam angelegten Simulations-umgebung; deutet sie doch an, daß im empirisch-experimentellen Bereich, der gegenwärtige Trend zu sehr tiefen und sehr benutzernahen Fragestellungen Vorrang vor dem flächendeckenden, aber empirisch-oberflächlichen Ansätzen gewinnt (über solche, teil-weise sehr aufwendigen und tiefen Analysemethoden hört man in letzter Zeit kaum noch etwas in unseren Landen).

In der Abteilung der studentischen Arbeiten wurden zwei Beiträge vorgestellt: "End User Se-arching as an Innovation: Measurement of Percei-ved Attributes and Task Complexity" von Joanne Marshall, University of Toronto und "Directness and Indirectness of Womens' Negotiations in Organizati-ons".

Die beiden Arbeiten wurden nach Schwerpunk-ten auf die Untersuchungsmethoden gelegt, im er-sten Falle auf die Definition von Attributen und Auf-gabenstellungen in Beziehung zur Suchstrategie, im zweiten Fall auf die Darstellung der konversations-analytischen Methode für Zwecke der menschlichen Beratungssituation. Auch hier wurde der Eindruck untermauert, daß das qualitative, empirische Metho-denspektrum auf Informationssituationen angewen-det wird - wohl der Einschätzung folgend, daß die quantitative Methoden der vergangenen zwei Jahr-zehnte recht wenig von den kognitiven Prozessen der "Informationsverarbeitung" beim Menschen erhellen konnte.

2. Wissensrepräsentation

Die Veranstaltung: "Structures of knowledge Repr-ésentation" wurde von Robert Odaly, Syracuse Uni-versity mit dem Beitrag "On the Validation and In-terpretation of Representation of Anomalous States of Knowledge in Information Retrieval" eröffnet.

Hierbei ging es um die Visualisierung von Begriffsstrukturen und deren Vernetzung in Situationen des Information-Retrievals. Der Ansatz des "Anomalous State of Knowledge" ist bereits in den 70er Jahren von N. Belkin, Rutgers University, entwickelt worden und versucht aus dem anormalen Zustand des Wissens einen Bezugsrahmen zu entwickeln, wie die Begriffs- und Lernstruktur in der Informationssuche nun ablaufen kann. Dieser Neu-Aufguß einer Reihe nun schon älterer Arbeiten der Autoren Oddy/Belkin konnte nichts Neues vorweisen. Daß Begriffs- und Vernetzungsstrukturen kreisförmig, ringförmig oder sternförmig angeordnet sein können, wußte man schon vorher. Die Fragen, wie es möglicherweise zu diesen Figuren kommen kann, ob diese Begriffsmuster typenabhängig, oder prozessabhängig entstehen, konnte nicht auch nur annähernd erklärt werden.

Vom Ansatz der Fragestellung her wesentlich enger war der Beitrag von Susan Bozi, Syracuse University "Syntactic Patterns in Sublanguages: Sciences and Social Sciences".

Mit viel statistischem Aufwand, der bei der Fragestellung vielleicht gar nicht angemessen war, wurden hier größere Mengen an Fachtexten der Natur- und Sozialwissenschaften hinsichtlich ihrer Syntax vergleichend gegenübergestellt. Abgesehen von der Frage des Erkenntniswertes solcher Untersuchungen, blieb als bedeutender Mangel der Eindruck haften, daß unter syntaktischer Analyse lediglich das Zählen von bestimmten Wortformen verstanden wurde: ein sehr oberflächliches Verständnis des linguistischen Interpretationsrahmens, der an Textsorten anzulegen ist. Ähnlich arm im linguistischen Verständnis von Texten war der Beitrag von Elisabeth Liddy, Syracuse University: "Discourse Level Structure in Abstracts".

Mit dem einfachen Anzählen von z.B. Präsens oder Perfektformen, wie sie präsentiert wurden, fallen solche zeitlich sehr aufwendigen Analysen hinter dem Stand der Textlinguistik zurück. Es ist zu wünschen, daß zu dem neueren - ungleich reicheren Analysemöglichkeiten eines modernen Verständnisses von Fachsprachen aufgeschlossen wird.

In der "Great Vocabulary Debate" standen sich Davis Batty als engagierter Vertreter für kontrolliertes Vokabular (im Sinne des Thesaurus) und Bruce Croft, University of Massachusetts als Verfechter des automatischen Indexierens gegenüber. Unterstützt wurden beide Disputanten durch die längeren Ausführungen des Alt-Meisters G. Salton, der es immerhin verstand, die "klassischen" Bewertungsmaße (Recall Precision) in Abhängigkeit von der Untersuchungsmethode zu relativieren. Die Positionen der Kontrahenten waren klar: Einerseits ein vom Text unabhängiges (durch geistige Tätigkeit entstandenes) kontrolliertes Vokabular, welches Zuordnungen in größerer Abstraktion vom Basistext vornimmt. Andererseits die Position, daß mit den

4 Hauptmethoden der linguistischen Reduktion von Fachtexten mindestens gleich gute Ergebnisse zu erzielen seien.

Diese Grundpositionen wurden nicht aufgegeben; jedoch fand insofern eine Annäherung statt, daß für die Erstellung von kontrollierten Wortschätzen der Einsatz von computerlinguistischen Verfahren sinnvoll und nützlich ist. Diese Diskussion faßte lediglich die Argumente der letzten beiden Jahrzehnte zusammen. Meines Erachtens wurde es versäumt, auf die künftige Perspektiven des Information-Retrieval einzugehen.

3. Eindrücke - zusammengefaßt

Auffällig - selbst bei dem geringen Ausschnitt über den hier nur berichtet werden konnte - ist der Trend zur differenzierten empirischen Analyse in gegebenen technologischen Systemumgebungen. Dabei werden von der soziologischen und psychologischen Methodologie her verstärkt qualitative und sehr differenzierte Analysemethoden angewendet, von denen man sich Aufschlüsse auf den großen "Kognitiven Unbekannten" verspricht. Auch wenn die hier vorgestellten Arbeiten mit Bezug auf die Analyse von Fachtexten vielleicht nicht das ganze Spektrum der Arbeiten repräsentieren, so ist doch auffällig, daß linguistische Ansätze kaum den modernen Stand aufweisen. Über Technologie als solche (d.h. außerhalb der gesellschaftspolitischen Dimensionen) wird nicht mehr in dem Maße gesprochen, wie man es vielleicht erwartet hatte. Technologie ist offenbar reichlich und in modernster Form vorhanden (übrigens auch in den Ausbildungsinstitutionen, die der Berichterstatter Gelegenheit hatte zu besuchen): mit ihr etwas Professionelles anzufangen, sie spezifisch auf die Erfordernisse der Informationsarbeit zu nutzen und sie auf die Bedürfnisse der Nutzer hin auszugestalten scheint die gegenwärtige Priorität auszumachen.

Der Standard der Hochschularbeiten scheint - abgesehen von der Zahl der Arbeiten - mit den (vergleichsweise wenigen) Arbeiten bei uns zumindest vergleichbar.

Prof. Dr. Thomas Seeger, Fachhochschule Darmstadt, Fachbereich IuD, Schöffersstraße 8, D-6100 Darmstadt

ZUGANG ZU WERKSTOFF-FAKTENBANKEN

ANFORDERUNGEN UND KONZEPTIONEN – STAND UND PERSPEKTIVEN

Technische Hochschule Darmstadt, 21. bis 23. September 1987

Etwa 50 Teilnehmer aus Industrie, Informationspraxis und Hochschule, aus den Werkstoffwissenschaften, dem Konstruktionsbereich, der Informatik und den Informationswissenschaften diskutierten vom 21. bis 23. September 1987 an der TH Darmstadt über den "Zugang zu Werkstoff-Faktenbanken – Anforderungen und Konzeptionen, Stand und Perspektiven". Veranstalter dieses Workshops war das vom Bundeswirtschaftsministerium geförderte Projekt WeBeS, das an der Technischen Hochschule Darmstadt in Kooperation mit dem Fachbereich Information und Dokumentation der Fachhochschule Darmstadt durchgeführt wurde. Der Veranstalter hatte eine Reihe kompetenter Referenten eingeladen und stellte selbst die Arbeitsergebnisse und Schlußfolgerungen aus seiner einjährigen Arbeit vor. Die Vorträge waren in ein Zeitraster von 45 Minuten eingepaßt, das sich tatsächlich als etwas eng herausstellte. Die Diskussion war nämlich während des gesamten Workshops überaus rege und produktiv und trug damit entscheidend zu dem einhelligen positiven Eindruck der Teilnehmer bei. Vorführungen gab es für die Datenbanken STEELFACTS und POLYMAT, für PERITUS, ein Expertensystem zur Werkstoff-Auswahl und zu einem in WeBeS entwickelten Smalltalk-"Werkstoff-Browser".

Der Workshop wurde am Montag, 14⁰⁰ Uhr von Prof. G. Knorz, dem Projektleiter von WeBeS, mit einer Kurzvorstellung des einladenden Projektes eröffnet. Von zwei Beobachtungen ausgehend, nämlich daß "Werkstoffdatenbanken nicht leicht an den Kunden zu bringen sind" und daß es an einem "kompetenten, konstruktiven und zukunftsorientierten informationswissenschaftlichen Ansatz fehlt" charakterisierte er zunächst die Situation, in der es zur Beantragung und Förderung des WeBeS-Projektes kam. Auf eine Kurzformel gebracht sollte dieses Projekt auf der Basis empirisch gestützter Konzeptionen für problem- und benutzerangepaßte Werkstoffdatenbanken Möglichkeiten und Grenzen solcher Systeme untersuchen, die gegenwärtig verfügbar sind ("Stand der Technik"), entwickelt ("angewandte FuE") oder diskutiert werden ("Grundlagenforschung"). Der Workshop als ein wichtiger Bestandteil des Projektes zielte darauf, für die gewonnenen Einschätzungen eine breit diskutierte Grundlage zu finden, um daraus für die weitere Entwicklung Konsequenzen und Prioritätensetzungen ableiten zu können. Das Projekt WeBeS und auch der Workshop ordnet sich damit in eine von mehreren parallelen Maßnahmen ein,

die Situation im Bereich Werkstoff-Fachinformation mittel- und langfristig zu verbessern.

Die sich anschließenden Vorträge des ersten Nachmittagsblockes beleuchteten beide den Status quo von Entwicklung und Bedarf an Werkstoff-Faktenbanken (WFB). Zunächst stellte Dipl.-Ing. W. Queren vom Fachinformationszentrum Werkstoffe, Berlin den "Stand der Werkstoff-Datenbank-Entwicklung" dar. Er gab für den Werkstoffsektor einen Überblick über

- Historie und Stand der einschlägigen nationalen und internationalen Aktivitäten (frühe Studien (1970) in der BRD, Aktivitäten in den USA, CODATA-Conferenzen 84, 85, 88, ASTM/E49-Conferenz 87, EG-Programm Demonstrator 84-88, Werkstoff-Datenbanken im FI-Programm),
- gegenwärtig auf dem Markt angebotene Online-Datenbanken sowie
- PC-Datenbank-Systeme.

Die Bedeutung der PC-Datenbanksysteme, auf schmale Sektoren zielend oder als Marketing-Maßnahme für die Werkstoffe eines Herstellers, sollte nicht unterschätzt werden. Als besonders wichtige Maßnahme stellte er die Entwicklung einer einheitlichen Terminologie zur Kategorisierung von Werkstoffdaten heraus, die die Voraussetzung für die Portabilität von WDB's ist. Damit war man bereits mitten in der Diskussion, wie eine solche Standardisierung praktisch erarbeitet und durchgesetzt werden könne.

Dipl.-Ing. E. Steinsiek (GEWIPLAN) trug über ausgewählte Ergebnisse einer empirischen Untersuchung der Firma GEWIPLAN vor: "Bedarfs- und Anforderungsanalyse für Faktenbanken". Die Aussage, daß ein Grundbedarf von ca. 180.000 potentiellen Recherchen an Werkstoffdatenbanken pro Jahr nach mechanischen und chemischen Kennwerten, nach Werten über Korrosion, Reibung und Verschleiß, nach Konstruktionshinweisen und anderen Informationen besteht, stützte sich auf Angaben von ca. 1200 Fachleuten aus Industrie, Wissenschaft und Behörden. Die Untersuchung gab über zahlreiche Aspekte der auftretenden Bedarfssituationen, der Art der benötigten Informationen und der Anforderungen an das Informationsangebot detaillierte quantitativ belegte Auskunft. Die zahlreichen Diskussionsbeiträge dokumentierten eindrucksvoll das Interesse der Teilnehmer, die sich von dem umfangreichen Zahlenwerk nicht "erschlagen" ließen. Insbesondere wurde Kompetenz der 1200 Fachleute in Bezug auf eine Technik hinterfragt, die offensichtlich, und das belegte die GEWIPLAN-Studie zweifelsfrei, bisher kaum bekannt und faktisch nicht angewandt wird.

Der zweite Nachmittagsblock wurde von Dr. Behrens, FIZ Karlsruhe moderiert. In seinem Vortrag "Anforderungen an Werkstoff-Datenbanken aus der Sicht eines Industriebetriebes" wollte Dr.-Ing.

W. Betz (Motoren- und Turbinen-Union München) erklärtermaßen mehr Fragen als Antworten aufwerfen. An den Anfang stellte er eine Grobgliederung der Werkstoffdaten in Verwaltungsdaten (z.B. Bestellnummer), Herstellparameter (z.B. chemische Zusammensetzung), Versuchsparameter (z.B. Probenlage) und Ergebnisse (z.B. E-Modul) als die in der industriellen Praxis nachgefragten Daten für Werkstoff-Entwicklung und -Auswahl, Konstruktion, Qualitätssicherung etc. Er konkretisierte wichtige Anforderungen an die Speicherung und den Zugriff von Werkstoffdaten, wie sie auch im weiteren Verlauf des Workshops immer wieder genannt wurden (Effizienz bei Massendaten, gelenkter Zugriff auf Daten-Ausschnitte, flexible und vollständige Datendefinition, u.a.m.). Darauf aufbauend entwickelte er die Skizze eines Datenmodells für die Speicherung von Werkstoff-Fakten, wie es bei MTU als Grundlage eigener Entwicklungen z.Z. diskutiert wird: Eine flexible Material-Daten-Matrix von Grundelementen, die ihrerseits aus Text bzw. Kennzahlen oder aus Maßzahlen mit Typ-, Dimensions- und Genauigkeitsangaben bestehen. Die vorab geäußerte Hoffnung einiger Workshopteilnehmer, daß die Industrie in aller Stille den Stein der Weisen längst gefunden oder gar geschliffen hätte, mußte – das stellte Betz klar heraus – begraben werden.

Den Abschluß des ersten Tages bildete ein sehr anschaulicher Vortrag von Dr.-Ing. K.-H. Beelich (TH Darmstadt, Fachbereich Maschinenbau), der über "Werkstoffauswahl bei den Arbeitsschritten des methodischen Konstruierens" sprach. Er machte deutlich, daß die Selektion eines Materials in einen komplexen, aber (wenn man der Konstruktionslehre nach Pahl/Beitz folgt) gut strukturierten Konstruktionsprozess eingebettet ist und daß bessere Möglichkeiten der Werkstoffauswahl stets in Konkurrenz zu geeigneten Konstruktionsmaßnahmen stehen. Mit eigenen Worten und überspitzt formuliert: zu hohe Anforderungen an einen Werkstoff deuten auf nicht-werkstoffgerechtes Konstruieren, also auf Konstruktionsmängel hin. Selbst wenn also Werkstoffdatenbanken keine besonderen Probleme aufwerfen würden, das ist die eine Interpretation von Beelichs Botschaft, so sollte man sich über den Stellenwert der von einer solchen Datenbank unterstützten Dienstleistungen keine Illusionen machen. Eine mehr optimistische und konstruktive Interpretation wurde von Beelich und der WeBeS-Projektgruppe formuliert: Vielleicht sind es vielfach nicht die universellen Datenbanken, die der Benutzer braucht, sondern Systeme, die umfassendere Arbeitszusammenhänge (wie etwa das Konstruieren) unterstützen und in die das Werkstoff-Auswahlproblem geeignet eingebettet ist (Das PERITUS-System, Vorstellung am Mittwoch, scheint ein erfolgreicher erster Schritt in diese Richtung zu sein.) Das methodische Konstruieren könnte dann die theoretische und in Grundzügen auch die praktische Basis zur Erstellung einer Wissensbasis liefern, die den Kontext

und die "Logik" des Auswahlprozesses für spezialisierte Anwendungen beschreibt.

Mit einem Vortrag über "Effektive Werkstoffauswahl mit der Kuststoff-Datenbank POLYMAT" von Frau Dipl.-Chem. J. Wierer vom Deutschen Kunststoffinstitut, Darmstadt wurde der von Dr.-Ing. N. Fuhr (WeBeS-Projekt, TH Darmstadt) moderierte erste Vormittagsblock am Dienstagmorgen begonnen. POLYMAT ist eine vom DKI hergestellte und vom FIZ Chemie online angebotene Datenbank (implementiert mit ADABAS/NATURAL), die 1986 in einer ersten Ausbaustufe mit 2000 Werkstoffen freigegeben wurde. Sie enthält mittlerweile etwa 3000 Werkstoffe und zwar hauptsächlich Thermoplaste. Pro Werkstoff sind aus einem Spektrum von insgesamt 200 Eigenschaften ca. 30 bis 50 Angaben gespeichert. Im Endausbau sollen etwa doppelt so viele Werkstoffe und Eigenschaften recherchierbar sein. Die Daten selbst stammen aus öffentlich zugänglichen Publikationen. Es wurde sehr klar herausgearbeitet, wie sich die Vorstellungen über Schwierigkeiten und Aufwand beim Aufbau einer Werkstoff-Datenbank erst mit der tatsächlichen Implementierungsarbeit konkretisiert hatten und welchen Aufwand allein etwa der laufende Systeminput erfordert. "Da hätten wir recht konkret vorwarnen können", so der Kommentar eines Workshop-Teilnehmers mit einschlägiger und leidvoller eigener Erfahrung. Nicht zuletzt das BMWi müsse hier die in der Vergangenheit zugrundegelegten unrealistischen Kostenannahmen und die damit zu hoch gesteckten Erwartungen korrigieren.

"Faktenbank ist nicht gleich Faktenbank" war eine der Erkenntnisse aus dem sich anschließenden Vortrag "Modellierung bei Faktenbanken" von Dr. J.L. Staudt, Universität Konstanz. Zunächst wurde die im IuD-Bereich weithin anzutreffende Fehleinschätzung korrigiert, daß Faktendatenbanken neben den Literaturdatenbanken eher ein Schattendasein führten. Im Gegenteil – bei der Mehrzahl der gegenwärtig öffentlich angebotenen 3000 Datenbanken handelt es sich um Faktenbanken! Richtig ist allerdings, daß Entwickler und Anwender in beiden Bereichen aus historischen Gründen voneinander bisher wenig wissen. Am Beispiel von konkreten (und tatsächlich auch häufig benutzten) statistischen Datenbanken erläuterte er das Datenmodell des mehrdimensionalen Merkmalsraumes und die Art und Weise, mit der man mittels recht kryptographischer Retrievalsystematiken aus dem gespeicherten Gesamttraum denjenigen "Teilraum" herauschneidet, der gerade interessiert. Da es sich hierbei in der Regel um Zeitreihen handelt, sind nachfolgende Operationen zur Auswertung, Veranschaulichung oder tabellarischen Aufbereitung üblicherweise in die Retrievalsysteme integriert. Besonders ging Staudt auf die Bedeutung der Modellierungsphase für einen Datenbankentwurf ein und auf die Tatsache, daß – wie man bereits nach kurzer Beschäftigung mit die-

sem Thema erkennt – Werkstoff–Fakten tatsächlich ein ganzes Bündel zusätzlicher und schwieriger Probleme aufwerfen. Auch die Anwendungssituation stellt harte Anforderungen: Die Argumentation in einem Gutachten etwa mag sich bei fehlenden Daten in einer Statistik–Datenbank der vorhandenen Datenlage erfolgreich anpassen lassen, dem Konstrukteur ist mit einigen zusätzlichen Angaben zur Kurzzeitbelastung nicht gedient wenn er eigentlich Aussagen über Dauerbelastung benötigt.

Die zweite Hälfte des Vormittags wurde von der *WeBeS–Gruppe* bestritten: "Überblick über das Projekt *WeBeS*" von G. Knorz, "Der empirische Ansatz von *WeBeS*" von Dipl.–Psych. K. Ammersbach und "Anforderungen an Werkstoff–Datenbanken aus der Sicht von *WeBeS*" von Dr.–Ing. N. Fuhr. Diskussionsleiter war Prof. J. Krause, Universität Regensburg. An einem Beispiel wurde zunächst systematisch verdeutlicht, von welcher Art die konzeptionellen Hürden sind, die ein Informationssuchender bis zum (hoffentlich) befriedigenden Rechercheergebnis überwinden muß: Das Werkstoff–Problem muß in einem ersten Schritt in formale Anforderungen an Materialeigenschaften übersetzt werden, wobei neben prinzipiellen Problemen der Formalisierbarkeit in diesem Bereich insbesondere Wissen über Werkstoffe, Prüfmethoden, Einflußgrößen, Eigenschaften, Verarbeitungsmethoden usw. notwendig sind. Im einem zweiten Schritt muß die Anpassung an den Wissensausschnitt der Datenbank erfolgen. Dazu genügt es nicht, die von der Datenbank angebotenen Kategorien zu kennen. Vielfach sind die nachgefragten Werte oft nicht bekannt, oder die Werte sind in anderer Form gespeichert als erwartet. Eine wesentliche Schwäche gegenwärtig verfügbarer Faktenbanken ist es, daß sie diese typische Situation nicht berücksichtigen, den Benutzer hier nicht unterstützen und ihn in sehr vielen Fällen regelrecht in die Irre leiten. Als dritter Schritt steht dann noch die Anpassung an die ganz speziellen Möglichkeiten und Besonderheiten des Datenbanksystems aus. Wenn existierende Werkstoff–Faktenbanken bisher weitgehend mangelhaft zu beurteilen sind, so liegt das tatsächlich entscheidend an der Komplexität des Gegenstandsbereichs: Inhomogenität, Unvollständigkeit und Abhängigkeiten in den Daten sowie die Unschärfe und die spezielle Einbettung des Suchproblems stellen Anforderungen, die konventionelle Datenbanksysteme bisher nicht abdecken können. Schon die Informationsstruktur eines konventionellen Werkstoffblatts, so lückenhaft und unbefriedigend dieses sein mag, läßt sich nicht so in ein Datenbanksystem abbilden, daß die gespeicherten Werte zuverlässig erfragt werden können. Ziel von *WeBeS* war es nun, Konzeptionen für benutzer– und problemangepaßte Werkstoffdatenbanken zu entwickeln und dies durch empirische Untersuchungen zu stützen. Das Untersuchungsdesign ging von der (Video–)Beobachtung re-

aler problemorientierter Recherchen an Werkstoffdatenbanken aus. Aufgezeichnet wurden die Interaktionen zwischen Industriekunden und professionellen Rechercheuren als Informationsvermittler. In Vor– und Nachinterview schloß die Datenerhebung bei insgesamt 15 Fällen ab. Das globale Ergebnis, keine einzige zufriedenstellende Recherche an einer Werkstoffdatenbank, löste dann doch in dieser Eindeutigkeit Betroffenheit aus. Detailergebnisse und der *WeBeS*–Systementwurf wurden im weiteren Verlauf des Workshops vorgestellt.

Dipl.–Ing. A. Brandstätter, GID (Projektträger) leitete dann die erste Nachmittagssitzung, die Dr.–Ing. G. Dathe (Betriebsforschungsinstitut, Düsseldorf) mit einem "Beitrag zu einem konzeptionellen Schema für Werkstoff–Informationssysteme" eröffnete. "Information statt Konfrontation" war das Motto, das er seinem Vortrag voranstellte und sein konkreter Vorschlag zur Herstellung einer gemeinsamen Verständigungsbasis war das Entity–Relationship–Modell (ERM). Indem er Werkstoffe und Erzeugnisse in schrittweiser Vervollständigung mit den Ausdrucksmitteln des ERM modellierte, stellte er dieses Werkzeug zur Datendefinition anschaulich vor und demonstrierte an einem gut ausgearbeiteten Beispiel eindrucksvoll den Gewinn bei einer solchen Vorgehensweise: konzeptionelle Klarheit und Übersicht, und vor allen Dingen die Möglichkeit einer präzisen Kritik und konstruktiven Diskussion. Nicht so sehr die Wahl gerade des ERM, sondern prinzipiell die Benutzung einer geeigneten formalen Definitionssprache auf dem Weg zur Festlegung einer umfassenden Modellierung von Werkstoffdaten war der entscheidende Punkt. Aus diesem Rahmenschema könnten dann speziellere konzeptionelle Schemata abgeleitet werden. Die parallel dazu laufenden Aktivitäten zur Vereinheitlichung der Terminologie und des Kategorienschemas, über die *Queren* berichtet hatte, wären hier einzubringen. Fernziel könnte ein Standard sein, der vergleichbar etwa zu ODA oder SGML in der Bürowelt, die Austauschbarkeit der Daten und damit die weitgehende Kompatibilität auch unterschiedlicher Werkstoffdatenbanken ermöglicht. Dieser Punkt und vor allen Dingen die praktischen Voraussetzungen und Konsequenzen einer solchen Vorgehensweise, etwa für die Arbeit eines entsprechenden Kreises von Fachleuten wurden anschließend diskutiert.

Der Beitrag "Funktionale Anforderungen an Werkstoff–Datenbanken" von Dipl.–Phys. F. Mie, GID Heidelberg, war zunächst eine engagiert vortragene Schelte an die Konzeption und Realisierung gängiger Informationssysteme, die völlig am eigentlichen Benutzer vorbeibetrieben wurden. Er forderte dazu auf, sich z.B. das Mißverhältnis zwischen der Bedeutung der Benutzeroberfläche für den Anwender und dem relativen Anteil ihrer Entwicklung am Gesamtaufwand zu vergegenwärtigen. Die Tatsache, daß in einem Projekt wie *WeBeS* eine

Diplom-Psychologin arbeitet und daß dort Benutzerforschung tatsächlich betrieben werde, bestätigte ihn allerdings in seiner Einschätzung und Hoffnung, daß mittlerweile ein Umdenken im Gange ist. Die verschiedenen Anforderungen an Werkstoff-Faktenbanken, die Mie aus dem Bereich der Softwareergonomie und aus seinen eigenen umfangreichen Erfahrungen systematisch zusammengestellt hatte, stießen auf großes Interesse.

Der letzte Vortragsblock am Dienstag, moderiert von G. Knorz, wurde von N. Fuhr eröffnet. In seinem Beitrag "Systemkonzeption für zukünftige Werkstoff-Informationssysteme" stellte er ein Schalenmodell, bestehend aus einer Werkstoff-Datenbank (einschließlich allen datenbankspezifischen Wissens) als Kern, einer ersten Schicht mit einer Repräsentation von datenbankunabhängigem Werkstoff-Wissen, einer zweiten Schicht mit speziellen Retrievalfunktionen und Methoden und einer dritten Schicht mit Anwendungs- und Benutzerklassenspezifischen Funktionen. Als Lösungsansätze wurden das NF²-Modell als Erweiterung des relationalen Datenmodells, abstrakte Datentypen, ein "intelligentes" Data Dictionary, Verfahren zur Behandlung von Unsicherheit und von fehlenden Werten sowie Konzepte für graphische Schnittstellen diskutiert.

Daß es mindestens ein überzeugendes Beispiel für ein problem- und benutzerangepaßtes System zur Selektion von Werkstoffen bereits gibt, machte Dr. N. Swindells (Matsel Systems Ltd., Liverpool) in seinem Beitrag "The User Interface to the Peritus Knowledge Base" deutlich. Es handelt sich dabei nicht um ein Zugangssystem zu einer universellen Werkstoff-Faktenbank, sondern um ein System, das auf der durchdachten Konzeption einer Werkstoff-Auswahl in mehreren Schritten besteht. Nach Auswahl der gewünschten Material-Klasse (Kunststoffe, Stahl,...) und nach Eingabe eines Anforderungsprofils erzeugt das System eine Liste mit Werkstoff-Kandidaten. Auf dieser Ebene kann von vollständigen Daten ausgegangen werden, da die benötigte Genauigkeit hier beim Input durchaus die Schätzung fehlender Werte zuläßt. In einem sich anschließenden Evaluierungsschritt können die Kandidatenwerkstoffe bezüglich der interessierenden Eigenschaften ausgewertet und, graphisch unterstützt, verglichen werden. Ein bemerkenswerter Aspekt der Systemarchitektur ist die Tatsache, daß das unter GEM laufende Programm auf einem PC installiert wird und sich seine Daten online vom Host holt. Swindells unterstrich, daß die Entwicklung eines solchen Systems zu verschiedenen Zeitpunkten verschiedene Schwerpunktsetzungen verlangt. Nachdem zu Beginn der Arbeiten zum Thema "Organisation der Wissensbasis" viel experimentiert wurde und viele Erfahrungen gewonnen wurden, könne man sich jetzt wieder den Programmen und ihrer Funktionalität zuwenden.

Die Präsentation einer Benutzerschnittstelle in einem Vortrag ist in der Regel in ihren Möglichkeiten sehr beschränkt. Die Systemvorführung, angesetzt von 18⁰⁰ - 19⁰⁰ Uhr bot die Gelegenheit, das *Peritus-System* in Form einer PC-Demonstrationsversion kennenzulernen. Die Teilnehmer machten regen Gebrauch von dieser und den anderen vorbereiteten Möglichkeiten (Vorführung der *Steelfacts-Datenbank* und einer Benutzeroberfläche für die Werkstoffsuche, die von O. Starrowsky im Rahmen einer Studienarbeit entwickelt wurde).

Um 19³⁰ Uhr war in einem nahegelegenen Restaurant ein gemeinsames Abendessen vorbereitet. Der Szenenwechsel lud dazu ein, einmal tief Luft zu holen und an die Stelle von Konzentration und Anspannung eines langen und intensiven Workshop-Programms einen informellen Meinungsaustausch zu setzen. Die allgemeine Auffassung, wie sie Frau Brandstätter formulierte und wie sie ausnahmslos bestätigt wurde, war, daß die fachliche Diskussion bisher außerordentlich konstruktiv und produktiv verlaufen war. Das von manchem vorab kritisch betrachtete "Experiment eines so heretogenen Teilnehmerkreises" (Knorz) hatte auch die positiven Erwartungen übertroffen. Dies ergab eine Diskussion, die von Frau Brandstätter geleitet wurde und in der die Teilnehmer ihren Hintergrund, ihre Erwartungen an den Workshop und, soweit möglich, ihre bisherigen Schlußfolgerungen kurz umrissen.

Die Systemvorführungen, mit denen der Mittwochmorgen begann, hatten gegenüber dem Vorabend mit der Kunststoff-Datenbank *POLYMAT* einen weiteren Programmpunkt. Für den Sitzungsleiter Dr. Hitzemberger, Universität Regensburg, war es zunächst gar nicht einfach, wieder auf das Vortragsprogramm umzuschalten. Er hatte jedoch gute Argumente, denn es sollte im Folgenden nicht um Bestandsaufnahme oder Altbekanntes, sondern um die Art von "Zukunftsmusik" gehen, an der in der Forschung gerade gearbeitet wird. Auf der einen Seite also interessant - und auf der anderen Seite aus Sicht der Informationspraxis kritisch zu hinterfragen.

"Natürliche und simulierte Zeigegesten" war das Thema von Frau D. Schmauks, die vor dem Hintergrund von Arbeiten im DFG-geförderten *Sonderforschungsbereich 314 (Künstliche Intelligenz und wissensbasierte Systeme)* über Möglichkeiten und Probleme einer besonders wichtigen Klasse non-verbaler Kommunikation sprach. Zeigegesten treten in Alltagssituationen in verschiedensten Varianten auf und sie machen die verbale Kommunikation sicherer und effektiver (auch im Datenmaterial von WeBeS, den protokollierten Rechercsesitzungen, spielt das "Zeigen" eine wichtige Rolle). Aufbauend auf einer Systematik von Zeigegesten unterschiedlicher Form und Funktion diskutierte Schmauks die Möglichkeit, ausgewählte Fälle von Zeigen in einen

(natürlichsprachigen) Rechnerdialog mit einzubeziehen. Erfahrungen dazu liegen aus dem Projekt "XTRA – Ein natürlichsprachiges Zugangssystem zu Expertensystemen" vor, wo verschieden genaues Zeigen in einem (Lohnsteuer-)Formular ermöglicht wird (man muß sich klar machen, daß das Deuten auf ein Feld je nach Dialogkontext dieses Feld, die entsprechende Zeile, die entsprechende Spalte oder einen umfassenden Formular-Abschnitt meinen kann). Besonders hervorgehoben wurde die Tatsache, daß ein wie auch immer simuliertes "natürliches Zeigen" in einer technisierten Kommunikation zwischen Mensch und Rechner sich von einem natürlichen Zeigen wesentlich unterscheidet. Für Bewertung und Weiterentwicklung muß also untersucht werden, wie wirkliche Benutzer zur Lösung realer Probleme mit solchen neuen Möglichkeiten umgehen. Zwei Fehleinschätzungen sollten vermieden werden: Die Annahme, daß hier eine neue Option kurzfristig nur noch umzusetzen ist, oder daß hier eine für die Fachinformationspraxis nicht relevante Forschung betrieben wird.

Im ersten Teil des Beitrags "Von der benutzerfreundlichen Schnittstelle zum kooperativen Beratungsdialo" zeigte K. Ammersbach zunächst auf, inwieweit der Dialog, den ein Rechercheur mit einem Benutzer führt, ein "aufgabenorientierter Beratungsdialo" ist. Bereits in ihrer Diplomarbeit hatte Ammersbach solche Dialoge, damals am Beispiel der Befragung von Literaturdatenbanken, untersucht. Es zeigte sich, daß sowohl dem Zugang zu Literatur wie auch dem zu Faktenbanken ein gemeinsames Dialog-Grundmuster unterliegt und daß kooperatives Verhalten des Rechercheurs in allen Phasen der Recherche eine wesentliche Voraussetzung für den (relativen) Erfolg der Problemlösung ist. Wenn auch "kooperatives Systemverhalten" schon seit einiger Zeit gerade auch im Zusammenhang mit Benutzermodellierung ein aktives Forschungsthema ist, so muß doch klar gesagt werden, daß eine Beratungsleistung durch ein maschinelles System auf dem Niveau, das etwa der Rechercheur vorgibt, auch mittelfristig nicht zu denken ist: Eine Recherche nach Werkstoffen etwa, die mit der Adresse eines Erfinders endet, verlangt eine Kombination von Erfahrung, Überblickswissen und Detailkenntnis, die außerhalb der Reichweite heute diskutierter Expertensystem-Ansätze liegt.

Inwieweit auch wissensbasierte Systeme nicht die Lösung aller Probleme sein können wurde auch im zweiten Teil des Vortrags deutlich, in dem G. Knorz einen rekonstruierten realen Dialogabschnitt als Basis eines fiktiven Recherchedialogs an einem "WeBeS-Utopia"-System verwendete. Dieses System als Projektion der gegenwärtig diskutierten einschlägigen Konzepte aus der Künstlichen Intelligenz, der Software-Ergonomie und der Informationswissenschaft wurde an einem hochauflösenden farbigen "Folien-Bildschirm" demonstriert. Ein dynamisches

und für den Benutzer transparentes *Benutzermodell*, Wissen des Systems über den Gegenstandsbereich und über Recherchestrategien, ein flexibler, geführter Dialog, isolierte Elemente natürlichsprachiger Eingaben und vor allen Dingen graphisch-interaktive Dialogformen waren die kennzeichnenden Elemente. Obwohl in dem speziellen Fall das System schließlich doch "am Ende seiner Weisheit" war, so wurde andererseits sehr deutlich, daß bereits die heutige Technologie wesentliche Bausteine und Methoden einer neuen Generation von Informationssystemen bereitstellt (z.B. den sinnvollen Einsatz von Graphik) und daß wissensbasierte Systeme darüber hinaus einen wichtigen Beitrag liefern können und werden.

Den Abschluß des Workshops bildeten die Vorstellung und eine sich anschließende Plenumsvorstellung einer Liste von Statements, mit der G. Knorz die Schlußfolgerungen aus der in WeBeS geleisteten Arbeit und aus der bisherigen Diskussion der Veranstaltung zusammenfaßte:

- Gegenwärtig verfügbare Datenbank-Modelle und Systeme haben grundlegende Defizite hinsichtlich angebotener Datentypen und Operationen. Es ist mittel- und langfristig nicht sinnvoll, die Schwächen unzureichender Datenbanksysteme durch Frontend-Entwicklungen verdecken zu wollen.
- Wir wissen noch wenig über eine adäquate Modellierung von Werkstoff-Daten. Die Ausarbeitung und Erprobung eines konzeptionellen Schemas ist vordringlich. Analog etwa den anwendungsspezifischen Expertensystem-Shells sollten "Werkstoffdatenbank-Shells" angestrebt werden. Die Kombination öffentlicher und privater Datenbestände ist ein wichtiger Gesichtspunkt
- Verbesserte Datenbanktechnologie (s. o.) wird zu Systemen führen, deren Zielgruppe primär der professionelle Informationsspezialist ist.
- Informationssysteme, die sich an den Endbenutzer direkt wenden, sollten eine anwendungsspezifische Auswahl von Benutzerführungen (Skripts) und Methoden enthalten. Übersichtsinformationen, insbesondere der Vergleich mehrerer Werkstoffe müssen einfach und effektiv zur Verfügung stehen. Darüber hinaus muß eine Weiterverarbeitung der recherchierten Daten angeboten werden.
- Die Entwicklung besserer Benutzeroberflächen sollte das verfügbare Hardware-Angebot (speziell die hochauflösenden Bildschirme) nutzen. Stellt man die Entwicklungszeiten in Rechnung, so kann eine "konservative" Entscheidung sich bei Fertigstellung als völlig unangemessen herausstellen.
- Benutzeroberflächen sollten unter aktiver Einbeziehung der späteren Anwender entwickelt werden. Dazu sollte sich die Ein-

sicht durchsetzen, daß eine angemessene Entwicklungsumgebung (in Bezug auf Hardware, Software und Qualifikation der Entwickler) unverzichtbar ist. Ein Ansatz wie "rapid prototyping" ist bei weitem einem benutzerfernen Phasen-Vorgehen vorzuziehen. Low-cost Entwicklungsumgebungen verschwenden Geld bei späteren Nachbesserungen und durch schwache Akzeptanz und schlechte Erfahrungen bei der Benutzung.

- Eine Datenbank kann fehlendes Grundwissen beim Anwender über Werkstoffe, Werkstoff-Eigenschaften, Prüfbedingungen etc. nicht ersetzen.
- Expertensysteme sind dort ein sinnvoller Ansatz, wo sich ein spezialisierter Ausschnitt des Gesamtproblems definieren läßt. Mit entscheidend für den Erfolg wird die richtige Auswahl einer solchen Pilotentwicklung sein. Als Strategie ist vorzuschlagen: ein möglichst "kleines" Problem sollte gewählt werden, um dafür einen Prototypen zu erarbeiten. Die Aussagen in (5a) sind hier analog zu übertragen. Die weitere Entwicklung sollte von den Erfahrungen mit dem Prototypen abhängig gemacht werden.
- Dort, wo es tatsächlich auf eine fachliche Beratung ankommt, wie sie etwa durch eine Informationsvermittlungsstelle geliefert werden kann, ist diese durch ein Expertensystem nicht zu leisten (natürlich kann die Beratungsleistung des Informationsvermittlers durch ein XPS unterstützt werden). Dies liegt daran, daß der Verlauf des Beratungsgesprächs nicht prognostizierbar ist und der Gegenstandsbereich des Gesprächs die Art der zu verwendenden Informationsquellen nicht vorab eingegrenzt werden kann.
- Die Qualität der Daten muß ein wesentliches Argument für die Benutzung des Informationssystems sein. Die Anwender erwarten, daß sie mehr als nur Standardwerte (wie sie sie ohne Schwierigkeit in gedruckten Nachschlagewerken finden können) recherchieren können.

Die Diskussion bestätigte die vorgetragenen Feststellungen und versuchte, praktische Schlußfolgerungen daraus abzuleiten. So wurde überlegt, wie (und in welchen Bereichen) kostensparende und qualitätssteigernde Organisationsformen für einen gemeinsamen Dateninput von Herstellern, Anbietern und Nutzern gefunden werden können. Außerdem wurde die Einbindung der Werkstoffinstitute und der Hersteller in die weitere Diskussion für wesentlich gehalten. Die Vorbereitung des nächsten Planungsschrittes, Aussagen über die Zuordnung von Prioritätsstufen zu liefern, wurde an die WeBeS-Projektgruppe delegiert. Diese soll einen Vorschlag dazu schriftlich ausarbeiten und an die Workshopteilnehmer verschicken. Dieser Vorschlag und die einge-

gangenen Rückmeldungen werden in einem Arbeitskreis, den das BMWi etwa Ende dieses Jahres einberufen könnte, weiterbehandelt werden.

Damit war der Workshop "Zugang zu Werkstoff-Faktenbanken: Anforderungen und Konzeptionen - Stand und Perspektiven" zu Ende gegangen. Das Thema wird noch eine beträchtliche Zeit aktuell bleiben, doch ist für die weitere fachliche Diskussion als Vorbereitung praktischer Maßnahmen und Fortschritte ein ermutigender neuer Anfang gesetzt. Die Veranstaltung hat bewiesen: Wir brauchen für eine positive Entwicklung den Sachverstand aller beteiligten Disziplinen und Institutionen - und ein solcher Ansatz ist tatsächlich produktiv.

Prof. Dr. G. Knorz, Fachhochschule Darmstadt, Fachbereich Information und Dokumentation (IuD)

TINLAP-3 - THEORETICAL ISSUES IN NATURAL LANGUAGE PROCESSING

New Mexico State University, Las Cruces, NM, 7.-9. Januar 1987

Die TINLAP-3 steht als dritte Veranstaltung in einer Konferenzserie, die 1975 am MIT begonnen (SCHANK, WEBBER 1975) und 1978 mit der TINLAP-2 an der University of Illinois (WALTZ 1978) fortgeführt wurde. Anlaß für diese Konferenzserie war jeweils das Bedürfnis, sich nach einer gewissen Forschungsperiode über das zu besinnen, was erreicht wurde, und daraus Perspektiven für zukünftige Forschungsaufgaben abzuleiten: *It is therefore very timely for those in theoretical research in the area of natural language processing and understanding to stand back and survey the partial solutions available to problems in this field, the blind alleys that have been closed off, those that had been closed for many years but have now been declared reopend, and the new theories and representations that have been developed and published since the last such inventory was taken* (WILKS 1987, p.1). Aus dieser Motivation heraus wurde besonderes Gewicht auf Interdisziplinarität gelegt, d.h. es wurden auch Vertreter aus der Künstlichen Intelligenz, der Logik, Philosophie, Psychologie und Linguistik. Daß eine Grenzziehung zwischen diesen Bereichen und 'Computational Linguistics' ebenso schwierig und umstritten ist wie die Definition von 'Computational Linguistics' selbst, ist auch in Deutschland mittlerweile vielerorts diskutiert worden (z.B. BATORI et al. 1982, LAUBSCH 1985, S. 267ff) und wurde auf der diesjährigen TINLAP erneut deutlich.

Die Tagungsform der TINLAP-Konferenzen unterscheidet sich von anderen Konferenzserien dadurch, daß zu bestimmten Themenschwerpunkten Experten verschiedener Fachrichtungen oder theoretischer Ausrichtungen eingeladen wurden, die in

Form von Positionspapieren zu zentralen Problemen der Sprachverarbeitung Stellung beziehen. Dieser formale Rahmen ist für den programmatischen Charakter der Konferenz sicher förderlich. Für jeden Themenschwerpunkt war ein Organisator zuständig, der die inhaltlichen Schwerpunkte gesetzt, die Zusammensetzung des Expertengremiums bestimmt hatte und während der Tagung als Moderator fungierte.

Die Kenntnis der Positionspapiere, die in einem vor der Tagung versandten Proceedings-Band zusammengefaßt sind (WILKS 1987), wurde bei allen Teilnehmern vorausgesetzt, sodaß während der Konferenz oft nur Teilaspekte vorgetragen zu werden brauchten.

Insgesamt nahmen 169 Teilnehmer an der TINLAP-3 teil (lt. Registrierungsliste). Wie nicht anders zu erwarten kam die Majorität aus den Vereinigten Staaten (136), und andere Nationen waren nur gering repräsentiert (Großbritannien: 10, Frankreich: 6, Canada: 5, Irland: 4, Japan: 4, Bundesrepublik: 3, Jugoslawien: 1). Von den insgesamt 41 eingeladenen Referenten stammten 35 aus den USA, 4 aus Großbritannien und jeweils einer aus Canada und der Bundesrepublik.

Das Programm der TINLAP-3 gliederte sich in 9 Themenbereiche, die in Form von Podiumsdiskussionen gestaltet waren. Nach einer mehr oder weniger kurzen Einführung durch die jeweiligen Organisatoren stellten die eingeladenen Referenten ihre Beiträge in Kurzvorträgen vor (jeweils ca. 10 Min.), woran sich eine Diskussion unter den Referenten anschloß. Der letzte Teil bestand dann in der 'Öffnung des Podiums', d.h. das Publikum wurde einbezogen. Dieser letzte Teil war oft zu kurz bemessen, sodaß die Warteschlangen aus dem Publikum aus Zeitgründen – insgesamt waren nur 2 Stunden für jeden Themenbereich vorgesehen – meist abgebrochen werden mußten.

Generell kann man sagen, und dies ergibt sich bei einer von eingeladenen Vorträgen getragenen Veranstaltung ganz natürlich, daß das inhaltliche Niveau der TINLAP-3 deutlich über dem einer 'normalen' Konferenz, wie z.B. der COLING lag. Dadurch, daß eine thematische Konzentration schon vorgegeben war und hierzu die jeweiligen Experten zu Wort kamen, waren die drei Konferenztage von Anfang bis Ende durchgehend interessant. Es gab so gut wie keine 'Aussetzer' und durch die Besetzung der Podiumsdiskussion mit amerikanischen bzw. englischen Referenten war auch die Verstehbarkeit gewährleistet, eine Vorbedingung, die bei internationalen Tagungen oft Probleme bereitet. Bei fast allen Veranstaltungen war die Auswahl der Referenten gut getroffen, d.h. kontroverse Meinungen konnten vorgestellt, begründet und diskutiert werden. Für fast alle Veranstaltungen hätte man sich jedoch gewünscht, die doppelte Zeit zur Verfügung zu haben.

Sämtliche Podiumsdiskussionen wurden mit Videokameras aufgenommen. Eine Kopie kann bei A*C Publishing & Production, 1315 Colorado Ave., Las Cruces, N.M. 88001 erworben werden. Alle 5 Bänder kosten \$ 150,00, ein Einzelband \$ 40,00. Ferner ist geplant, überarbeitete Versionen der Tagungspapiere sowie wesentliche Diskussionsbeiträge als Buch herauszugeben.

Nun jedoch zu den 9 Podiumsdiskussionen im einzelnen. Nach dem Titel werden jeweils die eingeladenen Referenten aufgeführt, der Name des Organisators ist kursiv gedruckt.

I. WORDS AND WORLD REPRESENTATIONS

Don Walker (Bellcore)

Bran Boguraev (Cambridge)

Bob Amsler (Bellcore)

Jerry Hobbs (SRI)

Judy Kegl (Princeton)

Der Hauptdisput handelte davon, ob die ((computer-) linguistische) Tätigkeit, Lexika und sprachlich kodierte Enzyklopädien zu erstellen, überhaupt noch sinnvoll ist, ob nicht die Erforschung von Weltwissen, wie es im KI-Paradigma gefordert wird, einer rein sprachimmanenten Repräsentation den Boden entzieht. Die Mehrzahl der Referenten führte Argumente dafür ins Feld, daß maschinelle Lexikographie ein theoretisch wie praktisch sinnvolles Aufgabengebiet ist. Don Walker illustrierte diese Ansicht mit einer formalen Vorstellung aller weltweit bisher auf diesem Gebiet arbeitenden Gruppen und mit einer Vorstellung der Organisationsstruktur für die zukünftige Forschungsförderung (Lexical Advanced Research Projects Agency), die mit verschiedenen Direktoraten und Zuständigkeitshierarchien einer Organisation wie der DARPA nicht unähnlich erschien. Die mehr KI-orientierte Gegenposition wurde tapfer und mit glänzenden Beispielen von Jerry Hobbs verteidigt, der für eine Integration lexikalischer Semantik mit Ansätzen zur Formalisierung von Weltwissen plädierte. Die anderen Podiumsreferenten räumten zwar ein, daß Weltwissen auch notwendig sein, nur sei dies ein äußerst kompliziertes Unterfangen, und mit traditionellen Ansätzen würde man zumindest auf sicherem Boden stehen. Außerdem könne man den Bedürfnissen potentieller Abnehmer – Verlage wünschen sich z.B. riesige Lexika mit 5 Mio. Einträgen – besser gerecht werden.

II. UNIFICATION AND THE NEW GRAMMATISM

Fernando Pereira (SRI)

Gerald Gazdar (Sussex)

Steve Pulman (Cambridge)

Aravind Joshi (U. Penn)

Mitch Marcus (Bell Labs.)

Martin Kay (Xerox-PARC)

Bei dieser Podiumsrunde waren sich eigentlich alle Teilnehmer darin einig, daß unifikationsbasierte

Grammatikformalismen auf Grund ihrer mathematischen Eigenschaften anderen Ansätzen überlegen sind, in ähnlicher Weise wie kontextfreie Grammatiken über theoretisch 'schönere' Eigenschaften verfügen als kontextsensitive. Joshi's Einwand, man müßte noch abwarten, inwieweit dieses theoretisch saubere Konstrukt sich auch im semantischen Bereich als brauchbar erweisen würde und der Anspruch der Kompositionalität (Gazdar) eingelöst werden können, änderte wenig am grundsätzlichen Konsens. Die einzige (und einsame) Gegenposition wurde von Mitch Marcus vertreten der sich darüber wunderte, daß dieses Thema überhaupt auf einer theoretischen Tagung über 'Computational Linguistics' auftauchte, und nicht – wie er eher erwartet hätte – auf einer Tagung über Programmiersprachen oder theoretische Informatik. Seine Kritik an der Unifikation läßt sich wie folgt zusammenfassen (eine schriftliche Version ist leider im Tagungsband nicht enthalten): eine wohldefinierte (formale) Semantik des Unifikationsformalismus enthält keine Gewähr für das, was man mit den Ergebnissen anfangen kann; Unifikation ist grundsätzlich nicht modularisierbar; Unifikation hat wenig mit kognitiven Aspekten der Sprachverwendung zu tun. Ein Manko der Themenstellung (und der Teilnehmerauswahl) lag vielleicht darin, daß nur innerhalb eines Paradigmas diskutiert wurde, daß die Einwände von Marcus regelrecht auf Unverständnis stießen und daß kein Vertreter eines explizit anderen Grammatikformalismus zu Wort kam.

III. CONNECTIONIST AND OTHER PARALLEL APPROACHES TO NATURAL LANGUAGE PROCESSING

Dave Waltz (Thinking Machines)
 Martin Cottrell (UCSD)
 Gene Charniak (Brown Univ.)
 Jay McClelland (CMU)
 Wendy Lehnert (U. Mass.)

Kontroverser als zuvor ging es auf der letzten Podiumsrunde des ersten Tages her: ein relativ neues, aus der kognitiven Psychologie stammendes Theoriegebäude wurde diskutiert. Dankenswerterweise gab Dave Waltz zunächst eine knappe Einführung in das, was mit dem Begriff Konnektionismus verbunden ist und aus welchen Wurzeln sich diese Theorie herleitet; denn die intime Kenntnis des Grundlagenwerks (RUMELHART et al. 1986) konnte nicht vorausgesetzt werden. Die Referenten versuchten dann, Konnektionismus als allgemeines und kognitiv adäquates Prozeßmodell vorzustellen (Cottrell in Bezug auf Semantik, McClelland in Bezug auf Tiefenkasusinterpretation), ihre skeptische Haltung auszudrücken (Lehnert über den sog. TWITIT-approach (TWeak It Til It Thinks), Kay in der Hinsicht, daß – entgegen allen vernünftigen Bedenken – der Konnektionismus eines Tages möglicherweise doch Lösungen für Probleme bietet, bei denen man

sonst ratlos ist), oder sogar eine dezidiert ablehnende Haltung zu begründen (Charniak: ... *I think it is still fair to say that connectionist networks cannot represent propositions at all. To the degree they can it is like a horse walking on two legs – it does not do it very well, and you loose all that is distinctive about the creature. I use a horse because I want to gallop. If I have to walk I would rather do it in Lisp.*). Diese letzte Position wurde vor allem dadurch untermauert, daß keinerlei Einsicht in die tatsächlich ablaufenden Prozesse möglich ist, daß man also nicht erklären kann, wie ein konnektionistisches Verfahren zu einer bestimmten Entscheidung kommt.

Das Thema Konnektionismus tauchte während der weiteren Tagung noch häufiger in Form von Publikationsfragen auf, ohne daß daraus jedoch weitere Einsichten erwachsen. Die Antwort auf viele Fragen war 'that really is a hard problem and we have to see how to solve it'; hierbei schien – jedenfalls für die Konnektionismus-Vertreter (Waltz war eine erfrischend undogmatische Ausprägung dieses Typs) – die prinzipielle Lösbarkeit jeweils präsupponiert zu sein.

IV. DISCOURSE THEORY AND SPEECH ACTS

Barbara Grosz (SRI, Harvard)
 Julia Hirschberg (Bell Labs.)
 Ray Perrault (SRI)
 Bob Wilensky (Berkeley)
 Franz Guenther (Tübingen)

Wilensky führte eine Unterscheidung zwischen Linken und Rechten in Bezug auf die Pragmatik ein: rechts ist demnach eine wahrheitsfunktionale Semantik, die Pragmatik eigentlich ausklammert, links ist die Gegenposition, die Semantik aus der Pragmatik heraus zu begründen versucht. Perrault argumentierte dafür, daß eine bloße Etikettierung von Äußerungen durch Sprechaktbezeichnungen nicht ausreichend ist, sondern vielmehr die mentalen Zustände und Zustandsänderungen im Kommunikationsprozeß modelliert werden müßten. Julia Hirschberg betonte die notwendige und mittlerweile angehbare Integration prosodischer Faktoren in die Diskurspragmatik (man kann satzübergreifende Phänomene ohne Rücksicht auf Intonation überhaupt nicht sinnvoll studieren). Guenther versuchte schließlich, die Diskursrepräsentationstheorie als Repräsentation mentaler Modelle von Sprecher und Hörer vorzuschlagen.

Generell wurde hier – wie auch in der nachfolgenden Diskussion – versucht, die Grenze (und ihre Rechtfertigung) zwischen Semantik und Pragmatik zu definieren, und die Relation zwischen isolierten Sätzen und Diskurszusammenhängen aufzudecken (ist die Untersuchung von Einzelsätzen überhaupt sinnvoll, taugen Erkenntnisse daraus für irgendetwas?). Die Antworten für einen solchen komplexen Gegenstandsbereich waren dann auch unterschiedlich, aber nicht prinzipiell gegensätzlich. Ein Disput z.B. zwischen Perrault und Martin Kay wäre vermut-

lich pointierter gewesen. Als Tendenz jedoch, und dies gilt für die gesamte Tagung, soll festgehalten werden, daß mentale Modelle, d.h. die explizite Modellierung von Sprecher- und Hörerannahmen, als erfolgversprechende Strategie angesehen werden, wenn es auch noch keine Patentrezepte für ihre Realisierung gibt.

V. WHY HAS THEORETICAL NLP MADE SO LITTLE PROGRESS

Roger Schank (Yale)

Norm Sondheimer (USC-ISI)

Larry Birnbaum (Yale)

Ivan Sag (Stanford)

David Israel (SRI)

Diese Podiumsdiskussion war sicherlich diejenige mit dem größten Unterhaltungseffekt, gleichzeitig aber auch diejenige, die dem Ziel der TINLAP am nächsten stand. Der Unterhaltungseffekt war durch die Zusammensetzung der Referenten und ihre jeweils idiosynkratische Lebhaftigkeit bedingt; doch heißt dies nicht, daß entertainment die inhaltliche Diskussion dominiertem, sondern eher das Gegenteil war der Fall: die inhaltliche Diskussion profitierte von einer exzellenten Präsentation.

Norm Sondheimer, auf den die Themenformulierung ursprünglich zurückging, zeichnete zunächst ein ziemlich düsteres Bild über den Forschungsfortschritt. Eigentlich hätte man mehr Fehler gemacht als Erkenntnisse gewonnen. Das Operieren mit Gegenbeispielen sei gegenüber konstruktiven Lösungen zu ausgeprägt gewesen. Als Konsequenz wurde 'Let's treat our work as that of a successive approximation' empfohlen, und eine vehemente Aufforderung für zukünftige, bessere Arbeiten ausgesprochen: laßt uns große Grammatiken, große Wissensbasen implementieren und austesten, laßt uns große Systeme bauen, aus denen wir möglicherweise nur lernen, was wir falsch gemacht haben, denn anders lernen wir gar nichts.

Ivan Sag, mit nicht zu übersehenden Seiten blicken auf Roger Schank, hielt dann ein glänzendes Plädoyer für die Notwendigkeit, syntaktische Sachverhalte zu berücksichtigen. Er zeigt an einfachen Beispielsätzen, was passiert, wenn etwa Subject-Verb-Kongruenz, Reflexivierung oder Interpunktion nicht berücksichtigt wird, bzw. daß diese Information für Sprachverstehensprozesse essentiell ist und nicht aus anderen Quellen beschafft werden kann. Seine Prognose, insbesondere soweit es die linguistikbasierten Vorgehensweisen (Unifikation) betrifft, nahm sich gegenüber der von Sondheimer recht optimistisch aus.

Larry Birnbaum betonte die kognitive Ausprägung im Gegensatz zu einer linguistisch-deskriptiven Richtung. Letztere habe in der Vergangenheit zu stark vorgeherrscht, und es sei an der Zeit, dies zu ändern: 'Let's put the AI back in NLP. We might even put some of the fun back in at the same time'.

David Isarel war ein bißchen der Clown dieser Diskussion. Es seien riesige Fortschritte in den Gebieten Syntax (seit 1957, Chomsky), Semantik (seit 1879, Frege) und Pragmatik (seit 1957, Grice) erreicht worden, aber dies sein theoretischer Fortschritt. Praktischer Fortschritt sei kaum angebar.

Roger Schank schließlich bedauerte die Rückentwicklung zu traditionell linguistischen Bereichen: '*I am against 1001 researchers doing syntax, as syntax isn't that interesting*'. Er betonte jedoch gleichzeitig einen generellen Meta Fortschritt: '*It's not so much that we have developed solutions to all our problems, although we certainly have developed some solutions. The point is that we are starting to understand what the problem is, and this is much more important*' (WILKS 1987, p. 104). Eine zunächst mit Erstaunen, sodann mit Beifall bedachte Äußerung von Schank soll nicht unerwähnt bleiben: '*I do not think that syntactic restrictions are negligible, I only think that there are further and possibly more challenging issues at hand*'. Dies ist sicherlich eine Sinneswandlung bei einem der führenden Wissenschaftler, vor allem, wenn man früher publizierte Äußerungen damit vergleicht: '*The problem in language understanding work is doing that kind of reasoning (that bezieht sich auf script application in SAM). If we cannot do that, we cannot do anything more than play with words. We must forget about the words. What we have to do is to get to the ideas; the words themselves have to be gotten around.*' (SCHANK 1983, p. 138).

Generell ist der Appellcharakter dieser Podiumsdiskussion nicht zu überhöhen gewesen, der dahinging, nicht nur nachzudenken, danach Papier zu beschreiben, sondern vielmehr implementierte Systeme vorzuweisen, die dann auch möglichst noch kognitiv adäquat sein sollen.

VI. FORMAL VERSUS COMMON SENSE SEMANTICS

Yorick Wilks (NMSU)

Geoff Nunberg (Stanford/Xerox-PARC)

Wendy Lehnert (U.Mass.)

Karen Sparck-Jones (Cambridge)

Susan Stucky (Stanford-CSLI)

Wilks plädierte für das Ernstnehmen von formalen Semantikansätzen für die Gebiete, die bekanntermaßen nicht leicht formalisierbar sind (z.B. Referenz): '*we should notice the repeated offer, to our sloppy and heuristic discipline, of a real serious theory for the field ... The chances always are that this prescription is unrelated to the disease; has Chomsky really helped language processing? We do indeed need a good theory but these are quack cures trading chiefly on the fears and inadequacies of practitioners and patients in the field*'. Karen Sparck-Jones argumentierte für Bedeutungspostulate als das Rückgrat jeder vernünftigen Semantiktheorie und empfahl '*you have to define primitives and word*

senses, but don't worry about quantifier structures'. Wendy Lehnert ging davon aus, daß sowohl Chomsky als auch Schank eine Unterscheidung zwischen Semantik und Pragmatik fordern. Die Bewertbarkeit eines Semantikersystems wäre völlig beliebig, wenn man hierzu nicht pragmatische Kriterien heranziehen kann. Susan Stucky setzte sich mit Situationssemantik und Situationstheorie auseinander. Als Fazit ihrer Ausführungen ergab sich, daß eine *'situated perspective on natural language processing'* vor allem für theoretische Ansätze interessant ist. Für 'Computational Linguistics', insbesondere für die semantische Repräsentationsebene in Systemen, wäre jedoch keine direkte Übertragbarkeit gegeben und Aussagen hierzu wären noch verfrüht. Geoff Nunberg als 'common-sense linguist' betonte schließlich den Zusammenhang zwischen sozialer Verpflichtung und Wortbedeutung, ohne jedoch selbst irgendwelche Verpflichtungen hinsichtlich sprachverstehender Systeme eingehen zu wollen.

Die Diskussion drehte sich darum, was das Typische an KI-Methoden ist, wie sie heute in Erscheinung treten, und wie sie eigentlich aussehen sollten. Eine einheitliche Sichtweise war ebensowenig zu erkennen wie eine Menge dediziert einheitlicher Sichtweisen.

VII. REFERENCE: THE INTERACTION OF LANGUAGE AND THE WORLD

Doug Appelt (SRI)
Deborah Dahl (SDC Inc.)
Bonnie Webber (U. Penn.)
Amichai Kronfeld (SRI)
Brad Goodman (BBN)

Diese Podiumsdiskussion befaßte sich mehr mit Referenzproblemen als daß sie Lösungen in der Form von *'how might these insights be incorporated into understanding systems'* angeboten hätte. Es wurde sogar explizit gesagt, daß solche Lösungen bisher nicht vorstellbar seien. Die mit zahlreichen Beispielen angeführten Probleme (Kronfeld: Planungsaspekte; Goodman: Referenzfehler an Hand des bereits hinlänglich bekannten Wasserpumpendialogs (weder der Dialog noch die Probleme noch die vorgeschlagenen Lösungen haben sich in den letzten 5 Jahren geändert); Dahl: durch welche linguistischen Ausdrücke wird ein neuer Referent in den Diskurs eingeführt?; Webber: komplexe Diskursobjekte, event references) waren aus der Literatur allerdings oft schon bekannt und auch die Beispiele erzeugten oft einen déjà-vu Effekt. Insgesamt hat diese Veranstaltung wenig dazu beigetragen, neue Einsichten in das komplizierte Problem sprachlicher Referenz zu vermitteln.

VIII. METAPHOR

Deirdre Gentner (Univ. of Illinois)
Andrew Ortony (Univ of Illinois)
Ed Plantinga (Toronto)

George Lakoff (UC Berkeley)
Geoff Nunberg (Stanford/Xerox-PARC)

Das Verstehen metaphorischer Ausdrücke ist innerhalb der Sprachverarbeitung ein relativ neues und vergleichsweise wenig bearbeitetes Gebiet. Die Podiumsdiskussion zu diesem Thema diente dann auch dazu, die existierenden Ansätze zusammenzufassen und eine gemeinsame Problemlösungsform definieren zu können. Bis auf Gentner und Plantinga waren die Äußerungen hochgradig theoretischer Natur (um nicht zu sagen: literaturwissenschaftlicher Natur) und hätten ebenso auf jeder Philologentagung getätigt werden können. Einigkeit schien insofern zu herrschen, als man für die Behandlung von Metaphern keine Spezialprozessoren annimmt, die erst dann aktiv werden, wenn 'normales' Verstehen scheitert, sondern daß Metaphernanalyse in einen generellen Verstehensprozeß integriert werden muß.

IX. NATURAL LANGUAGE GENERATION

Aravind Joshi (U. Penn)
Dave McDonald (U. Mass. Amherst)
Doug Appelt (SRI)
Bill Mann (USC-ISI)
Mitch Marcus (Bell Labs.)
Tony Kroch (U. Penn.)

Die letzte Podiumsdiskussion der TINLAP-3 beschäftigte sich mit dem Thema Generierung sprachlicher Äußerungen. Insbesondere sollte geklärt werden, ob Analysegrammatiken in irgendwie 'umgedrehter Weise' auch für die Generierung eingesetzt werden können (Bidirektionalität). Doug Appelt stellte eine allgemeine Systemarchitektur vor, nach der die Anwendung einer Grammatik für Analyse und Generierung möglich erscheint. Dieses Prinzip ähnelt dem PATR-II Ansatz von Shieber. Bedingung hierfür ist jedoch, daß die Grammatik semantikkfrei ist. Bill Mann zeigte auf mehreren Ebenen, daß die Probleme in der Generierung ganz andere sind als die beim Verstehen von sprachlichen Äußerungen; deshalb Bidirektionalität auch keine sinnvolle Annahme. Gleichzeitig wies er darauf hin, daß Generierung vermutlich ein komplizierterer Prozeß ist als Analyse, zumindest dann, wenn man die Inhaltsbestimmung mit zur Generierung zählt. Dave McDonald sprach sich ebenfalls gegen ein einfaches Umdrehen aus: *it is conceivable to share the knowledge, but not at all the processes. In Generation we have a drastically different information flow.* Der Vortrag von Tony Kroch stammt aus dem Gebiet der Psycholinguistik, war eine Vorlesung des Proceedings-Textes, und es wurde nicht klar, welchen Bezug der Inhalt zum Thema der Veranstaltung hatte. Hierin kann man einen der wenigen 'Aussetzer' auf der TINLAP erkennen. Der letzte Vortrag, von Mitch Marcus, griff dann das an, was bisher in der Generierung geleistet wurde: *'Generation programs do not mean what they say'*. Über die frühen Ansätze in BABEL, bei dem

der Prozeß der Wortwahl wirklich modelliert wurde, sei man seither nicht hinausgekommen. Er schlug zwei Tests für die Güte eines Generators vor: erstens müßte ein System, das lexikalische Semantik enthält, in der Lage sein, Paraphrasen erzeugen zu können; zweitens dürfte die Änderung des Diskursbereiches keine einschneidenden Auswirkungen auf den Generator haben, in ähnlicher Weise wie dies für Parser gilt. Sein konstruktiver Vorschlag war, daß man Syntax durch eine bessere und allgemeinere Form von lexikalischer Semantik ersetzen müsse.

Mit dem Ende dieser Podiumsdiskussion war dann auch das Ende der Tagung definiert. Man hätte sich eine Zusammenfassung, oder auch nur persönliche Eindrücke aus dem Kreis der Teilnehmer gewünscht (Martin Kay ist für solche Gelegenheiten immer ein vielversprechender Kandidat), und nicht einfach ein Aufhören der Tagung. Für die noch nicht abgereisten Teilnehmer gab es am Samstag dann allerdings noch eine Party in der Wilks'schen Residenz, die von vielen Anwesenden für eine rückblickende Einschätzung genutzt wurde.

Literatur

I. Bátori, J. Krause, H.D. Lutz (eds.): *Linguistische Datenverarbeitung. Versuch einer Standortbestimmung im Umfeld von Informationslinguistik und Künstlicher Intelligenz. Tübingen: Niemeyer 1982*

J. Laubsch (ed): *Proceedings of the 8th German Workshop on Artificial Intelligence (GWA) 1984. Berlin: Springer 1985*

D.E. Rumelhart, J.L. McClelland and the PDP Research Group (eds.): *Parallel Distributed Processing: Explorations in the microstructure of cognition. Vol. 1 & 2. Cambridge: MIT Press 1986*

R. Schank: *Intelligent Advisory Systems. In: P. Winston, K. Prendergarst (eds.): The AI-Business. Cambridge: MIT Press, 1983, 133-148*

R. Schank, B.L. Nash-Webber (eds.): *Theoretical Issues in Natural Language Processing. 10.-13. June 1975, MIT, Cambridge, MA.*

D.L. Waltz (ed.): *Theoretical Issues in Natural Language Processing-2. 25.-27. July, 1978, Illinois, Urbana*

Y. Wilks (ed.): *Theoretical Issues in Natural Language Processing-3. Position Papers. 7.-9. January 1987, Las Cruces, NM.*

Prof. Dr. Wolfgang Hoepfner, EWH Rheinland-Pfalz, Abt. Koblenz, Universität Hamburg

FIRST EUROPEAN WORKSHOP ON LANGUAGE GENERATION

Abbaye de Royaumont, Paris, 24. - 26. Januar 1987

Ich muß zugeben, ich war mit gemischten Gefühlen nach Paris gefahren. Konnte es nach dem erfolgreichen Generierungs-Workshop im August in Nijmegen nun bereits im Januar wieder einen interessanten Workshop geben? Der Verlauf dieses Workshops hat mich diese Frage bejahen lassen. Dies lag sicherlich mit daran, daß es dem Veranstalter Michael Zock vom LIMSI, Paris gelungen war, ein anregendes Ambiente zu organisieren und diskussionsbereite und interessante Referenten zu gewinnen.

Auch erwies sich der europäische Rahmen als ziemlicher Gewinn. Während auf internationalen Konferenzen und Workshops normalerweise die Amerikaner sehr dominierend sind und damit die Generierung von Englisch zum primären Thema wird, zeigte sich hier auf der Tagung, daß bei europäischen Wissenschaftlern doch soviel Kenntnisse auch anderer Sprachen vorhanden sind, daß verschiedenartige Systeme zur Generierung von Italienisch, Französisch, Deutsch etc. vorgestellt und gewürdigt werden können. Dies, glaube ich, gibt mit eine Berechtigung für eine solche europäische Veranstaltung in Ergänzung zu den internationalen.

Auch wurde mit der Zahl der Teilnehmer die kritische Grenze nicht überschritten, bei der es dann zu typischem Konferenzverhalten kommt. Durch die überschaubare Teilnehmerzahl stellte sich schon am ersten Abend eine lockere und ungezwungenen Atmosphäre ein, sicher begünstigt durch das Flair der ehemaligen Zisterzienser-Abtei (bei der man nicht verwundert gewesen wäre, wenn jeden Moment die drei Musketiere um die Ecke gekommen wären) und die - sieht man vom Frühstück ab - exzellente Küche des Hauses.

Zum eigentlichen Workshop-Programm: An 2 1/2 Tagen wurden insgesamt 18 Vorträge gehalten. Versucht man die Vorträge etwas zu gruppieren, so bieten sich die folgenden Kategorien an:

- Vorträge, die linguistisch oder grammatische Formalismen betrafen
- Vorträge, die über Systemimplementationen berichteten
- Vorträge, die psycholinguistisch orientiert waren
- Vorträge, die aktuelle oder geplante Anwendungen von Generationen behandelten

Zur ersten Kategorie: In *Concerning the Logical Component of a Language Generator* entwickelte Simon DIK (Amsterdam) an Beispielen, warum die Sprache der funktionalen Grammatik ein möglicherweise geeignetes Mittel für die einem Generator zugrundeliegende logische Komponente sein könnte.

Zur aktuellen Diskussion über lexikalisch-funktionale Grammatiken gab es zwei interessante Bei-

träge (von denen allerdings nur der erste tatsächlich gehalten wurde: Russel BLOCK (Hamburg) war an unerbitterlichen Zöllnern gescheitert, die ihn mangels Visa wieder zurückgeschickt hatten). Helmut HORACEK (Hamburg) diskutierte in seinem Beitrag *The Application of Unification for Syntactic Generation in German*, mit welchen Techniken im LOKI-Projekt versucht wird, bei der Generierung mit LFG-Grammatiken Backtracking zu vermeiden und wie bestimmte Phänomene des Deutschen, wie freie Wortordnung, durch Einführung geeigneter Metavariablen behandelt werden. In dem leider nur als Papier vorliegenden Beitrag von Block *Can a Parsing Grammar be used for Natural Language Generation?* – Untertitel: *A negative Example of LFG* – wurde eine sehr kritische Einschätzung lexikalisch-funktionaler Grammatiken vorgenommen, die sicherlich noch manche Diskussion nach sich ziehen wird.

Harry BUNT (Tilburg) stellte in seinem Vortrag *A Phrase-Structure Grammar with Discontinuous Unit* einen Grammatikformalismus vor, den er 'augmented discontinuous phrase structure grammar' nennt. Diese Grammatik produziert Syntaxbäume mit Diskontinuitäten.

Domenico PARISI (Rom) berichtete in *A lexically Based Sentence Generation System* über den von ihm verfolgten Ansatz, mit den Lexikoneinträgen sogenannte "positioning instructions" zu verbinden, die festlegen, in welcher relativen Position (bezogen auf andere Satzelemente) ein gegebenes Wort im Satz dann angeordnet sein wird.

Robin P. FAWCETT (Polytechnics of Wales) konnte in *Language Generation as Choice in Social Interaction* zweierlei verbinden: Einen Überblick über das unter seiner Leitung stehende und mit erheblicher Industriebeteiligung gestartete COMMUNAL-Projekt zur Sprachverarbeitung, dann die Darlegung der theoretischen Basis dieses Vorhabens – systemic grammar –, wobei deutlich herausgearbeitet wurde, warum diese linguistische Theorie sich besonders für Aufgaben der Generierung eignet. Zur Kategorie 'Berichte über implementierte Systeme' lassen sich die folgenden Beiträge zählen:

Hans-Joachim NOVAK (früher Hamburg, jetzt LILOG/Stuttgart) konzentrierte sich in *Referring Expressions in a Dynamic Environment* auf den Aspekt der Generierung von beschreibenden Nominalphrasen in dem von ihm implementierten Generierungssystem für das NAOS-Projekt (Beschreibung von Straßenszenen auf der Schlüterstraße).

George HOUGHTON (Sussex) berichtete in *The Use of Intonation in Computer Generated Dialogue* über einige der in seinem System zur Erzeugung gesprochener Sprache – Dialoge über die Mini-Welt eines Roboters – verwendeten Techniken (z.B. Akzentplatzierung und Tonfallbestimmung).

Dietmar RÖSNER (Stuttgart) gab in seinem Beitrag *The Generation System of the SEMSYN Project – Towards a Task-Independent Generator*

for German einen vergleichenden Überblick über die bisher implementierten Anwendungen des SEMSYN-Generators und zeigte in einer Synopse auf, welches Fragment des Deutschen bereits durch diesen Generator abgedeckt wird.

Nathalie SIMONIN (Paris) berichtete in *An Approach for Creating Structured Text* über ihre Arbeiten zur Generierung französischsprachiger beschreibender Texte über den Inhalt der geographischen Datenbank ATLASECO.

In *Some Experience on Natural Language Generation* präsentierte Giovanni ADORNI (Genua) zwei von ihm verfolgte Ansätze für taktische Generatoren. Erste Methode: Ausgangspunkt ist ein concept-dependency-Netz und eine transformationelle Grammatik. Zweite Methode: Lexikalisch dekomponierte semantische Repräsentation, bei der Wortbedeutungen intern von logischen Prädikaten repräsentiert sind.

Robert DALE (Edinburgh) sieht seine Arbeiten über *The Generation of Referring Expressions in Structured Discourse* als Weiterentwicklung der Ansätze von Grosz und Sidner, sein Gebiet ist die Generierung von Rezepten.

Chris MELLISH (Sussex) beschrieb in *Natural Language Generation from Plans* seine in Prolog ausgeführten Experimente zur Generierung von Kochrezepten aus den Repräsentationen der Pläne, die den Kochaktivitäten zugrundeliegen. Er ist bei diesen Arbeiten primär daran interessiert, eine Art obere Grenze für die Komplexität von Generierungsprozessen bestimmen zu können.

Zur Kategorie psycholinguistischer Referate lassen sich die folgenden zählen:

Vanda ZAMMUNER (Padua) rekapitulierte in *The Process of Discourse Planning and Production: a Model and some Data* ihr theoretisches Modell, das alle Aspekte, die in Textplanung und -produktion eingehen, zusammenfassen soll.

Annie PIOLAT und Fernand FARIOLI (Aix-en-Provence) berichteten in *Effect of the Macro-Control of Information on the Temporal Characteristics of Text Production* über computergestützte Experimente zur Textplanung, die sie mit Schülern gemacht hatten. Diesen wurden Textteile vorgegeben, die sie dann zu einem Text zusammenstellen sollten.

In *Automatic and Executive Processing in Semantic and Syntactic Planning: A Dual Process Model of Speech Production* stellte Trevor HARLEY (Warwick) sein Modell der Sprachproduktion vor. Er geht von zwei Typen von Verarbeitungsmechanismen aus: einem schnellen, parallelen Mechanismus, der nicht in Interferenz mit dem Inhalt des Arbeitsspeichers gerät, und einem anderen, 'executive processing' genannten, der langsam und seriell ist und der vom Arbeitsspeicher aus beeinflusst werden kann.

Zur Gruppe "Anwendungen von Generatoren":

In *Talking on the Top of their Heads* stellte Michael ZOCK (Paris) verschiedene seiner Überlegungen vor, wie Generatoren zu Studien- und Lernzwecken eingesetzt werden könnten.

V. KORST, BAKKER und V.SCHAAIK von der Universität Amsterdam präsentierten ihr System, mit dem Studenten durch die Benutzung eines Generators grammatische Formalismen lernen können (*A Generator for Teaching Linguistics*).

Vielleicht in keine der obengenannten Kategorien passend, aber dennoch ein wichtiger Beitrag zur Tagung war der humorvoll präsentierte Erfahrungsbericht von KWEE Tjoe-Liong (Amsterdam): *Natural Language Generation, one Individual Implementor's Experience*. Er berichtete hier von den Schwierigkeiten, denen er sich gegenüber sah, als er bei längeren Experimenten mit dem TEXT-System dieses ausweiten wollte.

Resume:

Daß die Tagung ein Erfolg war, kann wohl unbestritten behauptet werden. So herrschte in der Schlußdiskussion auch Einmütigkeit, daß im Jahre 1989 unbedingt eine Fortsetzungsveranstaltung stattfinden sollte. Als Organisationskomitee dafür erklärten sich Chris MELLISH, Robert DALE und Michael ZOCK bereit. Der SECOND EUROPEAN WORKSHOP ON LANGUAGE GENERATION wird voraussichtlich in der Nähe von Edinburgh stattfinden. Weitere Informationen (gegebenfalls auch Kopien der eingereichten Papiere) sind über Michael ZOCK erhältlich.

Adresse: Dr. Michael Zock, LIMSI, B.P. 30, F-91406 Orsay C dex France

Dr. Dietmar Rösner, Projekt SEMSYN, Institut für Informatik Universität Stuttgart, Herdweg 51, 7000 Stuttgart 1

MACHINE TRANSLATION SUMMIT

EINE BEWERTENDE BESCHREIBUNG DER VORGESTELLTEN SYSTEME UND PROJEKTE IM BEREICH DER MASCHINELLEN ÜBERSETZUNG (MT)

Hans Billing

GMD Zentralbereich Projektträgerschaft und -Förderung
Herriotstraße 5, D-6000 Frankfurt 71

Gliederung

Einleitung

A. Kommerzielle Systeme

- I. SYSTRAN
 - I.1 Präsentation
 - I.2 SYSTRAN aus Benutzersicht
- II. METAL
 - II.1 Präsentation
 - II.2 METAL aus Benutzersicht
- III. DLT
- IV. ATLAS
 - IV.1 Präsentation
 - IV.2 ATLAS II aus Benutzersicht
- V. HICATS/JE
 - V.1 Präsentation
 - V.2 HICATS/JE aus Benutzersicht
- VI. TAURAS
 - VI.1 Präsentation
 - VI.2 TAURAS aus Benutzersicht
- VII. PIVOT
 - VII.1 Präsentation
 - VII.2 PIVOT aus Benutzersicht

B. Projekte im Bereich maschineller Übersetzung: Status und Perspektiven

- I. EUROTRA
- II. ALVEY
 - II.1 Prototyp AIDTRANS
 - II.2 Prototyp NTRAN
 - II.3 Anhang: "The Englishmen's Keyboard"
- III. ODA
- IV. GETA
- V. MU
 - V.1 MU I
 - V.2 MU II
- VI. CMU
- VII. EDR

C. Schluß

Anhang

EINLEITUNG

Das Ziel des 1. Machine Translation Summits in Japan (Hakone, 17. bis 19. September 1987) war es, Benutzer, Entwickler und politische Entscheidungsträger zusammenzubringen, um die Fragen

zukünftiger Nutzung und der bereits erzielten Leistung bisheriger Systeme zu diskutieren. Das Zusammentreffen sollte auch auf dem Hintergrund jüngster Anstrengungen Japans (s.u.), die Wei-

terentwicklung praktischer Maschinenübersetzung zu fördern und mögliche neue Perspektiven in der Umsetzung dieser Systeme aufzuzeigen, gesehen werden.

Entsprechend dieser Zielsetzung war der Summit in 4 Teilbereiche gegliedert:

- **Kommerzielle MT-Systeme**
Hier wurde neben dem Entwicklungsstand kommerziell bereits angebotener MT-Systeme auch eine Benutzersicht des jeweiligen Systems präsentiert (7 Systeme).
- **Technische Aussichten der MT**
Führende Fachwissenschaftler aus verschiedenen Ländern zogen in einer Podiumsdiskussion Bilanz zum Stand der Entwicklung der MT-Systeme und versuchten, die Stoßrichtung und Erfolgswahrscheinlichkeit der Weiterentwicklung aufzuzeigen.
- **Gegenwärtiger Stand und Perspektiven von MT-Projekten**
Es wurden die 7 derzeit wohl größten Forschungs- und Entwicklungsprojekte im MT-Bereich präsentiert.
- **MT aus Regierungssicht**
Vertreter der Regierungen von verschiedenen Ländern diskutierten die Perspektive zukünftiger MT-Projekte und MT-Anwendungen bezogen auf das jeweilige Land.

Parallel zu den Podiumsvorträgen fand eine Demonstration von kommerziellen MT-Systemen statt. Da die Demonstrationen nicht zusätzlich auch außerhalb der Redezeiten der Hauptveranstaltung angeboten wurden, kann in diesem Bericht eine Bewertung der Systeme aufgrund von persönlichen Erfahrungen damit nur marginal gegeben werden. Der Anhang zu diesem Bericht gibt dafür eine kurze technische Beschreibung der Systeme.

Im Folgenden wird eine Beschreibung der auf dem MT-Summit im Plenum vorgestellten kommerziellen Systeme und der Forschungs- und Entwicklungsprojekte inklusive des Versuchs einer Bewertung gegeben. Die Ergebnisse der Podiumsdiskussion zu den technischen Aussichten der MT sind hierin implizit eingearbeitet und werden nicht nochmals separat dargestellt. Der interessierte Leser sei zu diesem Punkt genau wie zu "MT aus Regierungssicht" auf die Veröffentlichung der Kongreßvorträge verwiesen.

Der hier vorliegende Aufsatz entstand im Rahmen der Tätigkeit des Projektträgers Fachinformation, der für das BMFT unter anderem die Projekte EUROTRA-D und deren Begleitforschung betreut.

A. Kommerzielle Systeme

I. SYSTRAN

I.1 Präsentation

Über Grundidee und Geschichte von SYSTRAN als eines der Pioniersysteme im MT-Bereich ist bereits soviel veröffentlicht (siehe z.B. F. Slocum: A Survey of Machine Translation: Its History, Current Status and Future Prospects. Computational Linguistics, Vol 11, No. 1, 1985), daß hier auf eine inhaltliche Darstellung verzichtet wird. In Hakone wurde SYSTRAN von Frau Ryan (Mitarbeiterin von P. Toma, Aorangi International University) präsentiert, die auf die folgenden Punkte abhob:

1. **Geschwindigkeit**
max. 2 Mio. Worte pro CPU-Stunde Produktivitätssteigerung: Faktor 4 gegenüber Humanübersetzung
2. **Kosten**
bei voller Post-Edition vergleichbar mit Humanübersetzung, Rohübersetzung erheblich billiger
3. **Genauigkeit**
- Nach Angabe des Kernforschungszentrums Karlsruhe: bei uneditiertem französisch-englischem Output 95% verstehbar, nur 1% absolut unverständliche Sätze.
- Über die umfangreichen Lexika hohe Konsistenz der Terminologie
4. **Verfügbarkeit**
- Service-Büros in Luxemburg, Frankreich, Belgien und Italien die von der Roh- bis hin zur voll posteditierten Übersetzung alle Qualitätsstufen anbieten.
- 800 on-line Terminals (U.S. Air Force)
- Potentiell 4,5 Mio Minitel-Benutzer in Frankreich
- Geplant: on-line Zugang in Japan
- 15 operationelle Sprachpaare
- 6 Sprachpaare als Pilotsysteme
- Lexika für zahlreiche Sachgebiete
- weites Spektrum der verarbeitbaren Textsorten
5. **Flexibilität**
Modularer Aufbau ohne Beschränkung der Gesamtsystemgröße. Interessant hier ein Modul "Multitarget dictionary" (eine Ausgangssprache, viele Zielsprachen).

SYSTRAN zeichnet sich sicherlich auch durch seine bereits jetzt für viele Sprachpaare sehr umfangreichen Lexika aus. Hierbei kann für ein Wort oder einen Ausdruck potentiell unbegrenzt viel Information abgespeichert werden. Die Lexika werden, genau wie die syntaktisch-semantischen Werkzeuge, immer weiter ausgebaut. So ist z.B. für das derzeit 50 000 Grundeinträge umfassende japanische Lexikon ein Ausbau auf 500 000 Begriffe innerhalb von 2 Jahren geplant. SYSTRAN verfügt momentan über 500 hierarchisch strukturierte se-

mantische Kategorien. Es sind eine ganze Reihe linguistischer Programme z.B. für die Lösung der Ambiguitäts-Probleme (Homograph-Routine, lexikalische Routine) integriert. Hierbei ist die Analyse-Komponente für die Ausgangssprache als größtes der linguistischen Programme unabhängig von der Zielsprache. Dies ist der Grund, warum SYSTRAN in der Lage ist, ein "multitarget dictionary" (aus dem Englischen in 5 Zielsprachen) zu adaptieren, wie dies 1983 für Xerox gemacht wurde.

Da nach wie vor SYSTRAN ein offensichtlich Lexikon-orientiertes System ist, wobei die Geschwindigkeit auch dadurch erreicht wird, daß bei Nicht-Finden des Wortes abgebrochen und das Wort unübersetzt wiedergegeben wird, ist das Beispiel der Übersetzung im Gebiet der Nukleartechnologie mit nur etwa einem nicht gefundenen Wort pro Seite interessant. Offensichtlich handelt es sich bei SYSTRAN um ein System, das wohl aus der Sicht der linguistischen Theorie wenig Eleganz aufweist, aber über die enormen Lexika die oben unter 1 bis 5 aufgezählten Vorteile hat.

1.2 SYSTRAN aus Benutzersicht

Als SYSTRAN-Benutzer trug Herr Akiyama von der Firma Arthur Andersen & Co vor. Bei dieser Buchführungsfirma besteht im Bereich Management Information Consulting ein großer Bedarf für die Übersetzung von EDV-Handbüchern aus dem Englischen ins Japanische im Zusammenhang mit EDV-Installationen im wirtschaftlichen Umfeld der Kunden von Arthur Andersen & Co. Es handelt sich um umfangreiche Texte inklusive Tabellen und Diagrammen, die bereits in maschinenlesbarer Form vorliegen. Die graphischen Teile werden zunächst separat abgespeichert. Die Texte werden dann von der Übersetzungsfirma übersetzt und mittels Textverarbeitung posteditiert. Nur diese posteditierten Texte werden weitergegeben. Dann wird die Übersetzung von Mitarbeitern der Firma Arthur Andersen & Co auf eigene Terminologie und Fachtermini hin nochmals durchgesehen. Dies ist de facto eine zweite Postedition. Danach werden Text und graphische Teile wieder zusammengebracht. Akiyama hebt hierbei positiv die Geschwindigkeit und terminologische Konsistenz von SYSTRAN hervor. Die Konsistenz wurde auch dadurch erreicht, daß in SYSTRAN ein speziell für diese Anwendung zugeschnittenes Wörterbuch aufgebaut wurde. Weiter wird über die Möglichkeit der Abspeicherung ganzer Sätze der Posteditationsprozeß sehr einfach. Die Übersetzung mittels SYSTRAN ist nicht billiger als die Humanübersetzung. Es fällt der Gegensatz auf, daß trotz der Aussage, mittels implementiertem speziellen Wörterbuch terminologische Konsistenz zu erreichen, eine zweite Postedition durchgeführt wird. Leider wurden keine Aussagen darüber gemacht, wie die graphischen Teile separiert und mit dem Text wieder zusammengefügt werden.

II. METAL

II.1 Präsentation

METAL wurde, ursprünglich von der U.S.-Regierung gefördert, an der University of Texas ab 1961 entwickelt. Es handelt sich vom Ansatz her um ein System, das von Chomskys transformationellen Paradigma ausgehend mit einer syntaktischen Interlingua auf der Tiefenstruktur arbeitet. METAL versteht sich aber trotzdem als System mit Transfer-Ansatz. Seit 1978 besteht ein Kooperationsvertrag mit Siemens, das seit 1980 alleiniger Sponsor ist. Im Einzelnen wurde METAL wie folgt von Herrn Schneider, Firma Siemens, dargestellt:

1. Geschwindigkeit

Ursprünglich in Fortran realisiert, wurde METAL inzwischen in LISP umgeschrieben. Durch einen "Bottom-up chartparser", der unbrauchbare Pfade in einer möglichen Baumstruktur durch präferentielle Gewichtung anhand lexikalischer und grammatikalischer Daten ausschließt, und eine "weiche" Behandlung bei fehlenden Daten wurde neben guter Grammatikalität erhebliche Beschleunigung der Übersetzung erreicht. Weitere Beschleunigung wird durch effizienten Aufbau der Grammatikregeln erreicht; dabei werden die Strukturen entsprechend der Häufigkeit ihres Auftretens indiziert. Die Produktivitätssteigerung gegenüber der Humanübersetzung wird mit einem Faktor 3 inklusive Postedition angegeben. Die maschinelle Übersetzungsgeschwindigkeit liegt bei 1 Wort/Sekunde (etwa 200 Seiten/Tag)

2. Kosten

METAL wird überwiegend in hausinternen Pilotinstallationen bei Siemens eingesetzt. Es werden keine Angaben über Kosten für fremde Benutzer gemacht.

3. Genauigkeit

METAL gibt ohne kriterielle Basis eine Korrektheitsrate von mehr als der Hälfte aller übersetzter Sätze an. Der Rest sei mindestens so gut, daß er als Grundlage für die Postedition verwendet werden kann. Diese Genauigkeit wird auf Grammatik-Seite über rekursiven Einsatz der indizierten Phrasenstruktur-Regeln erreicht. Auf Lexikon-Seite geschieht dies durch das hierarchisch modulare Lexikon.

4. Verfügbarkeit

METAL ist implementiert in Commonlisp und läuft auf einer Symbolics 36-Series LISP-Maschine. Verschiedene, die Verarbeitung unterstützende Programme laufen auf Siemens PC-MX 2-Maschinen. METAL hat bislang außer hausinternen nur wenig externe Pilotinstallationen, die aber so erfolgreich laufen, daß das System inzwischen kommerziell angeboten wird, allerdings nur für

das Sprachpaar Deutsch-Englisch. Diese kommerzielle Version wird geliefert mit einem Funktionswortschatz, dem allgemeinen Wortschatz und dem allgemeinen technischen Wortschatz. Insgesamt sind im Lexikon ca. 15.000 Einträge. Dazu muß gesagt werden, daß es sich, bedingt durch den speziellen Aufbau des METAL-Lexikons, nicht um Einzelworteinträge handelt, sondern z.B. strukturelle Informationen mitgespeichert werden. Weitere Sprachpaare sind geplant. Da METAL mit monolingualen Lexika und je einem Transferlexikon für jedes Sprachpaar arbeitet, gleichzeitig auch die Grammatikkomponente strikt modular aufgebaut ist, dürfte die Integration geplanter weiterer Sprachpaare problemlos sein.

Eine Besonderheit von METAL ist die automatische Erkennung von Text(-teilen) z.B. auch Überschriften, wenn auf der gleichen Seite auch noch Tabellen oder Graphik etc. vorhanden sind, so daß nach der Übersetzung eine Option auf Reformatierung analog der Quellvorlage angeboten wird.

5. Flexibilität

Über den hohen Grad der Modularität dürfte METAL bezüglich Flexibilität wenige Einschränkungen haben.

Das Hauptproblem, das in der Präsentation von METAL angesprochen wurde, war der Mangel einer theoretischen Basis für sprachunabhängige semantische Repräsentation. Bezüglich der Disambiguierung von Ausdrücken/Sätzen befindet man sich auch bei METAL im Stadium der reinen Forschung. Es gibt genausowenig eine theoretische Fundierung für die Behandlung von Ellipsen. Hier ist METAL für die maschinelle Übersetzung auf heuristische Verfahren angewiesen. Nicht zuletzt bleibt dann noch das Problem, daß sich die Eingabe nicht nach der Übersetzungsmaschine richtet. Wie aber sollen grammatikalisch unrichtige Sätze übersetzt werden? METAL versucht dies mit den "fail-soft routines".

II.2 METAL aus Benutzersicht

Als METAL-Benutzer trug Herr Shah von der Firma COMPULEX, Computer Intergraded Translations, Zürich vor. Diese Firma ist spezialisiert auf multilinguale technische Dokumentation. Es handelt sich um die erste externe Testinstallation von METAL seitens Siemens, die im Februar 1987 durchgeführt wurde. Installations- und Anlaufphase betragen 3 Monate. Der Benutzer hat inzwischen 300 Seiten eines größeren Auftrags (Handbücher und Testberichte einer neuen Produktreihe, die international vermarktet werden soll) mit METAL übersetzt und gibt an, die Qualität der Maschinenübersetzung könne ein hohes Niveau dann erreichen, wenn insbesondere das System-Lexikon auf das Fachgebiet "adäquat getunt" sei. Vorteile gegenüber konventioneller anderssprachiger Dokumenterstellung seien:

- Drastische Reduzierung der Zeit für die Produktion der (mehrsprachigen) Dokumentation
- signifikante Verbesserung der Konsistenz der Terminologie, dadurch verbesserte Funktionalität der Dokumentation
- Signifikante Senkung der Produktionskosten, insbesondere auch deshalb, weil in METAL die Seitenformatierung integriert ist. Damit fallen die bis zu 40 % hohen Kosten für zusätzliche Textverarbeitung bei Einschluß von Tabellen, Graphik, etc. in der Dokumentation weg. Die Gesamtkosten für die Erstellung der mehrsprachigen Dokumentation werden sogar mit mindestens 40 % niedriger angegeben.

Kritik an und Probleme mit METAL wurden nicht ausgesprochen.

III. D L T

DLT (Distributed Language Translation) fällt insofern aus der Reihe der vorgestellten Übersetzungsprogramme, als es sich im Status Quo um ein Forschungs- und Entwicklungsprojekt handelt, bei dem die Kommerzialisierung nicht vor 1992 geplant ist. DLT wird von der Firma BSO seit 1982 zunächst von der Kommission der Europäischen Gemeinschaft, ab 1984 von der Niederländischen Regierung zu 30 % unterstützt entwickelt. Es soll sich um ein System für Übersetzung mit so hoher Qualität handeln, daß die Postedition nicht mehr notwendig ist. Dafür scheint aber eine maschinell unterstützte Präedition notwendig. DLT verfolgt einen echten Interlingua-Ansatz mit Esperanto als Zwischensprache. Aus Sicht von Chomskys Paradigma ist Esperanto eine natürliche Sprache und somit auf struktureller Ebene mit den zu übersetzenden Sprachen auf demselben Level. Gleichzeitig aber ist Esperanto diejenige Sprache, die bezüglich Struktur und Transparenz am wenigsten von formalisierten Systemen entfernt ist. Trotz der Notwendigkeit der durch den Interlingua-Ansatz bedingten zweifachen Substitution, die ja immer als Hauptkritik an diesem Ansatz für Übersetzungssysteme geübt wird, verspricht sich der Entwickler über den extensiven Gebrauch von Esperanto als Zwischenebene einen hohen Grad an Disambiguierung auf semantischer Ebene und eine klare Aufspaltung (hier auch im Sinne einer erstrebenswerten Modularität zu verstehen) der Übersetzung in zwei getrennte Teilprozesse. Weitere Besonderheiten von DLT werden sein:

- Weltweite Multilingualität
- Integration von künstlicher Intelligenz
- Interaktivität während der Texteingabe
- Netzwerktechnik bei relativ kleinen Maschinen mit verteilten, aber interaktiv nutzbaren Speziallexika

Entsprechend dem Entwicklungsstand können kaum Angaben über Geschwindigkeit, Kosten, Genauigkeit, Verfügbarkeit und Flexibilität gemacht werden. Das größte Problem für DLT ist die Ambiguität bzw. Polysemie bei der Übersetzung. Als Lösungsansatz will man Weltwissen integrieren und zwar auf der Interlingua-Ebene in Esperanto. Man verspricht sich davon die Möglichkeit intensiven Patternmatchings, das von den klaren Morphem-Strukturen des Esperanto profitieren könnte. Bis jetzt allerdings verbirgt sich unter dem Begriff "Wissensbank" die Kombination von den 800 Lexikonoriginaleinträgen zu ca. 50.000 Wortkombinationen, die dann als "Basiseinheiten" zur Übersetzung verwendet werden. Hauptsprachpaar ist momentan Englisch-Französisch. Vorbereitende Studien werden gemacht für Finnisch, Ungarisch, Chinesisch und Japanisch. Eine Darstellung von DLT aus Benutzersicht konnte in Hakone natürlich noch nicht gegeben werden.

IV. ATLAS

IV.1 Präsentation

Fujitsu hat zwei Maschinenübersetzungssysteme. ATLAS I basiert auf Syntax, ATLAS II auf der Semantik. Herr Uchida präsentierte nur ATLAS II. Ausgehend von einem extrem simplifiziertem "world model" des Menschen, das nicht zuletzt vom linguistischen Wissen geformt sein soll, wird ATLAS II als in dem Sinn dem Menschen "nachgebildet" beschrieben, daß das System mit "world model" und "language model" ausgestattet sei. Das "world model" definiert jede mögliche Beziehung zwischen Begriffen. Die Qualität dieses Ansatzes einmal dahingestellt, löst ATLAS II die Übersetzungsaufgabe so, daß die Satzanalyse der Quellsprache mittels Lexikon, Analyseregeln und dem "world model" als Beziehungsgefüge von Bedeutungen über einen begrifflichen Struktur-Transfer geschieht. Dahinter steht ein Apparat von begrifflichen Struktur-Übertragungs-Regeln. Generiert wird dann wieder mit Lexikon und Generierungsregeln plus einem Sprachmodell, das von gleichzeitig auftretenden Begriffsbeziehungen ausgeht. Diese Abbildung der Quellsprache mit Bedeutungsrelationen hin zur Zielsprache mit bloßen Begriffsrelationen ist die direkte Ausprägung der oben kurz benannten schwachen Grundlagentheorie. Dies allerdings wird in der Präsentation ganz und gar nicht als Schwäche des Systems benannt. Über die greifbaren Systemeigenschaften wie Geschwindigkeit, Kosten, Genauigkeit und Flexibilität, schweigt sich der Redner aus.

Kommerziell verfügbar ist ein System Englisch-Japanisch. Andere Sprachpaare sind lediglich im Experimentierstadium (Japanisch-Koreanisch, Japanisch-Englisch, Japanisch-Französisch, Japanisch-Swahili, Japanisch-Innuit). In einem Kooperationsprojekt wurde mit ATLAS an der Universität Stuttgart auch ein Japanisch-Deutsch-System entwickelt.

Uchida stellte in einem allgemein gehaltenen Teil die Vor- und Nachteile maschineller und Humanübersetzung dar. Positiv kann bei der MT die Geschwindigkeit und die Konsistenz der Terminologie sein. Je größer allerdings die Menge des im System zu integrierendes "Weltwissen" ist, je weiter also die für alle Systeme postulierte Integration der Semantik getrieben wird, desto länger wird die Arbeitszeit. Großes Problem ist generell auch die Qualität der Übersetzung. Hier sollen unterstützende Funktionen für Präedition, Postedition und umfangreichere Wörterbücher Verbesserung bringen. Uchida dicit.

Anmerkung: *Einige interessante Merkmale von ATLAS sind im Anhang zu finden.*

IV.2 ATLAS II aus Benutzersicht

Leider sehr knapp gab Herr Saeki von der Firma Mazda einen Erfahrungsbericht im Umgang mit ATLAS II. Nachdem bei der Übersetzung von Servicehandbüchern für Autos und Servicetrainingshandbüchern nur 40 % der Übersetzung durch ATLAS richtig oder wenigstens korrigierbar war, hat Mazda zusammen mit dem Entwickler Fujitsu u.a. die japanische Analysegrammatik, die semantischen Verarbeitungsfunktionen und die Wörterbücher überarbeitet. Nun seien 80 % der Übersetzung korrekt. Herr Saeki gab eine Art Checkliste für die Einführung eines MT-Systems, die meines Erachtens die realen MT-Probleme sehr gut beleuchtet:

1. Sachgebietsauswahl, die passend ist für ein MT-System
2. Präzise Anpassung des Systems auf das Arbeitsgebiet
3. Systematische Wartung der Wörterbücher und Einbau eines Kontrollsystems
4. Standardisierung und Reglementierung bei der Erstellung von Dokumenten und Aufbau eines Präeditionssystems
5. Anpassung und Integration existierender Textverarbeitungssysteme
6. Einbau von Trainingsprogrammen und Aufbau eines Vertriebssystems für die MT-Systeme bei den betroffenen Abteilungen.

Saekis Liste ist hier in Übersetzung fast wörtlich wiedergegeben.

V. HICATS/JE

V. 1 Präsentation

HICATS/JE (Hitachi Computer Aided Translation System/Japanese to English) wurde von Herrn Kaji von dem Systems Development Laboratory (Hitachi) präsentiert. Eine Besonderheit des Systems liegt darin, daß für die Analyse eine andere Grammatik ("Dependency Grammar") als für die Generierung (Phrasenstruktur-Grammatik) verwendet wird. Die dependency grammar erzeugt

einen semantischen Abhängigkeitsgraphen (Knoten mit indizierten Verzweigungen), mit dem die Begriffe und semantische Zusammenhänge intermediär repräsentiert für den Transfer zur Verfügung stehen. Der Abhängigkeitsgraph ist hierbei nahezu sprachunabhängig. Auf diese Weise löst HICATS das Problem der Übersetzung zweier strukturell verschiedener Sprachen. Von Kaji angesprochene Probleme des Japanischen für die Analysekomponente sind

- flexible Wortanordnung
- Bedeutungsvarianz von Begriffen durch postpositionale Wörter
- Auslassbarkeit ganzer Phrasen

Für die Generierung des Englischen ist vor allem die richtige Auswahl bei äquivalenten Begriffen problematisch. Weiteres Problem ist, daß die maschinell erfaßte Grammatik niemals vollständig sein kann und damit genau wie die Lexika permanente Wartung braucht. Da das System im jetzigen Zustand nicht beliebige Sätze übersetzen kann, das Ergebnis aber bei präeditierten Sätzen ganz passabel ist, wird der Vorschlag einer entsprechend restringierten Teilsprache gemacht. Über Geschwindigkeit, Kosten und Genauigkeit wurden in der Präsentation keine Angaben gemacht. HICATS/JE ist seit Mai 1986 auf dem Markt. Es läuft auf HICAT M series Computern unter dem Betriebssystem VOS3. Einsatzgebiet ist der wissenschaftliche und technische Bereich (Produkt-handbücher, Forschungsberichte, Patentabstrakte). Die Präsentation ging überwiegend auf realsprachliche Übersetzungsbeispiele ein. Der zugehörige theoretische Teil war auffällig unscharf in seiner Begrifflichkeit, bzw. etliche im ganzen Forschungsgebiet MT gebräuchliche Begriffe wurden verwendet, aber in anderer Bedeutung.

Anmerkung: *Weitere Merkmale werden im Anhang gegeben.*

V.2 HICATS/JE aus Benutzersicht

Die JAPIO (Japan Patent Information Organisation), vertreten durch Herrn Oshio, bietet als Informationseinrichtung Retrievendienste über Patente, Gebrauchsmuster, Design und Handelsmarken. Auf Anfrage durch das Japanische Patentamt übersetzt JAPIO Abstracts von veröffentlichten ungeprüften Patentschriften aus dem Japanischen ins Englische. Um der gewaltigen Menge Herr zu werden und den gesamten Übersetzungsprozeß zu beschleunigen, will die JAPIO ein maschinelles Übersetzungssystem einführen. Entgegen der Ankündigung, über den konkreten Testeinsatz von HICATS/JE zu berichten, wurde nur über allgemeine Probleme der Übersetzung bei JAPIO und spezielle Probleme der Textkorpora der Abstracts vorgetragen. Weiter wurde ein knapper Abriss über jetziges Vorgehen bei der Übersetzung versus MT und spezielle Anforderungen an eine Übersetzungsmaschine gegeben. Als Schlußfolgerung wurde ein Appell an

die MT-Entwickler ausgesprochen, für mehr Genauigkeit und höhere Funktionalität ihrer Systeme zu sorgen. Die JAPIO im Gegenzug verspricht passende Lexika. Es wurde noch nicht einmal klar, ob HICATS/JE überhaupt zu Testzwecken jemals eingesetzt wird.

VI. TAURAS

VI. 1 Präsentation

Taurus (Toshiba Automatic Translation System Reinforced by Semantics) wurde von Herrn Amano vom Forschungs- und Entwicklungslabor bei Toshiba präsentiert. Das System soll multilingual angelegt sein. Es enthält neben einem zweisprachigen Editor (Deutsch-Englisch) eine ganze Reihe von Software-Beiprogrammen (z.B. zweisprachige Textverarbeitung). Das System wird entwickelt zur Übersetzung technischer Texte. Das Besondere ist ein Grammatiksystem für die Übersetzung, wo syntaktische und semantische Analyse logisch völlig getrennt werden. Die Syntaktische Analyse mittels ATNG (Augmented Transition Network Grammar) liefert eine unique syntaktische Struktur ohne semantische Merkmale. Mit der Annahme, Bedeutung sei immer lexikalisch, werden im semantischen Schritt der Analyse für die rein syntaktische Struktur mit ihren kategorialen Strukturelementen verschiedene begriffliche Interpretationen durchgeführt (Pattern matching). Das heißt, semantische Regeln werden mit Begriffen im Lexikon über lexikalische Regeln verbunden. Ist die Analyse nach diesem 2. Schritt nicht erfolgreich, wird der zu analysierende Satz an die syntaktische Analyse zurückgewiesen. Bei erfolgreicher Annahme entsteht als Ergebnis der Analyse eine begriffliche Struktur, die dem Transfer zur Verfügung steht. In diesem Sinn sind in TAURAS in der Analysekomponente Syntax und Semantik unabhängig, aber interaktiv. Die Generierung, die trotz multilingualem Ansatz hier nur für das Japanische dargestellt wird, nimmt für morphologische Generierung die im Lexikon enthaltene Information und verbindet mit den Informationen, die aus der begrifflichen Struktur zur Verfügung stehen (also z.B. Modalität, Zeit, Typ). Für das Japanische ist dann noch das sprachliche Sonderproblem der postpositionierten Wörter (Joshi) zu lösen.

Auch diese Präsentation brachte keine Angaben über Geschwindigkeit, Kosten, Genauigkeit oder Flexibilität. Geplant ist einerseits ein multilinguales Übersetzungssystem, das aber andererseits im Status quo ausschließlich Englisch-Japanisch als Sprachpaar enthält. TAURAS ist in C mit Betriebssystem UNIX implementiert. Es ist lauffähig auf Toshiba Minicomputer AS 3000 (CPU MC 68020). Damit ist wenigstens ein hoher Grad an Portabilität gewährleistet.

VI.2 TAURAS aus Benutzersicht

Nach Darstellung von Herrn Yamaoka von der Übersetzungsfirma I.S.S. wird dort seit Juni 1986 ein CAT-System für Englisch-Japanisch von der Firma Toshiba eingesetzt. Nachdem TAURAS aber als MT-System präsentiert wurde, ist nicht klar, ob an dieser Stelle überhaupt von demselben System die Rede war. I.S.S. beabsichtigt mit der Einführung eines CAT-Systems folgendes:

- Standardisierung der Terminologie
- Benutzerspezifische Anpassung der Terminologie
- Leistungssteigerung des Humanübersetzers durch Computerunterstützung
- Computerisierung des Lexikons mit mind. 10 Mio. Begriffen.

Tatsächliche Erfahrungen mit dem Toshiba-System werden nicht geschildert. Merkwürdig, daß Toshiba hier einen Benutzer präsentiert, der wahrscheinlich gar nicht TAURAS verwendet und so wie das System offiziell dargestellt wird, durch den von I.S.S. so hoch angegebenen Bedarf an Lexikoneinträgen, gar nicht verwenden kann.

VII. PIVOT

VII.1 Präsentation

Das Zwei-Phasen-Übersetzungssystem PIVOT der Firma NEC wurde von Herrn Muraki präsentiert. Es handelt sich um einen typischen Interlingua-Ansatz, der momentan für Japanisch-Englisch und Englisch-Japanisch realisiert, aber wegen des Ansatzes erweiterbar ist. Integriert sind etliche interaktive Tools:

- Textverarbeitung (zweisprachig)
- Lexikonentwicklung und -Wartung (zweisprachig)
- Wörterbuch und Textausdruck
- Automatische Suche unbekannter Wörter

Die Interlingua enthält neben strukturellen Informationen auch solche über Knoten, Pragmatik, Bereich und Paraphrasenpostulate für sogenannte begriffliche Primitive. Damit sind eine ganze Reihe allgemeiner (z.B. fehlende Kasus-Struktur im japanischen) Probleme lösbar. Für die Weiterentwicklung will man für PIVOT Folgendes verbessern:

- Ausführlicher und besser definierte semantische Definitionen der sog. begrifflichen Primitive
- Verbesserung der Codierung von pragmatischen Informationen, um die Selektivität der begrifflichen Primitive sowohl bei der Analyse als auch für die begriffliche Oberflächenrealisierung bei der Generierung zu verbessern
- Logisch besser definierte Interaktion zwischen Wörterbuch und grammatischen Regeln

Dies geschieht auch im Hinblick auf die Multilingualität. Es wurden auch hier keinerlei Angaben über Geschwindigkeit, Kosten und Genauigkeit gemacht.

VII.2 PIVOT aus Benutzersicht

Als PIVOT-Benutzer trug Herr Arai von den Japan Convention Services (JCS) vor. Die JCS sind im Feld der internationalen Kommunikation von der Übersetzung über Dokumentation bis hin zur Nachrichtenstreuung über das Fernsehnetz tätig. Er schilderte, daß es von großem Nutzen sei, nicht als Endanwender ein fertiges System zu kaufen, sondern als Benutzer-Entwickler bereits in den Entstehungsprozeß involviert zu sein. Fertige Systeme haben bereits festgelegte Lexika und sind dadurch nur noch für bestimmte Sachgebiete einsetzbar. Die Richtigkeit dieser Aussage sei dahingestellt. Es wurde wiederum nichts über Erfahrungen in diesem Fall mit PIVOT ausgesagt.

B. Forschungs- und Entwicklungsprojekte im Bereich maschineller Übersetzung: Status und Perspektiven

I. EUROTRA

Bedingt durch den politischen Aufbau der Europäischen Gemeinschaft besteht die Notwendigkeit, in den neun offiziellen Amtssprachen und dadurch in 72 möglichen Sprachpaaren zu kommunizieren. Die EG ist gezwungen, den größten Übersetzungsservice der Welt zu unterhalten. So hatte man sich nach einigen Präliminarien entschlossen, endgültig ab 1982 dezentral ein eigenes Übersetzungssystem in 3 Phasen zu entwickeln, das hier von Sergei Perschke präsentiert wurde. Die Fertigstellung des präindustriellen Prototyps ist 1990 vorgesehen. Das Projekt hat ein zentrales Management und Koordinierung und prinzipiell jeweils eine nationale Entwicklungsgruppe, die verantwortlich ist für die Analyse- und Syntheseteile der eigenen Sprache und für den Transfer aus der je anderen Sprache in die eigene. Bewertung und Planung des Projekts wird gemeinsam von den jeweiligen Projektleitern in etwa monatlich stattfindenden Sitzungen vorgenommen.

Eurotra ist ein Transfer-basiertes System, bei dem über eine Reihe intermediärer Repräsentationen ein Paar von Interface-Strukturen für Quell- und Zielsprache erzeugt wird. Analyse und Generierung sind unifikationsbasiert und momentan in PROLOG implementiert. Es ist geplant, kritische Teile in C zu reimplementieren. Um den Transfer von dem Lexikon als größte Systemkomponente zu entlasten, sind sprachunabhängige interlinguale Begriffe mit Entsprechungen in den Einzelsprachen eingebaut. Als semantische Wissensbasis wird bloßes begriffliches Allgemeinwissen verwendet, dem keine sachgebietspezifische Basis zugrunde liegt. Man spricht von der Notwendigkeit der Integration von KI-Wissen (insbesondere für die Disambiguierung), ohne eine klare Vorstellung zu haben.

Trotz der durch die verschiedenen intermediären Repräsentationsebenen bedingten theoretischen

Notwendigkeit verschiedener monolingualer Wörterbücher neben dem eigentlichen bilingualen Übersetzungswörterbuch wurde eine Sicht auf diese Wörterbücher kreiert, die das Wörterbuch als unique und monolingual erscheinen läßt. Dies wird erreicht durch einen prinzipiell identischen Abbildungsmechanismus von einer Repräsentationsebene auf die andere.

Entsprechend der Janusköpfigkeit der Eurotra-Entwicklung, einerseits Systementwicklungs- und Wartungswerkzeug zu sein, andererseits auch schon das Übersetzungsprogramm selbst, hat man, um beiden Ansprüchen gerecht zu werden, eine Systemarchitektur entwickelt, wobei zwischen dem Benutzer und dem Regelwerk ein (kommerziell verfügbares) relationales Datenbankmanagementsystem (Unify mit SQL) zwischengeschaltet ist. Es ist geplant, bis zum Projektende ein Wörterbuch mit 20.000 Begriffen aus einem eingeschränkten Sachgebiet zu implementieren. Mit Verweis darauf, daß im Status quo Eurotra etwa in der Mitte seiner Laufzeit ist, werde es für nicht möglich erklärt, Aussagen über die zukünftige Ausstattung, Geschwindigkeit und weitere Möglichkeiten des Systems zu machen. Damit konnte man auch über eventuelle Entwicklungsschwierigkeiten schweigen.

II. ALVEY

ALVEY ist ein von der Britischen Regierung unter Industriebeteiligung gefördertes 5-Jahres-Programm für Forschung und Entwicklung von Informationstechnologien. Die vier Hauptgebiete des Programms sind:

- Software Engineering
- Man-Machine Interfaces
- VLSI (Very large-scale integration techniques)
- IKBS (Intelligent knowledge-based systems)

Innerhalb IKBS wird u.a. über NLP (natural language processing) in zehn Einzelprojekten gearbeitet. Hutchinson (Univ. of East Anglia, UK) trug ein Papier von Knowles (Aston Univ., Birmingham, UK) über das Projekt "Read and write Japanese without knowing it" vor, das aufgespalten in AIDTRANS (Sheffield) und NTRAN (Manchester) unter Knowles' (Birmingham) Monitoring durchgeführt wird. Dieses seit 1984 laufende 3-Jahres-Projekt hat alleine im letzten Jahr drei mal das Projektmanagement gewechselt. Die beteiligte Industriefirma hofft, mit den bis 10.87 fertiggestellten zwei Prototypen eine Überführung in vermarktete Produkte machen zu können.

II.1 Prototyp AIDTRANS

AIDTRANS ist ein Japanisch-Englisches M(A)T-System, das abgeleitet von einem bereits für andere Zwecke vorhandenen Analysesystem für japanische

Grammatik als Kern ein Integriertes Wörterbuch-System enthält. Grundidee ist, soviel als möglich Grammatik und Heuristik in das Wörterbuch miteinzutragen. Es wird für die Übersetzung eine lineare prädiktive Analyse durchgeführt mit einer einfachen links-nach-rechts Erfassung der Eingabe. Dieser Ansatz ist sicher veraltet. Außer einer Angabe von 200 verschiedenen Typen von Verknüpfungen wird keine Angabe über das Wörterbuch gemacht. Weitere Angaben über Ausstattung, Geschwindigkeit und weitere Möglichkeiten des Systems fehlen. Das System scheint in seinem Entwicklungsstand so zu sein, daß über bei der Übersetzung auftretende Probleme noch nicht einmal Aussagen möglich sind.

II.2 Prototyp NTRAN

Das System für Englisch-Japanisch besteht aus vier Teilen:

1. **Englisches Analysemodul**
basierend auf Lexical-Functional Grammar (LFG). Die F-Strukturen dienen als intermediäre Repräsentationen, die die (S-) Oberflächen-Konstituenten-Strukturen in sprachunabhängige Strukturen überführen. Dazu wird ein bottom-up Parser verwendet, der es dem top-down Interpreter in Prolog erlaubt, die kompilierten Grammatikregeln in bottom-up-Manier zu interpretieren. Der Prolog-Interpreter mit seinen Unifikationsalgorithmen überprüft die Wohlgeformtheit der F-Strukturen.
2. **Transfer**
Aus den lexikalischen Eintragungen in das monolinguale Wörterbuch kann über dort enthaltene Informationen für den Transfer mit seinem bilingualen Wörterbuch eine große Menge Redundanz vermieden werden. Das bilinguale Wörterbuch ersetzt dann die Werte der englischen S-Strukturen mit den semantischen äquivalenten im Japanischen. In einem 2. Transferschritt werden dann die attributiven Wertepaare übersetzt, die keine semantischen Werte haben. Dann werden bei Bedarf weitere kontrastive Informationen an die in diesem Stadium entstandenen S-Strukturen an die Ausgabekette für die Zielsprache angehängt, um aus der S-Struktur die japanische F-Struktur produzieren zu können.
3. **Japanische Generierung**
An die japanische F-Struktur wird weitere Information über Subkategorisierung angehängt und als Konsequenz daraus der entsprechende grammatikalische Partikel mit der betreffenden japanischen Lexikoneinheit verbunden.
4. **Mono- und bilinguale Wörterbücher**
Über die Angaben hinaus, daß für die englische Analyse ein Mono- und für den Transfer ein bilinguales Wörterbuch eingesetzt wird,

werden keine Angaben über Aufbau und Umfang dieser Wörterbücher gemacht.

An Problemen werden ohne jede Detaillierung genannt: Robustheit, Disambiguierung und Decodierungsgrad. Es handelt sich hier um ein offensichtlich State-of-the-art Teilprojekt, so daß ein merkwürdiger Kontrast zu dem Schwesterprojekt AIDTRANS entsteht.

II.3 Anhang: "The Englishmen's Keyboard"

Im Zusammenhang mit AIDTRANS wurde ein System entwickelt, das den für japanisch Nicht-Schriftkundigen in die Lage versetzt, nicht indirekt über äquivalente des Japanischen in lateinischer Schrift, sondern direkt über eine Auswahl von auf dem Bildschirm angebotenen häufigsten (Teil-) Schriftzeichen mit der "Maus" auszuwählen. Hier wird allerdings eine Grundkenntnis von Kanji vorausgesetzt.

Als Resume wird die vollautomatische maschinelle Übersetzung in absehbarer Zukunft als unmöglich bezeichnet. Maschinelle Hilfen für die Übersetzung haben neben der Realisierbarkeit den großen Vorteil der Geschwindigkeit. Für ein MT-System müssen Programmiersprachen und Technik, linguistische Theorie und formale Codierung von komplexem Wissen weiterentwickelt werden. So könnte man es schaffen, höhere Geschwindigkeit, terminologische Konsistenz, Grammatikalität und eine gute Lösung von Ambiguitätsproblemen zu erreichen.

III. ODA

ODA ist ein Forschungs- und Entwicklungsprojekt für maschinelle Übersetzung in Zusammenarbeit von Japan und den benachbarten Ländern China, Thailand, Malaysia und Indonesien. Das Projekt beginnt 1987 und hat eine geplante Laufzeit von 6 Jahren. Es wird versucht, für jedes Land zunächst eine Forschungsgruppe von je 50 Wissenschaftlern zu etablieren. Als Sachgebiet der Übersetzung sieht man hauptsächlich industrielle und technische Information. Übersetzungssprachen werden zunächst alle Sprachpaare mit Japanisch und der jeweiligen Sprache des Nachbarlandes sein. Die Sprachpaare zwischen den Nachbarländern sollen erst später behandelt werden.

Man geht von vornherein von einer Übersetzungsgenauigkeit von 80 bis 90% (inkl. Praediction) aus. ODA soll ein Multi-Interlingua-MT-System werden, bei dem neben Prä- und Post- auch Interim-Edition vorgesehen ist. Das Lexikon soll 50.000 Basisbegriffe und 25.000 Begriffe eines Spezialwörterbuchs enthalten. So die Planung.

Interessant ist dieses Projekt deswegen, weil es als flankierende Maßnahme zu einem parallelen wirtschaftlichen Zusammenschluß zwischen Japan und den genannten Nachbarländern zu sehen ist.

Möglicherweise ist der gesamte MT Summit unter dieser Perspektive einer erweiterten wirtschaftlichen Initiative Japans zu verstehen.

IV. GETA

GETA ist nicht eigentlicher Name eines Projektes, sondern Name einer Studiengruppe für maschinelles Übersetzen. Es handelt sich um eine 1971 gegründete Forschungsgruppe der USTMG (Scientific, Technological and Medical University of Grenoble) verbunden mit dem CNRS (Centre for National Research). Die Hauptmerkmale gegenwärtiger Entwicklungen bei GETA sind für MT-Systeme der 2. Generation:

- Transfer-Ansatz
- heuristische (und prozedurale) Techniken für linguistische Programmierung
- "Fail-soft"-Techniken bei der Behandlung der Übersetzungseinheiten
- Als Übersetzungseinheit werden ganze Sätze genommen.

Als Nahziel wird ein CAT-System angestrebt. In der Präsentation von C. Boitet wird eine ganze Reihe überwiegend experimenteller Aktivitäten bei GETA angesprochen, wo mit verschiedenen Sprachpaaren (das Ursprungssystem behandelte Russisch-Französisch), Maschinen unterschiedlicher Größe, verschiedenen Programmiersprachen, verschiedenen zugrundeliegenden DBMS etc. z.T. mit beachtlichen Teilergebnissen gearbeitet wird. Prinzipiell ist aber kein Ergebnis im Sinn eines lauffähigen MT-Systems erkennbar. Was Boitets Vortrag interessant macht, ist seine Liste von Forschungsschwerpunkten, um von Systemen der 2. zu solchen der 3. und 4. Generation weiterzukommen. Dies sind:

- Korrektursysteme, die "problem patterns" bei Source Interface Strukturen erkennen, um sie mittels der Wissensbank entsprechend modifizieren zu können.
- Repräsentation von "common sense"-Wissen in konsistenter Form für die Wissensbank.
- Interaktiver Ansatz bei der Übersetzung, weil trotz Wissensbank niemals 100% korrekte Analyse entstehen kann, da der zu übersetzende Text selbst normalerweise neues Wissen einführt.
- Normierte Eingabe, um wenigstens die Ambiguitäten bei der Analyse ausschließen zu können.
- Einführung von Metaregeln, um die Ambiguitätsprobleme mittels deklarativer Regeln von separaten Mechanismen lösen zu können.

Diese Punkte machen einerseits einige Probleme der heutigen MT-Systeme deutlich, andererseits aber sieht man, auch an ihrem heterogenen Charakter, daß sich selbst mit der sogenannten "4. Generation" lange noch keine endgültige Lösung für ein arbeitsfähiges MT-System abzeichnet.

V. MU**V.1 MU I**

Im MU-Projekt wurde zwischen 1982 und 1986, gefördert von der japanischen Regierung und unterstützt von STA (Science and Technology Agency), ein Prototyp-System als Machbarkeitsstudie für ein MT-System entwickelt. Es ist ein Transfer-basiertes System mit einem neutralen Lexikon, bei dem aus einer deklarativen Form das Analyse- und Synthesellexikon abgeleitet wird. Das Lexikon enthält momentan ca. 80.000 Einträge. Das System ist in der Lage, für die Sprachpaare J-E und E-J im Bereich des elektronischen Engineering für Abstracts von wissenschaftlichen und technologischen Papieren verstehbare Rohübersetzungen zu liefern. MU ist implementiert in einem LISP-Dialekt und läuft auf der Symbolics 3600 Serie. Die Verarbeitung findet nach heuristischen Prinzipien statt. Für die Grammatikbehandlung wurde eine spezielle, ebenfalls in LISP implementierte Sprache GRADE für flexible Baumstrukturbehandlung, Backtracking etc. entwickelt und implementiert. Ein Unterstützungssystem (Text- und Lexikoneditor, Wörterbuch-DBMS) ist integriert.

V. 2 MU II

Da einerseits das MU-System bezüglich der Machbarkeit als Erfolg zu werten ist, sich andererseits einige Defekte für den praktischen Betrieb zeigten, wurde MU als MU I deklariert und ein weiteres Vier-Jahres-Projekt (MU II) begonnen, in dem ein auf andere wissenschaftliche und technische Sachgebiete ausgedehntes MT-System zum Einsatz für Abstract-Übersetzung bei JICST (Japan Information Center for Science and Technology) entwickelt wird. Der Scope der Weiterentwicklung zeigt sehr gut die generellen Probleme von MT-Systemen:

GRADE wird reimplementiert in C, um die Geschwindigkeit zu erhöhen und die Arbeitsspeicherbelegung zu verringern. Außerdem werden einige mächtige Möglichkeiten, die von den Grammatik-Schreibern selten genutzt werden, aus GRADE wieder herangenommen, um das Verarbeitungsverhalten bezüglich Geschwindigkeit und Speicherplatzbelegung weiter zu verbessern. Das Lexikon wird auf ca. 300.000 Einträge erweitert. Für die Übersetzung wird bei der Eingabe eine nur eingeschränkte Sprache zugelassen, ohne den Aufwand für Präedition allzu hoch zu schrauben. Die Mensch-Maschine-Schnittstelle wird mit einem Übersetzungsorientierten Editor verbessert. Sachgebietsspezifische und benutzerorientierte Wörterbücher sollen den Deckungsgrad erhöhen. Wartungswerkzeuge für die Wörterbücher dienen hier als zusätzlich Unterstützung.

Probleme, die voraussichtlich auch nach Beendigung des MU II-Projektes bleiben, sind u.a. angemessene Ziellexikonauswahl und Verarbeitung von

den im Japanischen häufigen Ellipsen. Hierzu müßte das Umfeld der semantischen und kontextuellen Verarbeitung integriert werden.

Nach Auskunft von Jun-ichi Tsujii, der das MU-System präsentierte, findet die jetzige Weiterentwicklung bei JICST in Tokyo mit der fachlichen Beratung durch ihn, Nagao und Nakamura (Software) statt. Diese Kyoto-Gruppe beschäftigt sich inzwischen bereits mit Grundlagenprobleme im MT-Bereich.

VI CMU

Carbonell stellte unter diesem Titel eine der Aktivitäten der Carnegie Mellon University (CMU) vor, deren er zusammen mit Tomita nachgeht. Der gleiche Vortrag wurde auch auf der IJCAI 1987 in Mailand gehalten und in *"Natural Language"* im gleichen Jahre auf Seite 718-721 unter dem Titel *"The Universal Parser Architecture for Knowledge-Based Machine Translation"* veröffentlicht. Da die inhaltlichen Aspekte betreffend des Parsers in dem Vortrag und in der Veröffentlichung identisch wiedergegeben sind, möge man sich dort über die Details informieren.

Bei der CMU sind derzeit 30 Forscher in folgenden Projekten tätig:

- MT-Factory (Universeller Parser und Generator, multilingual, multi-domain)
- KI-Workbench
- Text-Analysis
- MT gesprochener Sprache

Generell wird ein semantischer Interlingua-Ansatz verfolgt, bei dem ein frameorientiertes semantisches System und ein wissensbasiertes Expertensystem eingesetzt werden soll. Die Philosophie ist es, bei einem absolut maximalen Anspruch ein System zu bauen, das funktioniert. Was eben den Japanern Schwierigkeiten mit dem Zeitbegriff macht (aus unserer Sicht!), ist den Amerikanern ihr "maximalist approach", geboren aus der Eigentümlichkeit der Verquickung von Pioniergeist und der "blissful simplicity" des Puritanismus. Auch dies aus mitteleuropäischer Sicht.

VII EDR

EDR steht als Abkürzung für Japan Electronic Dictionary Research Institute. Der von Kakizaki angekündigte Vortrag wurde von UCHIDA (Fujitsu Laboratories Ltd) gehalten. Von der generellen Linie, ein elektronisches Wörterbuch für Japanisch und Englisch zu entwickeln, einmal abgesehen, läßt sich im Hinblick auf MT-Systeme nebst ihren Entwicklungsschwierigkeiten nichts Fruchtbare sagen. Zum Abschluß soll der Leser aber mit 2 Zitaten aus Kakizakis Manuskript erfreut werden: *"In a compound word, how the concepts expressed by the compound word is expressed by concepts of component words"* (MT-Summit Programmheft S. 61). Das ist nicht

als Testsatz für ein MT-System gedacht, wäre als solcher allerdings kaum schlagbar. "Word dictionary is divided into two dictionaries. Basic dictionary and terminology dictionary. And Japanese and English word are the subjects the word dictionary, there are 8 word dictionaries" (Programmheft, S. 62). Die Ernsthaftigkeit von EDR stehe völlig außer Frage. Mit ihrer Präsentation beim MT-Summit hatte es einen schwarzen Tag.

C. Schluß

Der MT-SUMMIT ist weniger eine Veranstaltung gewesen, die neue Perspektiven oder Erkenntnisse gebracht hat, als eher eine Veranstaltung, bei der eine Bilanz gezogen wurde. Nach wie vor hört man, mit immer mächtigeren Maschinen die Laufzeit- und Hauptspeicherprobleme lösen zu können. Der Konflikt zwischen Interlingua- und Transferansatz spaltet die MT-Gemeinde in zwei Lager. Die sprachlichen Probleme wie Disambiguierung, Ellipsen, Anaphora etc. harren noch immer einer

tatsächlichen Behandlung. Analog des Glaubens an die magische Potenz der größeren Maschine geistert der Glaube der softwareseitigen Lösungen über KI-Ansätze durch die Forschungslandschaft. Die Analogie ist allerdings kontradiktorisch. So hätte es Kay leicht haben müssen mit seinen Warnrufen vor einem zweiten ALPAC-Report. Der obigen negativen Bilanz stehen aber eine ganze Reihe von Fortschritten entgegen. Dies zeigt sich z.B. darin, daß es inzwischen State of the Art ist, die einzelnen Übersetzungsschritte sauber in Sinn eines modularen Aufbaus zu trennen, möglichst spezielle Sprachen für linguistische Programmierungen zu nehmen, besondere Methoden, um das Ambiguitätsproblem trotz der Nicht-Lösbarkeit zu neutralisieren und nicht zuletzt entsprechend maschineller Verarbeitbarkeit strukturierte mono- und bilinguale Wörterbücher aufzubauen. Allgemein ausgedrückt: es sind eine ganze Reihe von Methoden der Fehlerbehandlung entwickelt worden. Die Zustandsbeschreibung der vorhandenen kommerziellen

ANHANG

Maschinelle Übersetzungssysteme

Name	LAMB	ATLAS I	ATLAS II	HICATS/JE	HICATS/EJ
Hersteller/ Entwickler	Cannon	Fujitsu	Fujitsu	HITACHI	HITACHI
Hauptmerkmal des Systems	wissensbasiertes Übersetzungs- system	Syntax- gesteuert	Transfer	Transfer	Syntaktischer Transfer
Übersetzte Sprachen	J - E	E - J	J - E	J - E	E - J
Übersetzungs- strategie	Phrasenstruktur- Grammatik Baum-Baum- Transformierung	Phrasen- struktur- grammatik	Dependency Grammar	Dependency Grammar Phrasenstruktur- turg.	Phrasenstruktur- grammatik Fallphrasen- struktur.
Wörter- bücher	J.-Analyse-W. Transfer-W. E.-Generierungs.-W. Sachgebiets-W.	E-J.-W. Basis 53000 Wörter Techn. Terminologie 25000 Wörter	J.-W./E.-W. Basiswörter- buch(50000 W) Techn. Terminologie 250000 Wörter	Basis-W. 50000 Wörter Techn. Terminologie 250000 W.	Basis-W. 50000 Wörter Techn. Terminologie 250000 Wörter
Ausstattung 1. Implementiert 2. auf OS 3. mit CPU (/Hardware)	Common Lisp Symbolics OS 61 Symbolics 3620	Assembler versch. OSIV FACOM M, FACOM S-3000	C OSIV/F4/MSP FACOM M	GDL u.PL/I VOS 3 HITAC M	GDL und PL/I VOS 3 HITAC M
Geschwin- digkeit	1000 w/h	60.000 w/h	60.000 w/h	20- 60.000 w/h	30-60.000 w/h
zusätzliche Merkmale	Wörterbuch- und Grammatik-Entwick- lungsumgebung Testtools	Bilingualer Editor Wörterbuch- Editor	Bilingualer Editor Wörterbuch- Editor	Wahlmöglich- keit -Satzstruktur -Alternatives Wort	

TAGUNGSBERICHTE

Übersetzungssysteme genauso wie Status- und Perspektivenbeschreibung der laufenden Projekte und Aktivitäten läßt kaum zwingende Empfehlungen für die weitere Forschungsstrategie zu. Im Status quo scheint es notwendig, vor allem im linguistischen Bereich erneute Anstrengungen in der Grundlagenforschung zu unternehmen. Nur so kann der Weg einer trivialen Wort-zu-Wort-Übersetzung über isolierte Satzübersetzung hin zu der notwendigen satzübergreifenden Analyse weitergegangen werden. Genausowenig dürfen die zu übersetzenden Sätze von ihrem einzelsprachlichen und somit letztendlich

kulturellen Hintergrund isoliert betrachtet werden. Hier ist viel Arbeit im Bereich kontrastiver Linguistik zu leisten. Die anwendungsorientierte Entwicklung jedenfalls scheint im Status quo nicht genug Substanz aus der Forschung ziehen zu können, um den fast im embryonalen Zustand befindlichen MT-Systemen glücklich auf die Welt verhelfen zu können. Es ist zu hoffen, daß die zum Abschluß durch den Leiter des Summits, Makoto Nagao, konstatierte Aufbruchstimmung zur erfolgreichen Weiterentwicklung sich bei dem für 1989 geplanten 2. MT-Summit in Deutschland zumindest ansatzweise verifizieren läßt.

Name	PAROLE	MU	MELTRAN-JE	PIVOT	TAURAS
Hersteller/ Entwickler	Wielers Research	Nagao Laboratory	DASAI	NEC	TOSHIBA
Hauptmerkmal des Systems	Transfer	Transfer	Transfer	Interlingua	Transfer
Übersetzte Sprachen	J - E	E - J J - E	J - E	E - J J - E	E - J
Übersetzungs- strategie	Fall-Grammatik mit Tree-Trans- ducer / Phrasen- struktur.	Tiefenseman- tische Repräsen- tation	Phrasenstruk- turierg. Tree-Trans- ducer	Dependency Grammar Augmented shiftre- duced Parser, tree- tree transducer	ATN
Wörter- bücher	Basis-W. 5000 Wörter	80000 Wörter (MU II: 30000 Wörter)	Basis-W. 50000 Wörter Techn. Termi- nologie 30000 Wörter	Basis-W. J: 40000 Wörter E: 53000 Wörter Sachgebiete 400000 Wörter	Basis-W. 50000 Techn. Termi- nologie 50000 Benutzerw. 30000
Ausstattung 1. Implementiert 2. auf OS 3. mit CPU (/Hardware)	Prolog u. Lisp Genera Symbolics	UTI-Lisp FACOM M xxx u. ZETA-Lisp Symbolics	ESP (Prolog) SIMPOS MELCOM PSI	C ACOS 4 ACOS Mainframe	C UNIX MC 68020
Geschwin- digkeit		4.000 w/h	5.000 w/h	60.000 w/h	7.000 w/h
zusätzliche Merkmale	Grammatik- Entwicklungs- umgebung	Texteditor, auch für Grammatik u. Wörterbuch	Grammatik- u. Wörterbuch- Entwicklungs- umgebung	Texteditor Wörterbuch-Update	Bilingualer Editor Wortprozessor Wörterbuch- Wartungssyst.

Name	PENSEE	RMT	SWP-7800	Sharp E-J MT
Hersteller/ Entwickler	OKI	RICOH	SANYO	SHARP
Hauptmerkmal des Systems	k.A.	Strukturtransfer	Transfer	k.A.
übersetzte Sprachen	J - E	E - J	J - E	E - J
Übersetzungs- strategie	Fall-Grammatik mit Tiefenstruktur	Augmented CFG Dependency trees	Augmented CFG	Augmented CFG
Wörter- bücher	Basis-W. 60000 Wörter Benutzerw. 40000	Basis-W. 30000 Benutzerw. 30000 Terminologie 50000	Basis-W. 55000 Wörter	Basis-W. 60000 Wörter + 40000 Wör- ter/Fachge- biet
Ausstattung				
1. Implementiert	C	C	C	C
2. auf OS	UNI PLUS +	UNIX System V	IRMX	UNIX System V
3. mit CPU (/Hardware)	MC 68010, MC 68020	3B2 (AT&T-CPU)	Intel 80186	MC 68010, MC 68020
Geschwin- digkeit	4.000 w/h	4.500 w/h	3.500 w/h	5.000 w/h
zusätzliche Merkmale	Bilingualer Editor Wörterbuch Editor	Wörterbuch- Entwicklungs- Werkzeug OCR-Option	Wörterbuch- editor Word- processor	Prä- und Posteditor



Studienführer LDV

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

DAS LDV-FORUM IM ABONNEMENT

Um das LDV-Forum regelmäßig zu beziehen, muß man/frau nicht Mitglied der GLDV sein: Seit etwa einem Jahr wird die Möglichkeit eines Abonnements angeboten. Ohne eigene Werbung dafür sind mittlerweile ca. 10 Abo's zusammengekommen. Der Vorteil gegenüber einem Einzelbezug besteht in einer gewissen Bequemlichkeit und in der Einsparung der Versandkosten. Der Nachteil gegenüber einer Mitgliedschaft besteht darin, daß Sie weder den elektronischen Newsletter, noch irgendwelche GLDV-Rabatte auf Veranstaltungsgebühren oder etwa auf die Olms-Reihe *Linguistische Datenverarbeitung* erhalten.

Sie bestellen ein Abonnement, indem Sie dies formlos der Redaktion mitteilen:

Redaktion des LDV-Forum, G. Knorz, c/o Fachhochschule Darmstadt, Fachbereich IuD, Schöfferstraße 8, D-6100 Darmstadt

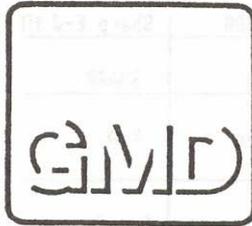
Den Jahrespreis von DM 30,00 incl. Versandkosten überweisen Sie bitte auf das Konto:

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

Vorkasse ist erforderlich.

d. Red.

STELLENANZEIGE



Das in Darmstadt gegründete
**Forschungsinstitut für
Integrierte Publikations- und Informationssysteme
der Gesellschaft für Mathematik und Datenverarbeitung**

befindet sich im Aufbau und erweitert seinen Mitarbeiterstab.

Die Aufgaben des Institutes umfassen alle Aspekte des elektronischen Publizierens und der Informationssysteme. Spezielle Schwerpunkte bilden objektorientierte, multimediale Wissensbanken, verteilte Systeme, Wissensrepräsentationsmethoden, integrierte Publikationssysteme und intelligente Mensch-Maschine-Schnittstellen einschließlich natürlichsprachlicher Ansätze.

Das Forschungsinstitut sucht

Wissenschaftliche Mitarbeiter,

die mithelfen, dem Institut schnell internationale Geltung zu verschaffen. Promotion ist erwünscht. Jüngeren Mitarbeitern wird die Gelegenheit zur Promotion geboten.

Informatiker(innen)

mit vertieften Kenntnissen in zumindest einem der Gebiete: Datenbanken, Informationssysteme, verteilte Systeme, Wissensrepräsentationen, Compilerbau, Mensch-Maschine-Interaktion.

Computerlinguisten(innen)

mit vertieften Kenntnissen in mindestens einem der Gebiete: wissensbasierte Verarbeitung von Texten, semantische Repräsentationen, Unifikationsgrammatiken, linguistische Pragmatik und Dialogsysteme

Die Bezahlung erfolgt nach BAT (Bundesangestellten-Tarif). Daneben werden die üblichen Sozialleistungen des öffentlichen Dienstes gewährt.

Bewerbungen mit Lebenslauf und Schriftenverzeichnis sind zu richten an:

G M D
Bereich Darmstadt
Postfach 10 43 26
Dolivostraße 15
6100 Darmstadt

GESELLSCHAFT FÜR MATHEMATIK UND DATENVERARBEITUNG (GMD)

Institut für Integrierte Publikations- und Informationssysteme, Darmstadt

Leitung: Prof. Dr. Erich J. Neuhold

Das Institut für Integrierte Publikations- und Informationssysteme wurde im Dezember 1986 als viertes Forschungsinstitut (F4) der Gesellschaft für Mathematik und Datenverarbeitung (GMD) mit Sitz in Darmstadt gegründet. Die GMD hat damit einen zusätzlichen Schwerpunkt ihrer wissenschaftlichen Arbeit geschaffen. Wie die anderen Forschungsstellen der GMD wird auch dieses Institut eng mit der Wirtschaft und den Hochschulen zusammenarbeiten. Zum 1. Januar 1988 übernimmt es einen Teil des Forschungs- und Entwicklungsbereiches der Gesellschaft für Information und Dokumentation.

Zum Leiter des Instituts ist Prof. Dr. Erich J. Neuhold berufen worden.

Das Institut hat die Aufgabe, wesentliche zukunftsweisende Beiträge zum elektronischen Publizieren, zu Informationssystemen und zur Entwicklung der Fachinformation insgesamt zu leisten. Langfristige Grundlagenforschung, mittelfristige Prototypenerstellung und kurzfristiger Transfer der Forschungs- und Entwicklungsergebnisse in die Industrie sollen verfolgt werden.

Um neue Ideen zu entwickeln, sie zu Büchern, wissenschaftlichen Texten, Tondokumenten, Bildern, Zeichnungen, Filmen oder anderen multimedialen Dokumenten auszuarbeiten und sie den Endbenutzern zugänglich zu machen, sind in allen Phasen Konzepte des Experimentierens, Illustrierens, Gestaltens, Edierens, Korrigierens, Druckens, Herstellens, Verteilens sowie des Zugriffs auf verschiedenste Informationen erforderlich. Dabei dürfen die Phasen von der Ideenfindung bis hin zur konkreten Publikation nicht isoliert betrachtet werden. Vorpublikatorische Tätigkeiten und solche, durch die das erstellte Dokument hinreichend einfach verteilt und genutzt werden kann, müssen vielmehr abgestimmt und integriert werden. Schließlich geht es nicht zuletzt darum, dem Einzelnen den Zugang zu den für ihn wichtigen Informationen zu erleichtern, damit ihm mehr Zeit für kreatives Arbeiten und für seine geistige Entwicklung bleibt. Voraussetzung hierfür ist es u.a., Systeme zu entwickeln, die einerseits durch anforderungsgerechte Unterstützungsmöglichkeiten zu qualitativ besseren Publikationen führen und andererseits ein schnelleres Lernen und Verstehen ermöglichen und damit das Verständnis von Texten erhöhen.

Inzwischen wurden sechs Forschungsbereiche gebildet, in denen im Rahmen verschiedener nach und nach anlaufender Projekte jeweils bestimmte Aspekte des gemeinsamen Zieles verfolgt werden. Diese Forschungsbereiche werden im folgenden kurz vorgestellt.

Integrated Information Systems

(Koordinator: Dipl.-Ing. Peter Fankhauser)

Der Kern zukünftiger Informationssysteme wird in diesem Forschungsbereich entwickelt. Dieser Bereich hat zum Ziel, Integrationsmethodiken zu entwickeln, mit deren Hilfe bereits bestehende und lokal verwaltete Datenbanksysteme bzw. verschiedene Fachinformationsbanken logisch zu einem Gesamtsystem integriert werden können. Ein solches System gibt dem Benutzer die Möglichkeit, Daten aus mehreren Datenbeständen oder Fachinformationsbanken gemeinsam zu verwalten und zu behandeln. Dies setzt aber die Kenntnis über die Beziehungen, die zwischen den einzelnen Datenbeständen bestehen, voraus. Meistens ist es aber nicht möglich, im vorhinein

sämtliche Beziehungen zwischen den einzelnen Datenbeständen explizit zu spezifizieren oder vom Benutzer ein entsprechendes Wissen anzunehmen. Im Forschungsbereich werden deshalb Konzepte entworfen, die es ermöglichen sollen, Beziehungen zwischen einzelnen Datenbeständen dynamisch, während der Benutzung des Gesamtsystems, zu erkennen und darzustellen.

Bei den Anwendungen, die es zu integrieren gilt, wird von einem breiten Spektrum ausgegangen. Es reicht von Literatur- und Volltextdatenbanken über Faktendatenbanken wie zum Beispiel Werkstoffdatenbanken oder Wirtschaftsdatenbanken bis hin zu hochgradig komplexen Wissensbasen. Die zu integrierende Manipulation verschiedenartiger Objekte, wie Texte, CAD-Objekte, Meßdaten, Tabellen und Sprachaufzeichnungen werden vom Forschungsbereich Datenbanktechnologie zur Verfügung gestellt.

Publication Systems

(Koordinator: Dr. Dr. Norbert Streit)

Die Aufgabe dieses Forschungs- und Entwicklungsbereiches besteht darin, eine verteilte integrierte Publikationsumgebung einzurichten, die als Modell für den gesamten Publikationsvorgang dienen kann. Diese Publikationsstraße soll alle an diesen Vorgängen beteiligten Personen und technischen Komponenten berücksichtigen. Dies erstreckt sich von der Einbeziehung des für die Autorentätigkeit herangezogenen Informationsangebotes über wissensbasierte Arbeitsplätze zur Erstellung, Gestaltung und Integration von Texten, Grafiken und Bildern, von Sprache und Video bis hin zum eigentlichen Druck- und Herstellungsvorgang und weiter bis zur Verteilung und Nutzung des fertigen Produktes. Bei dem Endprodukt kann es sich zum Beispiel um eine Zeitungsnachricht, einen wissenschaftlichen Artikel, ein gedrucktes Buch, ein multimediales elektronisches "Buch" oder um eine Fachinformationsbank handeln.

Die Projekte in diesem Forschungs- und Entwicklungsbereich werden sich u.a. mit der theoretischen Fundierung, der experimentellen und prototypischen Entwicklung sowie der empirischen Analyse und Validierung eines aktiven, wissensbasierten Autorensystems beschäftigen, das der gegenwärtig noch kleinen Gruppe von Ideenprozessoren und Hypertextsystemen zugerechnet werden kann. Diese Arbeiten sind am interdisziplinär ausgerichteten Ansatz der Cognitive Science orientiert und werden die Modellierung kognitiver Prozesse und Strukturen der Autorentätigkeit beinhalten.

Andere Arbeiten untersuchen die organisatorische Gestaltung und technische Realisierung einer Publikationsstraße durch existierende Komponenten. Dabei werden auch die Veränderungen traditioneller Rollen im Verlagsbereich untersucht und berücksichtigt, die sich bei der Erstellung, Verteilung und Rezeption von Dokumenten ergeben, wenn dies in elektronischer Form geschieht. In diesem Zusammenhang werden Dokumentstrukturen und Fragen des Hypertext-Publishing bearbeitet werden.

Die Publikationsstraße soll weiterhin dazu dienen, andere in F4 erzielte Forschungsergebnisse aber ebenso Lösungen von außen als Hardware- und Software-Komponenten zu einem integrierten System zusammenzufassen, mit dem sich die Möglichkeiten und Grenzen einzelner Bausteine beispielhaft demonstrieren lassen.

Knowledgebase Management Technology

(Koordinator: Dipl.-Ing. Wolfgang Klas)

Wissensbasen sollen in Zukunft die Speicherung und Handhabung von großen komplexen Datenbeständen, z.B. in Fachinformationsbanken aller Arten, erleichtern und dadurch auch den Komplexitätsgrad von Anwendungssystemen vermindern. Dazu müssen geeignete Datenmodelle und Formalismen entworfen werden, die sowohl die Struktur als auch die darauf definierten Operationen von komplexen Objekten wie z.B. Textdokumente, grafische Darstellungen und geometrische 3-dimensionale Objekte, aber auch Ton, Sprache, Animation und Video zu modellieren erlauben. Forschungsschwerpunkte sind dabei verschiedene erweiterte Mechanismen wie z.B. Versions-, Kopie- und Transaktionskonzepte, effiziente Strategien zur Objektverwaltung, Inferenzkomponenten, offene und inhaltsorientierte Schnittstellen zu Anwendungen. Um

Anforderungen, die mit großen dezentral gespeicherten Datenbeständen zusammenhängen, zu befriedigen, müssen in allen Fragen Aspekte verteilter Systeme berücksichtigt und untersucht werden. Um die Erweiterbarkeit und Kontrolle sicherzustellen, müssen das Datenmodell und die Abfragesprache theoretisch fundiert sein. Mit einem Prototyp soll das Modell für eine Reihe komplexer Objekttypen, darunter Grafik und Bild, entwickelt und überprüft werden.

Es wird angenommen, daß eine geeignete Kombination von objektorientierter Datenmodellierung und angepaßten Speichermanagementfunktionen zu Systemen führen, die Entwicklungszeiten verkürzen und außerdem gute Lösungsansätze für Repräsentationsprobleme in Wissensbasen, Büroinformationssystemen, Programmierumgebungen, CAD-Systemen usw. bieten. Kennzeichnend für alle diese Anwendungsbereiche ist, daß es um komplex strukturierte Objekte geht, die sich über einen längeren Zeitraum hinweg weiterentwickeln und verändern.

Cognitive User Interface

(Koordinator: Dr. Thomas King)

Ziel des Forschungsbereichs ist es, die Kommunikation und Interaktion zwischen Akteuren oder Systemen in Publikations- und Informationsprozessen zu erforschen, zu modellieren und die Ergebnisse dieser Arbeiten bei der Gestaltung anwenderfreundlicher Schnittstellen einzubringen. Das Ziel der Arbeiten besteht in der theoretischen Fundierung, experimentellen und prototypischen Entwicklung sowie der empirischen Überprüfung von anwenderorientierten multimedialen und multimodalen Interaktionsformen. Grundlegendes Paradigma dieser Arbeiten zur Mensch-Computer-Interaktion ist der interdisziplinär ausgerichtete Ansatz der Cognitive Science. Dabei werden die Möglichkeiten und Grenzen menschlicher Informationsverarbeitung und zwischenmenschlicher Kommunikation als Vorgaben für die Gestaltung ergonomischer Mensch-Computer-Schnittstellen angesehen. Dieser Ansatz der kognitiven Ergonomie erfordert einerseits, die für die intelligente Mensch-Computer-Interaktion geeigneten Formen der Wissensrepräsentation und Modellierung kognitiver Prozesse zu analysieren. Außerdem ist es notwendig, Werkzeuge für die Entwicklung von Komponenten für Benutzerschnittstellen und Anwendungsprogramme bereitzustellen. Zunächst sollen ein Wissenseditor, mit dem der Benutzer die Fachwissensbasis modifizieren bzw. ergänzen kann, ein Wissensnavigator, der den Benutzer durch die Wissensbasis führt, und eine Komponente zur Erkennung von Benutzerplänen entwickelt werden.

Innerhalb dieses Rahmens werden sich die Arbeiten vor allem auf psychologische und linguistische Aspekte der Mensch-Mensch- bzw. Mensch-Computer-Interaktion und ihre formale Modellierung beziehen. Darüberhinaus sollen prototypische Dialogschnittstellen für die in F4 zu entwickelnden Anwendungssysteme gestaltet werden.

Multimedia Knowledge Processing

(Koordinatorin: Dr. Karin Haenelt)

Dieser Forschungsbereich befaßt sich mit Texten als Publikations- und Informationsobjekten, die unter Berücksichtigung ihrer inhaltlichen Bedeutung und ihrer Zusammenhänge untereinander maschinell verarbeitet werden sollen. Verarbeitung wird dabei sehr umfassend verstanden. Es kann sich dabei ebenso um Übersetzungen in andere Sprachen handeln, als auch um die automatische Generierung von Texten zu gegebenen grafischen oder symbolischen Darstellungen oder umgekehrt die Erzeugung solcher bildhaften Darstellungen aus textlichen Beschreibungen. Eingeschlossen ist zum Beispiel auch die Erstellung von Kurzfassungen eines gegebenen Textes unter unterschiedlichen fachlich, inhaltlichen Blickwinkeln und für unterschiedlich fachkundige Leser mit entsprechend unterschiedlichem Wortschatz. Wichtige Anliegen sind es, den Wissenserwerb aus Texten zu ermöglichen, um Wissensbasen, zum Beispiel Faktendatenbanken, aufzubauen, sowie die durch die Auswertung erzielten Ergebnisse mit Publikationsmethoden erneut bereitzustellen. Daneben sollen auch Autoren unterstützt werden, die komplexe Gedanken so aufarbeiten müssen, daß sie linear als Text darstellbar sind.

Für die angestrebten natürlichsprachlichen Komponenten eines integrierten Publikations- und Informationssystems, die die genannten Aufgaben unterstützen sollen, sind zunächst theoretische Fundierungen zu erarbeiten. Unabdingbar ist es dabei, interdisziplinär zusammenzuarbeiten. Es

gilt zum Beispiel, Theorien, Modelle und Konzepte der Informatik und der Linguistik zusammenzuführen.

Hauptaufgabe der zunächst anlaufenden Projekte ist es, textorientierte Verfahren zu entwickeln. Das sind solche Verfahren, die auf der Ebene der Textstruktur arbeiten. Das bedeutet, daß die Textkonstitution die Operationen bestimmt, also nicht einzelne Einheiten wie Wörter, Sätze, Punkte oder Linien.

Publication and Information System Architecture

(Koordinator: Prof. Erich J. Neuhold)

Der Forschungsbereich PISA hat zum Ziel, die Anforderungen an Hardware und Software zu untersuchen, die für integrierte Publikations- und Informationssysteme zur Verfügung stehen sollten. Aus diesen Anforderungen soll dann eine Referenzarchitektur und daraus wieder eine Reihe von offenen Systemalternativen entwickelt und prototypisch realisiert werden. Werkzeuge für eine genaue (formale) Spezifikation der Systemeigenschaften und der auftretenden Schnittstellen werden eingesetzt bzw. falls notwendig auch entwickelt werden. Dieser Forschungsbereich unterstützt die übrigen Bereiche direkt in ihren Architekturfragen.

Die Institutsanschrift lautet:

Gesellschaft für Mathematik und Datenverarbeitung (GMD)
Institut für Integrierte Publikations- und Informationssysteme (F4)
Dolivostraße 15

6100 Darmstadt

Telefon (0 61 51) 8 75-800

e-mail: (z.B. NEUHOLD@DDAGMD11.BITNET)

Stand 21. 12. 1987

PROJEKTBERICHT AUS DEM BEREICH "ELECTRONIC PUBLISHING"

Projektstudium an der FH Darmstadt

Wie im Editorial der Juni-Ausgabe bereits angekündigt wurde, wird im Rahmen des im Wintersemester 87/88 am Fachbereich Information und Dokumentation der FHD angebotenen Projektstudiums ein Projekt aus dem Bereich "Electronic Publishing" durchgeführt.

Ziel des Projektes ist es, die Konzeption für den Prozeß der Produktion und des Angebotes des LDV-Forums unter Berücksichtigung des aktuellen Standes von Wissenschaft und Technik zu verbessern.

6 Studenten (*Peter Ditzer, Antje Kress, Claus Ritzler, Volker Scriba, Magit Stapf-Matlock, Ina Tauer*) des 5. Semesters befassen sich mit folgenden Teilprojekten:

a) Schnittstelle Autor – Redaktion

- technische Möglichkeiten der Textübermittlung (Mailbox, Diskette, Hardcopy & Scanner u.a.m.)
- Entwurf, Erprobung von Autorenrichtlinien in Anlehnung an SGML, wobei im Vordergrund steht, das Markup für den Autor so gering wie möglich zu halten. Dies soll dadurch erreicht werden, daß die Textauszeich-

nung hauptsächlich mittels einer in MProlog realisierten Grammatik automatisch durchgeführt wird, so daß der Autor beim Erstellen des Textes lediglich geringe Regeln beachten muß.

b) Spezifikation der redaktionellen Dienstleistungen

- Überlegungen zu Veränderungen bzw. Verbesserungen einzelner Rubriken des LDV-Forums hinsichtlich Gestaltung und Informationswert.
- Umsetzung dieser Überlegungen zusammen mit den Kommilitonen der Teilprojektes a)

c) Das LDV-Forum – sein Umfeld und seine Verbindungen

- Zusammenarbeit mit fachlich verwandten Zeitschriften (dazu: Übersicht erstellen und organisatorische Planung)
- Leseranalyse

Die einzelnen Arbeitsgruppen beabsichtigen, Sie über die Resultate der Arbeiten weiterhin zu informieren. Es ist außerdem beabsichtigt, ein Folgeprojekt im 6. Semester zu bearbeiten.

Projektgruppe "Electronic Publishing" FH Darmstadt, Fachbereich IuD, WS 87/88

DGD-KOMITEE

"WISSENSVERARBEITUNG – KI UND INFORMATIONSWISSENSCHAFT (KWV)"

Konstituierende Sitzung am 18. November 1987 im Fachbereich IuD der Fachhochschule Darmstadt.

Am 18. November 1987 fand im Fachbereich IuD der Fachhochschule Darmstadt die konstituierende Sitzung des DGD-Komitees "Wissensverarbeitung – KI und Informationswissenschaft" statt. Das Komitee tritt die Nachfolge von KLD ("Komitee Linguistische Datenverarbeitung") an und betreibt dessen Arbeit mit fortgeschobenem Programm weiter. Es ist aus der seit 1982 bestehenden Arbeitsgruppe "KI und Informationswissenschaft" hervorgegangen. Bisher wurden folgende Arbeitsthemen für KWV vorgeschlagen:

- Shells
- Rapid Prototyping
- Terminologie der Wissensverarbeitung
- Wissensrepräsentation und Klassifikationstheorie
- Methodik der Dokumentrepräsentation

Während des Symposiums "Computerlinguistik und ihre theoretischen Grundlagen" vom 09.-11.03.88 in Saarbrücken diskutiert das Komitee nochmals über die Repräsentation des SUSO-Szenarios, das eine Entscheidungssituation im Vorstand der SUSO-AG beschreibt. Seine Sitzung während des Dokumentartags (28.-30.09.88 in Aachen) will KWV schwerpunktmäßig dem Knowledge Engineering widmen. Ein vorläufiges Programm ist dazu schon ausgearbeitet.

Wer bei KWV mitarbeiten möchte, wende sich an: *Brigitte Endres-Niggemeyer, FH Hannover, Fachbereich BID, Hanomagstr. 8, D-3000 Hannover 91, Tel.: (0511) 444344* oder an: *Gerhard Knorz, FH Darmstadt, Fachbereich IuD, Schöffnerstr. 8, D-6100 Darmstadt, Tel.: (06151) 317881.*

AUTOCAT II ENTWICKLUNG EINES WISSENS- BASIERTEN ARBEITSPLATZSYSTEMS ZUR FORMALKATALOGISIERUNG

Die an der TH Darmstadt abgeschlossenen Arbeiten werden in einer Kooperation der Softwarefirma Danet (Darmstadt), der FH Darmstadt und dem FIZ Karlsruhe fortgeführt.

Von 1985 bis Ende 1987 arbeiteten Mitarbeiter des AUTOCAT-Projektes an der TH Darmstadt an der Aufgabe, die Formalkatalogisierung (d.h. die professionelle bibliographische Erfassung von Literatur) in Routinefällen einem Expertensystem zu übertragen. Als Arbeitsergebnis liegt ein in MPROLOG implementierter Prototyp vor, der vom optischen Leser kommende Zeitschriften analysiert und der die gefundenen Fachbeiträge nach den Regeln des INIS-Systems (International Nuclear Information System) bibliographisch erfaßt.

Ziel des Folgeprojektes AUTOCAT II ist es, aufbauend auf den Erfahrungen von AUTOCAT

- ein für Dokumentationszwecke brauchbares Arbeitsplatzsystem für den Dokumentinput zu konzipieren,
- in Zusammenarbeit mit FIZ Karlsruhe als Pilotanwender einen funktionalen Prototypen mit voll ausgebaute graphischer Benutzeroberfläche zu entwickeln und anschließend
- auf einer geeigneten low-cost Hardware zu implementieren.

Das Rapid Prototyping wird in Knowledge Craft auf einer Lisp-Machine durchgeführt, die Hardware-Auswahl für das Anwendungssystem steht noch aus. Der Fachbereich Information und Dokumentation (IuD) der FH Darmstadt ist für die wissenschaftliche Begleitung des bei der Danet laufenden Projektes zuständig und deckt insbesondere auch den Aspekt der Integration einer inhaltlichen Erschließung in das Arbeitsplatzsystem ab. Das Projekt hat am 1. Februar 1988 begonnen, dauert 2 Jahre und soll vom BMFT gefördert werden. Im ersten Halbjahr wird die Prototypentwicklung abgeschlossen sein und das Ergebnis vorgestellt werden.

Prof. Dr. G. Knorz, Fachbereich IuD, FH Darmstadt

MARKTSCHÄTZUNG EXPERTENSYSTEME

BRAINWARE veröffentlicht seit Dezember 1985 halbjährlich Marktschätzungen über die Größe des Expertensystemmarktes

Die neuesten Zahlen der Marktschätzung Expertensysteme (BRD, Stand: Dez. 87) zeigen, daß der Zuwachs an Lisprechnern weiterhin stark abnimmt (die Installationszahlen seit Dez. 85: 100, 300, 360, 390, 400). Dagegen ist ein kontinuierliches Wachstum bei den UNIX (Anz: 450) und PC-Installationen (Anz: 3300) zu verzeichnen. Bei den Großrechner-Implementationen ist das Wachstum noch sehr gering (140 Installationen).

Bei der personellen Ausstattung ist ein Trend hin zu AI-Gruppen mit 2 oder mehr festen Mitarbeitern zu beobachten. Diese Gruppen unternehmen (evtl. mit externer Beratung) erste Gehversuche und bemühen sich um die Realisierung eines später auch einsetzbaren Projektes.

Von einem breiten Durchbruch der nicht-prozeduralen Programmierung in den deutschen DV-Abteilungen kann nicht die Rede sein. Die Phase der Loslösung von der Forschung dauert immer noch an. Die Anzahl der Expertensysteme ist immer noch vergleichsweise gering. Während in den USA von ca. 5000 fertigen Systemen gesprochen wird, gibt es in der Bundesrepublik lediglich 150. BRAINWARE schätzt, daß von diesen lediglich etwa 25 eingesetzt werden, während es sich bei den anderen um "nützliche Lernerfahrungen" handelt.

Die Marktgröße hat sich bisher in einem Jahr jeweils verdoppelt (1987 90 Mio. DM). Es spricht alles dafür, daß sich das Wachstum weiter entwickeln wird. *Pressemitteilung BRAINWARE, Berlin (20. 1. 88)*

REZENSIONEN

Bátori, István; Weber, Heinz J. (Hgg):

NEUE ANSÄTZE IN MASCHINELLER SPRACHÜBERSETZUNG:

Wissensrepräsentation und Textbezug.

(New Approaches in Machine Translation: Knowledge Representation and Discourse Models.

Sprache und Information. Beiträge zur philologischen und linguistischen Datenverarbeitung, Informatik und Informationswissenschaft. Band 13. Tübingen: Max Niemeyer Verlag, 1986. 266 Seiten.

Der Band vereinigt 10 Aufsätze namhafter Autoren. Die Herausgeber verfolgen das sehr hoch gesteckte Ziel einer "Vorstellung von neuen Methoden und Perspektiven, die über die heute geltenden Standards in MÜ hinausgehen. Die Herausgeber wollten also ... solche Ansätze und Systeme präsentieren, die die Grenzen des heute Machbaren zu überwinden streben." (VIII) Dabei steht man vor dem Dilemma aller Veröffentlichungen in diesem sehr dynamischen Arbeitsgebiet, daß das Veröffentlichte zum Zeitpunkt des Erscheinens sehr oft bereits veraltet ist. Die Herausgeber verweisen z.B. auf das SIEMENS-System LITRAS, das schon nicht mehr vertrieben wird, auf die EUROTRA-D-Begleitforschung (aus der 2 Aufsätze stammen), deren Förderung inzwischen eingestellt ist, auf den SFB 100 der Universität Saarbrücken (aus dem ebenfalls 2 Artikel hervorgegangen sind), der seit Ende 1986 aufgelöst ist. Vor diesem Hintergrund ist eine Rezension des Bandes zum gegenwärtigen Zeitpunkt im Grunde notwendig ungerecht. Dennoch mag vorweg kritisch angemerkt werden, daß die hier vorgestellten "neuen Ansätze" einiger Autoren für den informierten Leser auch schon 1986 durchaus nicht mehr so neu waren.

Den Anfang des Bandes macht ein knapper Überblick über die Forschungsgeschichte der MÜ von I. Bátori, der drei Entwicklungsphasen unterscheidet: a) die Phase der Dominanz von Informatikern ohne oder mit nur geringen Linguistikkenntnissen; b) die Phase der linguistischen Dominanz, die von einer Überschätzung der Rolle der Theorie, insbesondere der Syntax, geprägt war; c) die Phase der Beeinflussung durch die KI-Forschung. Ganz entscheidend ist für Bátori, daß Arbeiten zur MÜ nur als interdisziplinäre Forschung betrieben werden können.

U.L. Figge ("Gedächtnis und Maschinelle Übersetzung") stellt als neuestes Ergebnis seiner fast 20 Jahre¹ währenden Forschungen zur algorithmischen Grammatik heraus: "Ich bin der Überzeugung, daß Sprachproduktion im Gedächtnis ihren Ausgang nimmt, daß Sprachrezeption auf das Gedächtnis zielt und daß menschliche Übersetzung über das Gedächtnis läuft." (29/30) In seiner Sicht "besteht

... ein Gedächtnis aus Konzepten und gehen diese Konzepte vielfältige Beziehungen miteinander ein." (30) Figge macht die bedeutsamen Feststellungen: "Übersetzung ist eine Art von Sprachverarbeitung. Ein Übersetzer verarbeitet Texte: ... Mit anderen Worten: Übersetzung ist ein zweisprachiger Prozeß, der über das Gedächtnis des Übersetzers läuft." (37) Diese Sprachverarbeitung hält er für "im Prinzip maschinell" simulierbar (36). So kommt er schließlich zu der den Leser sicherlich überraschenden Aussage: "Nach meiner Überzeugung hat ein maschinelles Übersetzungssystem, das im Rahmen des Möglichen versucht, einen menschlichen Übersetzer zu simulieren, die größten Chancen, hochqualifizierte Übersetzungen zu produzieren." (37)

P.E. Pause ("Zur Modellierung des Übersetzungsprozesses") kritisiert eingangs die derzeit herrschende euphorische Grundstimmung im MÜ-Umfeld und verlangt eine realistische Einschätzung der Möglichkeiten auf der Grundlage von mindestens drei Fragen. Diese beziehen sich 1. auf die Angabe der Textsorten und den Status der Übersetzungen, 2. auf die Aussagen der Übersetzungstheorie und 3. auf Fragen zur Effizienz. Pause behauptet, daß "die Optimisten die erste Frage nicht allzu präzise beantworten wollen" (46). Wer mit "Optimisten" gemeint sein mag, soll dahingestellt bleiben. Der Rezensentin sind keine MÜ-Projekte bekannt, die nicht in der Lage oder bereit sind, die ausgewählte Textsorte und den angestrebten Übersetzungsstatus exakt zu benennen. Pause beklagt, daß die 3. Frage in den Vordergrund gestellt wurde; das mag für die Finanziere von MÜ-Projekten zutreffen, ganz sicher aber nicht für die Projekt-Mitarbeiter. Unverständlich ist der Rezensentin Pauses Forderung in der für ihn zentralen Frage 2, die Übersetzungstheorie müsse Aussagen zur Praxis (Realisierbarkeit) von MÜ-Arbeiten machen. Wünschenswert ist allerdings in der Tat die Entwicklung einer modernen fachsprachenbezogenen Übersetzungstheorie. – Pause geht schließlich auf die Problematik von Satz- und Text-Übersetzung ein, entwickelt ein komplexes Text-Verstehensmodell, das er am Beispiel anaphorischer Wendungen erläutert. Dieses Modell erklärt er als ein Übersetzungsmodell, das derzeit mangels hinreichenden linguistischen Wissens nicht vollständig maschinell simulierbar sei. Das Ergebnis seiner diffizilen Ausführungen ist die Feststellung, daß für in der Praxis verwendbare Übersetzungssysteme nur 2 Wege offen blieben: "entweder (a) stark normierte Texte in stark normierten Sprachen (...) zu übersetzen, ..., oder (b) intelligente komplexe Wörterbücher zu simulieren als Hilfe für den menschlichen Übersetzer." (72) – Um was sonst hat sich denn die MÜ-Forschung und -Praxis bisher bemüht?

Chr. Rohrer ("Maschinelle Übersetzung mit Unifikationsgrammatiken") stellt in seinem Beitrag die Möglichkeiten der Unifikationsgrammatik, insbesondere in der Form der LFG, für den Einsatz in der MÜ vor. Zu Beginn erläutert er Konzept und Notationen dieses Grammatiktyps, beschreibt die Vorteile bei ihrer Anwendung auf die MÜ (mit Hinweis auf die amerikanischen Vorbilder Bresnan und Kay) und exemplifiziert Vorstellungen zur Entwicklung der Transfer- und Synthesekomponente eines MÜ-Systems, die im Analyse- und Transferbereich bereits zu Implementierungen geführt haben. Trotz erfolgreicher Versuche mit einem minimalen Datensatz warnt Rohrer vor allzu großem Optimismus und vertritt die Auffassung, "daß eine gute maschinelle Übersetzung von beliebigen Texten zu vernünftigen Preisen in absehbarer Zeit nur durch eine Einschränkung des Sprachausschnitts und des Gegenstandsbereichs möglich ist." (97/98) Bedauerlich ist, daß die Projektförderung durch das BMFT inzwischen eingestellt wurde, erfreulich hingegen, daß die Arbeit im Rahmen anderer Projekte weitergeführt werden kann.

Chr. Boitet und R. Gerber ("Expert Systems and Other New Techniques in M(a)T") berichten von der stufenweisen Entwicklung von Softwaretools für die MÜ, bei denen besonderer Wert auf eine benutzerfreundliche Programmierumgebung gelegt wurde. Für die MÜ selbst wurde dabei der Aufbau eines geeigneten Expertensystems für unabdingbar erkannt. Intensiv bemüht war man überdies um möglichst unproblematische Datenaustauschmöglichkeiten mit anderen Systemen. Von besonderer Bedeutung für ihre Arbeiten waren Erkenntnisse aus der KI-Forschung. – Dieses Paper ist die erweiterte Fassung eines Vortrags auf COLING 1984.

D. Rösner ("SEMSYN - Wissensquellen und Strategien bei der Generierung von Deutsch aus einer semantischen Repräsentation") stellt das Generierungssystem SEMSYN vor, das zusammen mit dem ATLAS/II-System des japanischen Partners FUJITSU ein japanisch-deutsches Übersetzungssystem ergibt, bisher vor allem für Titel aus dem Bereich der Informationstechnologie. Das System besticht durch den Vorteil, so variabel angelegt zu sein, daß es dem Linguisten leicht möglich ist, neue Beschreibungskomponenten zu integrieren, ohne daß Programmieränderungen erforderlich werden. Grundlage des Generierungssystems ist die Überführung von FUJITSUs vorgegebenen semantischen Netzen in eine neu entwickelte Repräsentationssprache (Frames u.a.). Es folgt die Entscheidung, in welcher syntaktischen Form die in diesen semantischen Netzen enthaltenen Inhalte im deutschen Text wiedergegeben werden sollen. Endelement ist die Produktion einer dementsprechenden syntaktisch und morphologisch korrekten Oberflächenstruktur. Im Unterschied zu den meisten anderen MÜ-Systemen ist SEMSYN weitgehend unabhängig von einer bestimmten Ausgangssprache; das Endprodukt wird inhaltsorientiert

erzeugt, die syntaktischen Grundlagen des Ausgangstextes bleiben ohne Einfluß. In SEMSYN ist ein höchst interessanter und völlig andersartiger Zugriff auf MÜ-Aufgaben dokumentiert.

M. Saluveer und Haldur Öim ("Rules and Reasoning in Text Comprehension") beschäftigen sich nicht mit Fragen der maschinellen Übersetzung oder der Übersetzungstheorie. Ihr Anliegen ist die Untersuchung von argumentativen Interaktionsstrukturen. Als Hauptergebnis stellen sie heraus, daß die menschliche Kommunikation in versteckter Weise den gleichen Regeln unterliegt wie monologische Texte. Inwieweit dies einen Beitrag zu den zahlreichen Problemen der MÜ darstellt, muß hier unentschieden bleiben.

Christa Hauenschild stellt in ihrem Beitrag "KIT/NASEV oder die Problematik des Transfers bei der Maschinellen Übersetzung" Überlegungen zur Diskussion, die man in dem 1985 begonnenen Berliner EUROTRA-D-Begleitforschungsprojekt zur Transferkomponente des projektierten MÜ-Systems angestellt hat. Kernauftrag des Berliner Teilprojekts war es, die Eignung der Generalized Phrase Structure Grammar für ein MÜ-System zu prüfen. Da bald erkannt wurde, daß diese allein unzureichend war, ergänzte man das System um weitere Komponenten im Analyse- und vor allem im Transferbereich. Hauenschild stellt hier vor allem die Möglichkeiten der dort entwickelten Semantic-Representation-Language insbesondere bei den diversen zu unterscheidenden Transferzwischenstufen ausführlich vor. Die Autorin erklärt abschließend ehrlicherweise, daß aufgrund des Projekt-Anfangsstadiums naturgemäß noch sehr vieles nur spekulativen Charakter besitzt.

Anneli Rothkegel berichtet in "Textverstehen und Transfer in der Maschinellen Übersetzung" über den Versuch, im Rahmen des SFB 100 über die Satz-Analyse und -Übersetzung hinauszugelangen zu einer satzübergreifenden Text-Analyse und -Übersetzung. Sie stellt als die beiden wesentlichen Elemente dieses Versuchs, der zur Entwicklung des Programmsystems TEXAN (auf der Grundlage des in SUSY integrierten Syntaxparsers SATAN) geführt hat, zunächst ein Textverstehensmodell vor, das sich auf die Textsorte EG-Abkommen konzentriert. Anhand konkreter Beispiele werden die Vorstellungen zu einer Strukturbeschreibung von "Texthandlungen" dieses Texttyps entwickelt, die maschinell erzeugt werden kann. Den dazu erforderlichen Formalismus erläutert A. Rothkegel im 2. Teil des Artikels. Grundlage ist das Frame-Konzept, das hier zu einem System von Texthandlungsframes weiterentwickelt und wiederum an konkreten Beispielen erläutert wird. Die Bedeutung dieser Arbeiten für das EUROTRA-D-Projekt liegt auf der Hand.

Auch H.J. Weber ("Faktoren einer textbezogenen Maschinellen Übersetzung: Satzstrukturen, Kohärenz- und Koreferenz-Relationen, Textorganisation") stellt nach einer kurzen Zusammen-

fassung divergierender bisheriger Ansätze zu Problemlösungsversuchen in der MÜ den Textbezug als Voraussetzung für eine verbesserte automatische Übersetzung in den Vordergrund. Ausgangspunkt dafür bleibt ihm jedoch die syntaktisch-morphologische Satzanalyse, für die er auf das bereits bekannte System SATAN zurückgreifen kann. Von hier aus wird als nächst größere Einheit der Textabschnitt durch die Erfassung der semantischen Relationen zwischen den Sätzen beschrieben. Materialgrundlage der Arbeiten sind EG-Nachrichten, ergänzt um Rundfunk- und Presse-Nachrichten, die ihm für eine weitgehend einheitliche Systematisierung der Kohärenz- und der Koreferenz-Beziehungen besonders geeignet erscheinen. Für die Untersuchung dieser Beziehungen ist COAT mit 4 Systemelementen verantwortlich (TEXTOR, CHOPPR, SEPARÉ² und LETRA), die bereits implementiert sind und die das Ergebnis von SATAN zur Weiterverarbeitung übernehmen. Das Ergebnis der Analyse von Relationen zwischen Textabschnitten, die schließlich zur Gesamtanalyse eines Textes führt, soll dann von OVERCOAT übernommen werden, einem Modul, das bei Manuskriptabschluß jedoch noch nicht implementiert war. Dieses neue Systemelement soll den Nachteil "einseitig verstehensorientierter" Systeme (251) mit ihren monotonen Aneinanderreihungen von Textteilen überwinden. Allerdings soll auch mit OVERCOAT noch nicht die fertige Übersetzung geleistet, sondern erst eine Textbeschreibung als "Zwischenbereich"

für eine adäquate Übersetzung geschaffen werden. – Die Arbeitsweise von COAT und OVERCOAT wird zwar theoretisch erläutert, praktische Systemablauf-Beispiele (etwa des implementierten COAT-Teils) werden leider jedoch nicht gezeigt. Nicht nur wer SATAN kennt, wäre wohl auf eine solche Demonstration der vielversprechenden Weiterentwicklung gespannt gewesen.

Daß dieser Band, der auch aus heutiger Sicht vor allem wegen der Vorstellung einiger Versuche, in der MÜ von der Satzebene zur Textebene vorzudringen, immer noch interessant erscheint, als fotomechanischer Abdruck von Typoskripten teuer verkauft wird (DM 86,00), mag selbst bei einem so renommierten Verlag aus Kostengründen verständlich erscheinen; daß Einzelteile der Beiträge offensichtlich (nicht immer geschickt) stückweise zusammengeklebt wurden, mag aus der Arbeitsüberlastung der Autoren begrifflich sein; daß bis auf D. Rösners Beitrag kein einziger Artikel "professionell" geschrieben wurde, verwundert; die Vielzahl an orthographischen und grammatischen Regelverstößen aus der Produktion einiger der Linguisten verursacht zumindest ungläubig-ärgerliches Kopfschütteln.

Martina Schwanke, Germanistisches Seminar der Universität Kiel

¹ Vgl. Herbert E. Bruderer: *Handbuch zur maschinellen und maschinenunterstützten Sprachübersetzung natürlicher Sprachen und mehrsprachigen Terminologiedatenbanken*. München/New York 1978, 131.

Herrmann, Thomas:

ZUR GESTALTUNG DER MENSCH-COMPUTER-INTERAKTION:

SYSTEMERKLÄRUNG ALS KOMMUNIKATIVES PROBLEM.

Sprache und Information. Beiträge zur philologischen und linguistischen Datenverarbeitung, Informatik und Informationswissenschaft. Band 14. Tübingen: Max Niemeyer Verlag, 1986. XIV, 259 Seiten.

Da diese Dissertation am Institut für Kommunikationsforschung und Phonetik der Universität Bonn von Prof. Lenders und am Fachbereich Informatik der TU Berlin von Prof. Floyd betreut wurde und im Rahmen der Tätigkeit des Autors bei der Gesellschaft für Mathematik und Datenverarbeitung in Birlinghoven entstand und beeinflusst wurde von Diskussionen im Arbeitskreis Rationalisierung Bonn, beruht auf keinem Zufall. Dieser "Zufall" hat Methode. Und diese Methode ist dem Untersuchungsgegenstand angemessen: sie behandelt Probleme der Software-Gestaltung der Mensch-Maschine-Kommunikation im Sinne der Informatik

ebenso wie kommunikationstheoretische Einsichten und kommunikationspraktische Fragestellungen, die den gesellschaftspolitischen und organisationsstrukturellen Hintergrund des Einsatzes neuer Technologien am Arbeitsplatz bis eben hin zur Rationalisierung von Erklärungsleistungen durch ein System nicht aussparen.

Die Grundfragen die der Autor stellt, gehen nicht aus von einer technokratischen Kosten-Nutzen-orientierten Prämisse der Leistungsoptimierung und Aufwandwandsminimierung. Sie nehmen zum Ansatz vielmehr einerseits die Betroffenen selbst (ihre Erklärungsbedürfnisse gegenüber Computersystemen und ihre Vorstellungen und Modellbildungen vom Computer), andererseits aber auch das Thema, welche kommunikativen Leistungen Menschen erbringen müssen, wenn sie den Erklärungswünschen anderer genügen wollen, und die daran anschließende Frage, ob diese zwischenmenschliche Kommunikation überhaupt an den Computer delegiert werden kann. Dabei denkt der Autor an Benutzer, die mit dem Rechner nicht nur Routinearbeiten ausführen, sondern mit dem Rechner Probleme bearbeiten, die eine jeweils kreative Lösung bedingen.

Der Autor arbeitet das weite Themenspektrum zwischen Software-Gestaltung, Künstlicher Intelligenz, Linguistik, Kommunikationsforschung und Benutzerforschung in sechs Kapiteln auf.

Im ersten klärt er Grundbegriffe und diskutiert die bislang vorgeschlagenen Modelle zur Mensch-Computer-Interaktion (= MCI: diese Wortwahl ist nicht zufällig).

Im zweiten Kapitel wird ein vom Autor entwickeltes Modell der MCI diskutiert, das das Computerbild des Benutzers zum Ausgangspunkt hat.

Im dritten Kapitel werden verschiedene Inhalte des Konzepts 'Selbsterklärung' erörtert und unterschieden nach "indirekter, deskriptiver und argumentativer Selbsterklärung".

Im vierten Kapitel geht es – ausgelöst von der Frage, ob argumentative Erklärung durch den Rechner überhaupt leistbar ist oder ob sie nicht gerade das konstitutive Element von zwischenmenschlicher Kommunikation ist und damit dieser Art der Interaktion vorbehalten ist – um prinzipielle Unterschiede zwischen Mensch-Mensch-Kommunikation und MCI.

Im vorletzten Kapitel wird dieser Faden weitergesponnen. Diskutiert wird nämlich der sprechakt-orientierte Ansatz, wie er in der KI etwa seit 1980 unter dem Stichwort 'kommunikative Adäquatheit' vertreten wird und der eine wenigstens teilweise Übertragung zwischenmenschlicher Kommunikationsleistungen auf den Rechner propagiert. Der Autor zeigt nicht nur, daß "die derzeitigen Defizite der linguistischen und kommunikationstheoretischen Ansätze für dieses Vorhaben zu groß sind" (S. 5), sondern darüber hinaus, "daß das Paradigma zwischenmenschlicher Kommunikation unter wesentlichen Aspekten für die Software-Gestaltung unbrauchbar ist" (S. 3). Dabei bezieht er sich vor allem darauf, daß Kommunikation den Verständigungserfolg prinzipiell nicht mit letzter Sicherheit kontrollierbar macht; daß die Realisierung von Verständigungsprozessen bei sich wandelnden sozialen Kontextbedingungen immer wieder neu gelernt werden muß; daß kommunikative Verständigung innerhalb sozialer Strukturen auf der Basis von Vertrauen gerade in die Leistungsfähigkeit menschlicher Kommunikation funktioniert; daß sich eine hörer-orientierte Modellierung einer Verständigungsaushandlung nicht mit einem Ansatz vereinbaren läßt, der einen Ausschnitt einer "formalen" Kommunikation herausgreift, für den darüberhinaus alle potentiellen Äußerungen des Sprechers und ihre Wirkungen auf den Hörer im Vorhinein vorhergesagt und damit beschrieben werden müßten; daß Argumentation in zwischenmenschlicher Kommunikation nicht nach formal-logischen Prinzipien aufgebaut sein muß, sondern zu Mitteln der Persuasivität (bis hinein in sprachliche Formulierungen) greift.

Im sechsten und letzten Kapitel entwickelt der Autor als Alternative zur "kommunikativen Benut-

zung" des Rechners sein "Prinzip der intervenierenden Benutzbarkeit ..., bei dem sich der Benutzer seine Arbeit mit dem System selbst mit Hilfe des Rechners zum Zweck der Systemerklärung dokumentieren kann ..., vorhandene Meldungen und Erklärungstexte nach seinen Vorstellungen modifizieren können (soll) und ihm Dialogbeispiele und exemplarische Datenbestände anzubieten (sind), um ihn beim Experimentieren mit dem System zu unterstützen" (S. 5).

In einem Anhang wird schließlich ein transkribiertes Erklärungsgespräch wiedergegeben, um zu zeigen, wie das im zweiten Kapitel entwickelte Kategoriensystem auf empirisches Material angewendet werden kann, um aber auch zu illustrieren, wie "mächtig" eine Software für eine "kommunikative Benutzung" eines Systems sein müßte, die alle potentiellen Erklärungsaufforderungen (die in den wenigsten Fällen in Frageform formuliert sind) antizipierte. Herrmanns Konzept der intervenierenden Benutzung basiert auf dem Gedanken, daß nicht der Rechner sich selbst erklärt, sondern daß ein System so transparent aufgebaut ist und so vielfältige Auskunft- und Interaktionsmodi (aber nicht im Sinne von argumentativen Klärungsdialogen) zur Verfügung stellt, daß sich der Benutzer das System selbst erklären kann.

Aufgrund des derzeitigen Forschungs- und Entwicklungsstandes ist es nicht überraschend, daß der Autor an dieser Stelle nicht in der Lage ist, einen umfassenden Gestaltungsvorschlag zu machen (er befindet sich dabei in guter Gesellschaft; vgl. Terry Winograd/Fernando Flores, *Understanding Computers and Cognition: A New Foundation for Design*. Norwood: Ablex 1986). Er ist vielmehr gezwungen, es bei der kursorischen Nennung und kurzen Erläuterung einzelner möglicher Elemente für eine derartige Schnittstelle zu belassen und darauf hinzuweisen, daß "eine Umsetzung des vorgeschlagenen Gestaltungskonzepts 'intervenierende Benutzbarkeit' ... letztlich nur im Kontext empirischer Evaluation vorgenommen werden (kann)" (S. 218), eines Forschungsbereichs, der noch in den Kinderschuhen steckt. Soviel aber wird deutlich. Herrmann entwirft den Prospekt eines Arbeitsplatzes, den der Betroffene selbstbestimmt aufbauen und modifizieren kann und an/in dem der Rechner zwar nicht als Partner fungiert, aber als eine Einheit von Maschine, Nachricht und Medium so ausgelegt ist, daß die Anpassungsleistung des Benutzers an den Rechner gegenüber heutigen Systemen einerseits verringert würde und die Anpassung, die nicht zu vermeiden ist, andererseits direkt erfahrbar wird.

Herrmann versucht auch deutlich zu machen, daß sein Ansatz eine Computer-Kompetenz aufseiten des Benutzers entstehen lassen würde, die die zwischenmenschlichen kommunikativen Fähigkeiten (sprachliche Kompetenz, kommunikative Kompetenz, kommunikationstheoretische

Kompetenz) um eine Dialogkompetenz, Benutzungskompetenz und interaktionstheoretische Kompetenz für MCI ergänzen würde, ohne daß diese Computerkompetenz in ein Konkurrenz- oder (noch schlimmer) Interferenzverhältnis zu den zwischenmenschlichen kommunikativen Fähigkeiten treten würde.

Diese bemerkenswerte Arbeit ist jedem zu empfehlen, der sich mit der Problematik der MCI auseinandersetzt oder sie kennenlernen will, wobei es nicht erheblich ist, ob er dies unter informatischen, kommunikationstheoretischen oder sozialwissenschaftlichen oder welchen Fragestellungen auch immer tut. Er wird über die jeweils anderen Bereiche viel Wissens- und Bedenkenswertes (einschließlich eines Verzeichnisses der relevanten Literatur) finden und seine eigenen Problemstellungen in einen größeren Zusammenhang zu stellen in der Lage sein. Darüber hinaus wäre zu wünschen, daß sie einen Anstoß gibt zu einer verstärkt empirisch abgesicherten Forschung im Bereich der MCI, der zwar nicht hinreichenden, aber doch notwendigen Voraussetzung für die Gestaltung qualitativ besserer MCI-Schnittstellen.

Dr. Hans Dieter Lutz, Erziehungswissenschaftliche Hochschule Rheinland-Pfalz, Abteilung Koblenz, Studiengang Angewandte Informatik, Schwerpunkt Linguistik

DER REDAKTION ...

... liegen folgende Bücher als Rezensionsexemplare vor:

- Kuhlen, R.: Informationslinguistik. Sprache und Information. Tübingen: Max Niemeyer Verlag, 1986. XV, 256 Seiten.
- Wilss, W.; Schmitz, K.-D.: Maschinelle Übersetzungsmethoden und -Werkzeuge. Sprache und Information. Tübingen: Max Niemeyer Verlag, 1986. XVI, 324 Seiten.
- Luckhardt, H.-D.: Der Transfer in der maschinellen Sprachübersetzung. Sprache und Information. Tübingen: Max Niemeyer Verlag, 1986. XVIII, 221 Seiten.

Auch an Rezensionen der inzwischen erschienenen Bände der Olms-Reihe "Linguistische Datenverarbeitung" besteht naturgemäß ein besonderes Interesse. Dies ist jedoch keinesfalls einschränkend zu verstehen: Jedes einschlägige Buch kann der Redaktion zur Besprechung vorgeschlagen werden.

Publikationen

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

GID-Informationszentrum (GID-IZ)
Herriotstraße 5
D-6000 Frankfurt 71

die Standardprofile *Linguistische Datenverarbeitung* und *Künstliche Intelligenz* der Datenbasis INFODATA regelmäßig zur Verfügung stellt. 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 subjectiv 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 unter Angabe der Bestellnummern an die Bibliothek des GID-IZ:

GID-IZ Bibliothek, Herriotstraße 5, D-6600 Frankfurt-71
Tel.: (069) 66 87-374 oder 66 87-373

Verarbeitete Ausschnitte der INFODATA-Datenbank: Standardprofile vom ersten Halbjahr 1987 (genauer: ab Dez. 86 (Erf.-Nr. 4001) bis Juli 87 (Erf.-Nr. 2800)).

Aufbau einer Dokumentationseinheit:

Autor, A.: **Der Titel in Fettdruck, mit einem Punkt abgeschlossen.** Die Quelle, mit Seitenangabe, Deskriptoren, durch Strichpunkt getrennt – Ref.Nr.: Bestellnummer

Falls mehrere Zitate aus einer Quelle stammen, erscheinen nach dem ersten alle folgenden Referenzen eingerückt, wobei die Quellenangabe sich dann auf die Seitennummern beschränkt.

Selfridge, M.: **Integrated processing produces robust understanding.** Computational linguistics, 12(1986)2, S. 89-106, *Man-machine communication; Language; Syntax; Semantics; Computational linguistics; Natural language processing; User interface; Testing; MURPHY* – Ref.Nr.: 86-04001

Kalita, J. K.; Jones, M. L.; MacCalla, G. I.: **Summarizing natural language database responses.** Computational linguistics, 12(1986)2, S. 107-124, *Man-machine communication; Pragmatics; Logic; Search result; Database; Natural language processing; Heuristics* – Ref.Nr.: 86-04002

Knorz, G.: **IR-Konzeptionen für Bürosysteme auf der Basis des AIR-Ansatzes.** Sprache und Datenverarbeitung, 10(1986)1, S. 29-36, *Automatic indexing; Information system; Use; Information retrieval; Office organization; Draft* – Ref.Nr.: 86-04028

Hendry, I. G.; Willett, P.; Wood, F. E.: **INSTRUCT: a teaching package for experimental methods in information retrieval. Part 1 – The users' view.** Program, 20(1986)3, S. 245-263, *Information retrieval; Man-machine communication; Search; Software; User training; Weighting; Ranking; INSTRUCT* – Ref.Nr.: 86-04046

Salton, G.: **Another look at automatic text-retrieval systems.** Communications of the ACM, 29(1986)7, S. 648-656, *Information retrieval system; Indexing techniques; Automatic indexing; Comparison; Freertext searching; Vocabulary; Control (Check); Descriptor; Effectiveness; Testing* – Ref.Nr.: 86-04105

Gaines, B. R.: **From ergonomics to the fifth generation. Thirty years of human-computer interaction studies.** In: Shackel, B. (Hrsg.); International Federation for Information Processing (IFIP), Task Group on Human-Computer Interaction: Human-computer interaction – INTERACT '84. Proceedings of the IFIP conference, London, U.K., September, 1984. North-Holland (1985), S. 3-7, *Man-machine communication; Hardware; Software; Artificial intelligence; Software engineering; Human factor* – Ref.Nr.: 86-04118

Higgins, C. A.; Whitrow, R.: **On-line cursive script recognition.** S. 139-143, *Manuscript; Pattern recognition; Text analysis* – Ref.Nr.: 86-04125

Benyon, D.: **Monitor: a self-adaptive user interface.** S. 335-341, *User interface; Artificial intelligence; Frame; Prototype; MONITOR* – Ref.Nr.: 86-04135

Helbig, G.: **Geschichte der neueren Sprachwissenschaft. – Unter dem besonderen Aspekt der Grammatik-Theorie.** Leipzig, DD: Bibliograph. Inst. (1986), 392 S., *Linguistics; Grammar; Generative Grammar; History* – Ref.Nr.: 86-04205

Giordana, A.; Saitta, L.; Laface, P.: **Discrimination of words in a large vocabulary using phonetic descriptions.** International journal of man-machine studies, 24(1986)5, S. 453-473, *Speech processing; Pattern recognition; Phonetics; Dictionary; Vocabulary; Italian; Testing* – Ref.Nr.: 86-04215

Konrad, E.; Reiner, U.: **Zur Semantik von Anfragesprachen für Dokumentennachweissysteme.** In: Strohl-Göbel, H. (Bearb.); DGD: Deutscher Dokumentartag 1985, Nürnberg, Oktober 1985. Saur (1986) S. 180-197, *Systems research; Information retrieval; Command language; Semantics; Model; Method* – Ref.Nr.: 86-04274

Fuhr, N.: **Ranking-Experimente mit gewichteter Indexierung.** S. 222-238, *Information retrieval; Automatic indexing; Ranking; Weighting; Testing; Search result; Evaluation; AIR* – Ref.Nr.: 86-04276

Wessoly, R.: **Evaluation von SUSY-DJT.** S. 239-245, *Automatic translation; German; Japanese; English; Evaluation; Empirical study; Journal; Title; Federal Republic of Germany; SUSY-DJT; TITRAN* – Ref.Nr.: 86-04277

- Korostelev, L. Yu.: **Features of processing of unidentified words in a machine translation system.** Automatic documentation and mathematical linguistics, 19(1985)2, S. 95-102, *Automatic translation; Orthography; Error; Correction; Dictionary; Testing* – Ref.Nr.: 86-04310
- Haslett, B. J.: **A critical analysis of van Dijk's Theory of discourse.** In: Dervin, B. (Hrsg.); Voigt, M. J. (Hrsg.): Progress in communication sciences. Volume 7. Ablex (1986) S. 153-193, *Communication theory; Communication; Verbal; Text; Structure; Context; Text analysis; Cognition* – Ref.Nr.: 86-04314
- Smith, L. C.: **Knowledge-based systems, artificial intelligence and human factors.** In: Ingwersen, P. (Hrsg.); Kajberg, L. (Hrsg.); Pejtersen, A. M. (Hrsg.): Information technology and information use. Towards a unified view of information and information technology.: Taylor Graham (1986) S. 98-110, *Expert system; Information retrieval; Trend* – Ref.Nr.: 86-04347
- Campbell, H. (Hrsg.): **Computer information systems in China. Current advances and prospects to 1995.** China Society for Scientific and Technical Information (CSSTI), Beijing, CN: International Chinese symposium on computerized information retrieval 1, Beijing, CN: 1985.10 Toronto, CA: China Business Consulting (1986), *Information policy; Planning; Dialog service; Reference database; Factual data bank; Information network; Character; Chinese; Automatic indexing; Training in librarianship; China* – Ref.Nr.: 86-04422
- Young, T. Y. (Hrsg.); Fu, K. S. (Hrsg.): **Handbook of pattern recognition and image processing.** Academic Press (1986), 725 S., *Pattern recognition; Figure; Computational linguistics; Speech recognition; Classification characteristic; Mathematical method* – Ref.Nr.: 86-04427
- Grishman, R. (Hrsg.); Kittredge, R. (Hrsg.): **Analyzing language in restricted domains. Sublanguage description and processing.** Hillsdale, NJ: Erlbaum Associates (1986) 263 S., *Computational linguistics; Terminological research; Text analysis; Syntax; Artificial language; Semantics; Freetext searching* – Ref.Nr.: 86-04438
- Sager, N.: **Sublanguage: linguistic phenomenon, computational tool.** S. 1-17, *Computational linguistics; Terminological research; Text analysis; Syntax; Typology* – Ref.Nr.: 86-04439
- Lehrberger, J.: **Sublanguage analysis.** S. 19-38, *Terminological research; Artificial language; Language; Syntax; Comparison* – Ref.Nr.: 86-04440
- Friedman, C.: **Automatic structuring of sublanguage information.** S. 85-102, *Computational linguistics; Text analysis; Freetext searching; Data input; Format; Syntax; Medicine* – Ref.Nr.: 86-04441
- Hirschman, L.: **Discovering sublanguage structures.** S. 211-234, *Computational linguistics; Terminological research; Text analysis; Semantics; Relation; Classification characteristic; Pattern recognition* – Ref.Nr.: 86-04442
- Salton, G.; Zhang, Y.: **Enhancement of text representations using related document titles.** Information processing and management, 22(1986)5, S. 385-394, *Indexing techniques; Automatic indexing; Search strategy; Weighting; Relation; Citation; Evaluation; Statistical survey* – Ref.Nr.: 86-04473
- Paice, C.: **Expert systems for information retrieval?.** Aslib proceedings, 38(1986)10, S. 343-353, *Expert system; Information retrieval; Trend* – Ref.Nr.: 87-00122
- Connolly, J. H.; Edmonds, E. A.; Guzy, J. J.; Johnson, S. R.; Woodcock, A.: **Automatic speech recognition based on spectrogram reading.** International journal of man-machine studies, 24(1986)6, S. 611-621, *Speech recognition; Speech processing; Man-machine communication; Graphic representation; Pattern recognition* – Ref.Nr.: 87-00128
- Boot, M.: **BOBRA: an expert system for automated language description as a basis for statistical studies.** Literary and linguistic computing, 1(1986)2, S. 55-62, *Expert system; Computational linguistics; Implementation; Use; Statistical survey; Netherlands; BOBRA* – Ref.Nr.: 87-00148
- Hockey, S.: **OCR: the Kurzweil Data Entry Machine.** Literary and linguistic computing, 1(1986)2, S. 63-67, *Input device; Character reader; Data acquisition; Data input; Computational linguistics; UK* – Ref.Nr.: 87-00149
- Uchida, H.: **Fujitsu machine translation system: ATLAS.** Future generations computer systems, 2(1986)2, S. 95-100, *Automatic translation; Software; Syntax; Semantics; English; Japanese; ATLAS* – Ref.Nr.: 87-00157
- Nitta, Y.: **Problems of machine translation systems. Effect of cultural differences on sentence structure.** S. 101-115, *Automatic translation; Translation; Comparison; Trend; Japanese; English* – Ref.Nr.: 87-00158
- Nagao, M.; Tsujii, J.; Nakamura, J.: **Science and technology agency's Mu machine translation project.** S. 125-139, *Automatic translation; Software; Japanese; English* – Ref.Nr.: 87-00159
- Parrott, J. R.: **Expert systems for reference work.** Microcomputers for information management, 3(1986)3, S. 155-171, *Library automation; Expert system; Library service; Trend; University library; Canada* – Ref.Nr.: 87-00160
- Mercure, G.: **A personal MIS on microcomputer.** Microcomputers for information management, 3(1986)3, S. 199-216, *Library organization; Decision support system; Microcomputer; Hardware; Software; University library; Canada* – Ref.Nr.: 87-00162
- Brookes, B. C. (Hrsg.): **Intelligent information systems for the information society.** Proceedings of the sixth International Research Forum in Information Science (IRFIS 6), Frascati, Italy, September 1985. North-Holland (1986), 259 S., *Information retrieval; Artificial intelligence; User interface; Software* – Ref.Nr.: 87-00303
- Pollitt, A. S.: **Expert systems and the information intermediary. Tackling some of the problems of naive end-user search specification and formulation.** S. 100-108, *User interface; Software; Information retrieval; Artificial intelligence; User support; Descriptor; Expert system; Medicine; CANSEARCH; MEDLINE* – Ref.Nr.: 87-00308
- Brooks, H. M.: **Developing and representing problem descriptions.** S. 141-161, *User interface; Artificial intelligence; Information retrieval; Problem solving; Model; Semantic network* – Ref.Nr.: 87-00310

- Daniels, P. J.: **The user modelling function of an intelligent interface for document retrieval systems.** S. 162-176, *User; Model; Frame; User interface; Artificial intelligence; Information retrieval* – Ref.Nr.: 87-00311
- Michelet, B.; Turner, W.: **Co-word search: a system for information retrieval.** S. 177-191, *Information retrieval; Software; User interface; Search strategy; Semantic field; Statistics; Method Leximappe* – Ref.Nr.: 87-00312
- Hutchins, W. J.: **Machine translations. Past, present, future.** Horwood u. Halsted Press (1986) 382 S., *Automatic translation; Information system; Computational linguistics; Linguistics; History; Trend; USA; Europe; UK; USSR; Japan; China; Mexico* – Ref.Nr.: 87-00333
- Zarri, G. P.: **Constructing and utilizing large fact databases using artificial intelligence techniques.** S. 141-160, *Factual data bank; Database organization; Knowledge representation; Inference; Method; Project; Information retrieval; RESEDA* – Ref.Nr.: 87-00370
- Reimer, U.: **A system-controlled multi-type specialization hierarchy.** S. 173-187, *Knowledge representation; Frame; Semantics; Data structure; TOPIC* – Ref.Nr.: 87-00371
- Chakravarthy, U. S.; Fishman, D. H.; Minker, J.: **Semantic query optimization in expert systems and database systems.** S. 659-674, *Expert system; Database system; Search strategy; Semantics; Knowledge representation* – Ref.Nr.: 87-00386
- Fugmann, R.: **Die Funktion von semantischen Kategorien in Indexierungssprachen und bei der Indexierung.** In: Degens, P. O. (Hrsg.); Hermes, H. J. (Hrsg.); Opitz, O. (Hrsg.); Gesellschaft für Klassifikation: Die Klassifikation und ihr Umfeld. Proceedings, 10. Jahrestagung der Gesellschaft für Klassifikation, Münster/W., Juni 1986. INDEKS (1986), S. 51-60, *Indexing techniques; Semantics; Precoordination; Postcoordination; Relation* – Ref.Nr.: 87-00407
- Rescheleit, W.; Menner, L.: **Vergleich der Wissensrepräsentationssprache FRL mit Dezimalklassifikation und Facettenklassifikation.** S. 68-76, *Knowledge representation; Expert system; Frame; UDC; Faceted classification; Comparison* – Ref.Nr.: 87-00408
- Riesthuis, G. J. A.; Colenbrander-Dijkman, A. M.: **Subject access to central catalogues. Incompatibility issues of library classification systems and subject headings in subject cataloguing.** S. 120-128, *Information retrieval; Subject catalog; Online; Man-machine communication; Software; Search strategy* – Ref.Nr.: 87-00413
- Riekert, W. F.: **Systemkomponenten zum Wissenserwerb.** S. 203-228 *Systems research; Knowledge representation; Expert system; Draft; Fact documentation; User interface; Software; Dal* – Ref.Nr.: 87-00450
- Hahn, U.: **Curriculum Informationslinguistik. Zur inhaltlichen und didaktischen Konzeption informationeller Sprachanalyse am Beispiel des Diplom-Aufbaustudiums Informationswissenschaft an der Universität Konstanz.** Konstanz, März(1986), 61 S., *Training in information science; Computational linguistics; Text analysis; Automatic indexing; Automatic translation; Information retrieval; Artificial intelligence; Natural language processing; Training course; Draft; FRG* – Ref.Nr.: 87-00523
- Hahn, U.; Reimer, U.: **TOPIC essentials.** TOPIC-19-86, Konstanz, April(1986) 33 S., *Computational linguistics; Text analysis; Compression; Text; Automatic abstracting; German; TOPIC* – Ref.Nr.: 87-00526
- Hahn, U.: **A generalized word expert model of lexically distributed text parsing.** TOPIC-21-86, Konstanz, April(1986) 16 S., *Text analysis; Computational linguistics; Grammar; Frame; Model* – Ref.Nr.: 87-00527
- Thiel, U.; Hammwöhner, R.: **Graphical interaction with a full-text oriented information system. The retrieval component of the end user interface TOPOGRAPHIC.** TOPOGRAPHIC-5-86, Konstanz, Mai(1986), 19 S., *Information system; Fulltext database; Dialog service; Graphic representation; User interface; Text analysis; Computational linguistics; Artificial intelligence; TOPOGRAPHIC* – Ref.Nr.: 87-00528
- Musa, F. A.: **A system for processing bilingual Arabic/English text.** J.ASIS, 37(1986)5, S. 288-293, *Publication; Preparation; Text; Multilingual; Arabic; English; Text analysis; Character; Character set; Code; Mathematical model* – Ref.Nr.: 87-00543
- Berry, D. C.; Broadbent, D. E.: **Expert systems and the man-machine interface.** Expert systems, 3(1986)4, S. 228-231, *Expert system; User interface; Trend; UK* – Ref.Nr.: 87-00597
- Zifonun, G. (Hrsg.): **Vor-Sätze zu einer neuen deutschen Grammatik.** Narr (1986) 302 S., *Linguistics; Grammar; German; Semantics; Semantic field; Syntax* – Ref.Nr.: 87-00609
- Nickerson, R. S.: **Using computers. The human factors of information systems.** MIT Press (1986) 448 S., *Computer; Use; User; Society; Implication; Information system; Man-machine communication; Information technology; Telecommunications; Software; Artificial intelligence; Expert system; Human factor* – Ref.Nr.: 87-00613
- Davies, R. (Hrsg.): **Intelligent information systems. Progress and prospects.** Horwood u. Halsted Press (1986), 300 S., *Artificial intelligence; Expert system; Information retrieval; User interface; Library automation; Software; Model; Trend* – Ref.Nr.: 87-00654
- Coelho, H.: **Library manager. A case study in knowledge engineering.** S. 33-53, *Expert system; Linguistics; Artificial intelligence; Natural language processing; Portuguese; Portugal; TUGA* – Ref.Nr.: 87-00656
- Davies, R.: **Cataloguing as a domain for an expert system.** S. 54-77, *Expert system; Cataloging rules; Automatic cataloging; Knowledge representation; Trend; AACR2* – Ref.Nr.: 87-00657
- Lebowitz, M.: **An experiment in intelligent information systems: RESEARCHER.** S. 127-149, *Expert system; Patent documentation; Software; Knowledge representation; Natural language processing; Information retrieval; RESEARCHER* – Ref.Nr.: 87-00659
- Vickery, A.; Brooks, H. M.; Vickery, B. C.: **An expert system for referral: the PLEXUS project.** S. 154-183, *Expert system; Referral; Software; User interface; Natural language processing; Knowledge representation PLEXUS* – Ref.Nr.: 87-00660

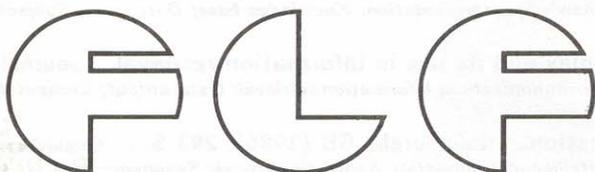
- Davies, R.: **Classification and ratiocination. A perennial quest.** S. 262-289, *Classification research; Classification system; History; Artificial intelligence; Trend* – Ref.Nr.: 87-00664
- Meijs, W.: **Lexical organization from three different angles.** ALLC journal, 6(1985)1/2, S. 1-10, *Lexicology; Computational linguistics; Dictionary; Model* – Ref.Nr.: 87-00715
- Brandon, F. R.: **Microcomputer software tools for a bilingual dictionary and an automatic bilingual dictionary.** S. 11-13, *Dictionary; Multilingual; Software; Microcomputer* – Ref.Nr.: 87-00716
- Juillard, M.: **A quantitative approach to semantic and morphosemantic fields in a literary work.** S. 14-23, *Computational linguistics; Vocabulary; Semantics; Literary studies; Frequency analysis* – Ref.Nr.: 87-00717
- Paikeday, T. M.: **Text analysis by microcomputer.** S. 29-32, *Text; Downloading; Microcomputer; Text analysis; Lexicology* – Ref.Nr.: 87-00718
- Jiaju, M.: **The function and formation of semantic systems. A new Chinese Thesaurus of Synonyms.** *Multilingua*, 5(1986)4, S. 205-208, *Linguistics; Synonym; Thesaurus; Chinese; China* – Ref.Nr.: 87-00726
- Ladd, P.: **Pre-editing of machine-translation input. An experimental evaluation.** S. 217-224, *Automatic translation; Data output; Error; Evaluation; Data input; Control (Check); Method* – Ref.Nr.: 87-00727
- Picken, C. (Hrsg.); Aslib: **Translation and communication.** Proceedings of a conference, November 1984, London, Translating and the computer 6: Translation and communication, (1985) 198 S., *Training; Translator; Database system; Terminology; Automatic translation; Translation; Word processing; Electronic mail; Speech recognition* – Ref.Nr.: 87-00744
- Bolc, L. (Hrsg.); Jarke, M. (Hrsg.): **Cooperative interfaces to information systems.** Springer (1986) 339 S., *Systems research; Artificial intelligence; Man-machine communication; User interface; Language; Software; DIAGRAM; HAM-ANS* – Ref.Nr.: 87-00751
- Thomas, P.; Ahmad, K.; Judge, A.; Barbour, S.; Reeves, N.: **The British Term Bank project.** Terminologie et traduction, Luxembourg, (1986)3, S. 45-100, *Database; Terminology; Database system; Data structure; Natural language processing; Artificial intelligence; English; UK; British Term Bank* – Ref.Nr.: 87-00753
- Arthur, J. D.: **A descriptive/prescriptive model for menu-based interaction.** International journal of man-machine studies, 25(1986)1, S. 19-32 *User interface; Software; Requirement; Man-machine communication; Model; User support; Smalltalk; Zog; Omni* – Ref.Nr.: 87-00755
- Habel, C.: **Prinzipien der Referentialität. Untersuchungen zur propositionalen Repräsentation von Wissen** Springer (1986) 317 S., *Artificial intelligence; Cognition; Knowledge representation; Artificial language; Logic; Semantics* – Ref.Nr.: 87-00760
- Hausser, R.: **NEWCAT: parsing natural language using left-associative grammar.** Springer (1986) 540 S., *Computational linguistics; Language; Grammar; Software; German; English; NEWCAT; DCAT; ECAT* – Ref.Nr.: 87-00761
- Dixon, G.: **An editor's expert system.** Literary and linguistic computing, 1(1986)3, S. 136-142, *Expert system; Software; Microcomputer; Publishers; SYNICS* – Ref.Nr.: 87-00801
- Hidley, G. R.: **Some thoughts concerning the application of software tools in support of old English poetic studies.** S. 156-162, *Software; Computational linguistics; Natural language processing; Clustering; Computer graphics; Text analysis; Literary studies MARK; CALCULIT* – Ref.Nr.: 87-00802
- Neuhaus, H. J.: **Phonetic character sets. Printing, sorting and computing.** S. 163-167, *Phonetics; Character set; Transcription; Code; Data input; Data output; Arrangement method* – Ref.Nr.: 87-00803
- Jochum, F.: **Wissensgestützte Benutzerschnittstellen für Informationssysteme. Bericht über das Projekt WISBI.** In: Müller, H.; Strauch, D. (Hrsg.): 5. KTK-Workshop 1986. Expertensysteme und Wissensorganisation in der Fachinformation. Berlin, (1986), S. 12-16, *User interface; Artificial intelligence; Software; Information retrieval; Microcomputer; Information retrieval system; Host* – Ref.Nr.: 87-00807
- Jackendoff, R.: **Semantics and cognition.** MIT Press (1986), 296 S., *Linguistics; Language; Cognition; Model; Semantics; Grammar; Lexicology* – Ref.Nr.: 87-00836
- Panyr, J.: **Information-Retrieval-Systeme. State of the art.** Handbuch der modernen Datenverarbeitung, 24(1987)133, S. 15-36, *Information retrieval system; Software; Indexing techniques; Automatic indexing; Searching; Ranking; Search strategy; Evaluation; GOLEM; STAIRS; GRIPS/DIRS; UNIDAS; FAIRS; STATUS; TELDOK; CONDOR* – Ref.Nr.: 87-00908
- Gebhardt, F.: **Dokumenten-Retrieval und Künstliche Intelligenz.** Handbuch der modernen Datenverarbeitung, 24(1987)133, S. 37-46, *Expert system; Information retrieval; Searching; User interface; Speech processing* – Ref.Nr.: 87-00909
- Capurro, R.: **Was ist Information? Hinweise zum Wort- und Begriffsfeld eines umstrittenen Begriffs.** Handbuch der modernen Datenverarbeitung, 24(1987)133, S. 107-114, *Information; Definition; Concept; Semantic field* – Ref.Nr.: 87-00914
- Lester, S.; Rada, R.: **A method of medical knowledge base augmentation.** Methods of information in medicine, 26(1987)1, S. 31-39, *de Automatic indexing; Medicine; Semantic network; Natural language processing; Method; Comparison; MeSH; CMIT* – Ref.Nr.: 87-00976
- Last, R. W.: **The potential of artificial intelligence-related CALL at the sentence level.** Literary and linguistic computing, 1(1986)4, S. 197-201, *Computer aided learning; Artificial intelligence; Foreign language; German; Software* – Ref.Nr.: 87-00978
- Weber, N.: **Computer-gestütztes strukturiertes Wortschatzlernen. Beispiel eines Lernprogrammes für fremdsprachliche Vokabeln aus dem nominalen Bereich.** Literary and linguistic computing, 1(1986)4, S. 207-213, *Computer aided learning; Vocabulary; Foreign language; Polish; Algorithm; Software; VOCARL* – Ref.Nr.: 87-00979

- Cooper, M.: **Human factor aspects of voice input/output.** S. 3-13, *Speech output; Speech recognition; Human factor* – Ref.Nr.: 87-00992
- Richter, W.; Steinschmidt, W.: **Applications and problems of voice i/o in office workstations.** S. 15-24, *Office organization; Speech processing; User interface; Speech input; Speech output* – Ref.Nr.: 87-00993
- Magnusson, L. E.: **Interactive speech response systems and its applications.** S. 41-55, *Speech processing; Computer; Host; Speech output; Interface; Telecommunications; Tell-Me 2000* – Ref.Nr.: 87-00994
- Viglione, S. S.: **Speech recognition: consumer products.** S. 109-122, *Speech recognition; Microcomputer; Telecommunications; Private household* – Ref.Nr.: 87-00996
- Pickup, M. N.: **Voice response.** S. 135-143, *Speech output; User interface; Telecommunications; Industry; UK Talk Back* – Ref.Nr.: 87-00997
- Meisel, W. S.: **Talkwriters: fact or fiction?.** S. 153-161, *Speech processing; Speech input; User; Requirement; Phonetics* – Ref.Nr.: 87-00998
- Picheny, M. A.; Das, S. K.: **A real-time IBM PC based large-vocabulary isolated-word speech recognizer.** S. 173-182, *Speech recognition; Word; Microcomputer; Dictionary; Office organization* – Ref.Nr.: 87-00999
- Hvelplund, H.: **The ESPRIT Eurohelp project. An intelligent help system.** In: Online International, Knowledge based systems KBS 86. Proceedings of the international conference held in London, July 1986. Online Publ. (1986) S. 61-70, *Computer aided learning; Artificial intelligence; User interface; Software; Electronic mail; EUROHELP* – Ref.Nr.: 87-01005
- Jenkins, P. G.: **TUTOR: a shell for intelligent computer aided instruction.** S. 71-80, *Computer aided learning; Software; Artificial intelligence; Student; TUTOR* – Ref.Nr.: 87-01006
- Stratil, M.; Burkhardt, D.; Chikken, P.: **Knowledge based intelligent video.** S. 81-86, *Computer aided learning; Videotape; Speech processing; Artificial intelligence; Foreign language* – Ref.Nr.: 87-01007
- Pettitt, S.: **Browsing is ... decision support.** S. 285-299, *User interface; Artificial intelligence; Information retrieval; Software; Alvey DHSS* – Ref.Nr.: 87-01016
- Black, W. J.: **Software for developing expert systems.** In: Gibb, F. (Hrsg.); Library Association (LA), Information Technology Group,; Library Association (LA), Library and Information Research Group,; Expert systems in libraries. Proceedings of a conference, November 1985.: Taylor Graham (1985) S. 22-35, *Expert system; Library automation; Software; User interface; Programming* – Ref.Nr.: 87-01030
- Davies, R.: **Expert systems and cataloguing. New wine in old bottles?.** S. 67-82, *Expert system; Cataloguing; System environment; Human factor; Trend* – Ref.Nr.: 87-01033
- Rieger, B. B.: **On generating semantic dispositions in a given subject domain. Stereotype representation and dynamic structuring of fuzzy word meanings for contents-driven semantic processing.** In: Agrawal, J. C. (Hrsg.); Zunde, P. (Hrsg.): Empirical foundations of information and software science. Proceedings of the second symposium, held in October, 1984, in Atlanta, Georgia. Plenum Press (1985) S. 273-291, *Natural language processing; Knowledge representation; Method; Semantics; Algorithm; Semantic network* – Ref.Nr.: 87-01076
- Brainerd, B.: **Textual analysis and synthesis by computer.** *Abacus*, 4(1987)2, S. 8-18, *Text analysis; Concordance; Language statistics; Software; Mainframe; Microcomputer; Literary studies; CLOC; BYU* – Ref.Nr.: 87-01101
- Haton, J. P. (Hrsg.): **Fundamentals in computer understanding. Speech and vision.** Cambridge Univ. Press (1987), 288 S., *Speech processing; Image analysis; Man-machine communication* – Ref.Nr.: 87-01123
- Grosz, B. J.; Sidner, C. L.: **Attention, intentions, and the structure of discourse.** *Computational linguistics*, 12(1986)3, S. 175-204, *Computational linguistics; Natural language processing; Model* – Ref.Nr.: 87-01134
- Grishman, R.; Hirschmann, L.; Nhan, N. T.: **Discovery procedures for sublanguage selectional patterns. Initial experiments.** *Computational linguistics*, 12(1986)3, S. 205-215, *Text analysis; Natural language processing; Testing* – Ref.Nr.: 87-01135
- Basu, A.; Dutta, A.: **Computer based support of reasoning in the presence of fuzziness.** *Decision support systems*, 2(1986)3, S. 235-256, *Systems research; Decision support system; Knowledge representation; Frame; Inference; Algorithm* – Ref.Nr.: 87-01180
- O'Connor, J.: **Computer use of a medical dictionary to select search words.** *Information processing and management*, 22(1986)6, S. 477-486, *Search strategy; Dictionary; Medicine; Word formation; Search result; Recall; Evaluation; Testing* – Ref.Nr.: 87-01227
- Zoeppritz, M.: **A framework for investigating language-mediated interaction with machines.** *International journal of man-machine studies*, 25(1986)3, S. 295-315, *Computational linguistics; Man-machine communication; Language; Programming language; Model* – Ref.Nr.: 87-01275
- Zvegintsev, V. A.: **Language as a factor in the computer revolution.** *Automatic documentation and mathematical linguistics* 19(1985)5, S. 1-11, *Artificial intelligence; Expert system; Language* – Ref.Nr.: 87-01279
- Driyanskii, V. M.; Zhogov, G. V.; Koltun, A. Ya.: **Use of quantitative estimators for paradigmatic relationships in a documentary information-retrieval system.** S. 59-80, *Mathematical model; Relation; Semantics; Information retrieval* – Ref.Nr.: 87-01285
- Leont'eva, N. N.: **An information model for a machine translation system.** S. 92-105, *Information science; Model; Automatic translation; French; Russian; USSR* – Ref.Nr.: 87-01287
- Nagao, M.; Tsujii, J.; Nakamura, J.: **Terminology dictionary for machine translation.** In: Krommer-Benz, M. (Hrsg.); Internationales Informationszentrum für Terminologie (Infoterm), Wien, AT: Networking in terminology. International co-operation in terminology work. Second Intoterm symposium. Proceedings. Saur (1986) S. 286-295, *Automatic translation; Dictionary compilation; Japanese; English* – Ref.Nr.: 87-01307

- Broens, U.: **A terminological data base system for the translation services at the Council of the European Communities.** S. 413-420, *Database system; Automatic translation; Translation service; EC; EC-countries* – Ref.Nr.: 87-01311
- Lehtinen, E.; Lyytinen, K.: **Action based model of information system.** *Information systems*, 11(1986)4, S. 299-317, *Information system; Decision support system; Model; Systems research; Draft; Semantics; Pragmatics; Information process; Communication process* – Ref.Nr.: 87-01319
- Spillner, B.: **Typen der semantischen Beschreibung im Wörterbuch.** In: Strohl-Göbel, H. (Bearb.); DGD: Deutscher Dokumentartag 1984. *Perspektiven der Fachinformation Programme - Praxis - Prognosen.* Darmstadt, vom Oktober 1984. Saur (1985) S. 262-272, *Lexicology; Dictionary; Semantics* – Ref.Nr.: 87-01353
- Link, E.: **Wortbildung im Fachwörterbuch.** S. 288-307, *Lexicology; Word formation; Guideline; Terminology* – Ref.Nr.: 87-01355
- Hüther, H.: **Zur Aufwandsabschätzung beim Aufbau eines automatischen Indexierungssystems.** S. 499-515, *Automatic indexing; Dictionary compilation; Language statistics; Document; Frequency analysis; Information retrieval; WAI; AIR; AIR/PHYS* – Ref.Nr.: 87-01364
- Walker, D. E.: **Knowledge resource tools for information access.** *Future generations computer systems*, 2(1986)3, S. 161-171, *Artificial intelligence; Information source; Knowledge base; Dictionary; Machine-readability; Dictionary compilation; Language statistics; Knowledge representation* – Ref.Nr.: 87-01378
- Vickery, A.; Brooks, H.; Robinson, B.; Vickery, B.: **A reference and referral system using expert system techniques.** *Journal of documentation*, 43(1987)1, S. 1-23 *Information retrieval system; Expert system; Prototype; Referral; Frame; Semantics; Database organization; PLEXUS; MIRABILIS* – Ref.Nr.: 87-01409
- Morrison, P. R.; Noble, G.: **Individual differences and ergonomic factors in performance on a videotex-type task.** *Behaviour and information technology*, 6(1987)1, S. 69-88, *Man-machine communication; Software; Ergonomics; User interface; Human factor; Empirical study; Curriculum; Training in computer science; Interactive videotex; Australia* – Ref.Nr.: 87-01417
- Lustig, G. (Hrsg.): **Automatische Indexierung zwischen Forschung und Anwendung.** Olms (1986), 195 S., *Automatic indexing; Research; Use; Input; Dictionary compilation; Information retrieval; Ranking; Weighting; WAI; AIR; AIR/PHYS; ALIBABA; DAISY* – Ref.Nr.: 87-01426
- Biswas, G.; Bezdek, J. C.; Marques, M.; Subramanian, V.: **Knowledge-assisted document retrieval. (Part 1). The natural-language interface.** *J.ASIS*, 38(1987)2, S. 83-96, *Systems research; Information retrieval; Natural language processing; Mathematical model; Semantic network* – Ref.Nr.: 87-01506
- Biswas, G.; Bezdek, J. C.; Subramanian, V.; Marques, M.: **Knowledge-assisted document retrieval. (Part 2). The retrieval process.** *J.ASIS*, 38(1987)2, S. 97-110, *Systems research; Information retrieval; Natural language processing; Semantic network; Mathematical model; Ranking; Inference; Search result; Prototype* – Ref.Nr.: 87-01507
- Philip, G.; Young, E. S.: **Man-machine interaction by voice: developments in speech technology. Part 1. The state-of-the-art.** *Journal of information science*, 13(1987)1, S. 3-14, *Man-machine communication; Speech recognition; Speaker identification; Speech input; Speech output; Trend* – Ref.Nr.: 87-01546
- Philip, G.; Young, E. S.: **Man-machine interaction by voice: developments in speech technology. Part 2. General applications, and potential applications in libraries and information services.** *Journal of information science*, 13(1987)1, S. 15-23, *Man-machine communication; Office organization; Speech recognition; Information and documentation; Trend* – Ref.Nr.: 87-01547
- Belyaeva, L. N.; Matorina, L. V.; Piotrovskii, R. G.; Yashenko, T. V.: **Semantic modules in a machine translation system. Complex term analysis.** *Automatic documentation and mathematical linguistics*, 19(1985)4, S. 52-61, *Automatic translation; Frame; Thesaurus structure; Term* – Ref.Nr.: 87-01575
- Aleksandrova, N. V.: **Linguistic analysis of an invention formula.** *Automatic documentation and mathematical linguistics* 19(1985)4, S. 93-100, *Patent; Text analysis; Syntax; Semantics* – Ref.Nr.: 87-01579
- Bevzenko, E. A.; Gal'perina, T. A.; Kruchenitskaya, E. A.; Kuznetsov, B. A.; Khoroshilov, A. A.: **Automatic compilation of frequency dictionaries for key words and word combinations from document search prescriptions.** *Automatic documentation and mathematical linguistics*, 19(1985)4, S. 113-119, *Dictionary compilation; Language statistics; Database; Science and technology; USSR* – Ref.Nr.: 87-01580
- Dietrich, R.: **Der Computer in den Geisteswissenschaften.** *IBM Nachrichten*, 37(1987)287, S. 15-23, *Computer; Humanities; Text analysis; Software; PRO TEXT* – Ref.Nr.: 87-01606
- Wilss, W. (Hrsg.); Schmitz, K. D. (Hrsg.): **Maschinelle Übersetzung – Methoden und Werkzeuge.** Akten des 3. Internationalen Kolloquiums 'Elektronische Sprachforschung' Saarbrücken, September 1986. Niemeyer (1987), 337 S., *Automatic translation; Computational linguistics; Natural language processing* – Ref.Nr.: 87-01679
- Agafonov, V. N.: **Basic concepts and definitions in indexing languages.** *Automatic documentation and mathematical linguistics* 19(1985)6, S. 1-11, *Documentary language; Semantics; Search strategy; Postcoordination; Compatibility* – Ref.Nr.: 87-01753
- Pshenichnaya, L. E.: **Relevance of a scientific document to an information query as regards the relation between the title and the complete text.** *Automatic documentation and mathematical linguistics* 19(1985)6, S. 24-31, *Title; Relevance; Text; Comparison; Information retrieval; Testing; Text analysis* – Ref.Nr.: 87-01754
- Mehl, S.: **Word Expert Parsing und Disambiguierung. Can inquiring experts be helpful?.** *LDV-Forum*, 4(1986)2, S. 3-7, *Natural language processing; Model; Word meaning; Semantics; Context* – Ref.Nr.: 87-01824
- Gust, H.: **Strukturiertes Wissen als Grundlage für Sprachverstehensprozesse.** S. 9-14, *Knowledge representation; Logic; Natural language processing; Knowledge base; Model* – Ref.Nr.: 87-01825
- Rösner, D.: **Systeme zur Generierung von Texten. Eine Übersicht.** S. 15-20, *Natural language processing; Software; Knowledge representation; Semantics; BABEL; IPG; KAFKA; PHRED; XPLAIN; TEXT; YH; MUMBLE; HAM-ANS; SEMTEX* – Ref.Nr.: 87-01826

- Mehdi, S. A.: **Arabic language parser.** *International journal of man-machine studies*, 25(1986)5, S. 593-611, *Natural language processing; Arabic; Grammar; Syntax* – Ref.Nr.: 87-01858
- Catach, N.: **The automatic phoneticization of the French language.** *Computers and the humanities*, 20(1986)3: special issue, S. 159-166, *Phonetics; French; Transcription; Automatic; Computational linguistics; France; HESOPHONIC* – Ref.Nr.: 87-01874
- Carbonell, N.; Haton, J. P.; Pierrel, J. M.: **Artificial intelligence in speech understanding. Two applications at C.R.I.N.** S. 167-172, *Speech processing; User interface; Speech recognition; Expert system; France* – Ref.Nr.: 87-01875
- Zingle, H.: **Linguistic preprocessing in vocal synthesis. Example of orthographic/phonetic transcription of German.** S. 173-177, *Text analysis; German; Computational linguistics; Transcription; Phonetics; Speech output; France* – Ref.Nr.: 87-01876
- Can, F.; Ozkarahan, E. A.: **Computation of term/document discrimination values by use of the cover coefficient concept.** *J.ASIS*, 38(1987)3, S. 171-183, *Automatic indexing; Indexing; Weighting; Clustering; Weighting; Method; Information science* – Ref.Nr.: 87-01881
- Humphrey, S. M.; Miller, N. E.: **Knowledge-based indexing of the medical literature. The Indexing Aid Project.** *J.ASIS*, 38(1987)3, S. 184-196, *Expert system; Medicine; Automatic indexing; Knowledge representation; Personnel; User participation; Project; Prototype* – Ref.Nr.: 87-01882
- Meunier, J. G.; Bertrand-Gastaldy, S.; Lebel, H.: **A call for enhanced representation of content as a means of improving online full-text retrieval.** *International classification*, 14(1987)1, S. 2-10, *Text analysis; Method; Typology; Fulltext database; Fulltext searching; Cognition; User interface* – Ref.Nr.: 87-01901
- Johansen, T.: **Elements of the non-linguistic approach to subject-relationships.** *International classification*, 14(1987)1, S. 11-18, *Text analysis; Model; Relation* – Ref.Nr.: 87-01902
- Mauch, H. U.; Yu, Y.; Samulowitz, H.: **MAP-DacS – ein mikrocomputergestützter Arbeitsplatz zum Aufbau und zur Nutzung von Datenbanken mit chinesischen Schriftzeichen.** *Nachrichten für Dokumentation*, 38(1987)2, S. 73-76, *Microcomputer; Information retrieval; Translation; Chinese; Arabic; Software* – Ref.Nr.: 87-02079
- Waters, S. T.: **Answerman, the expert information specialist. An expert system for retrieval of information from library reference books.** *Information technology and libraries*, 5(1986)3, S. 204-212, *Expert system; Information retrieval; Reference book; Agriculture; Software; Microcomputer; USA; Answerman* – Ref.Nr.: 87-02149
- Jaepinen, H.; Ylilampi, M.: **Associative model of morphological analysis. An empirical inquiry.** *Computational linguistics*, 12(1986)4, S. 257-272, *Text analysis; Computational linguistics; Morphology; Model; Finnish; Implementation; Empirical study* – Ref.Nr.: 87-02165
- Goodman, B. A.: **Reference identification and reference identification failures.** *Computational linguistics*, 12(1986)4, S. 273-305, *Man-machine communication; Natural language processing; Speech processing; Error; Correction; Method* – Ref.Nr.: 87-02166
- Herrmann, T.: **Zur Gestaltung der Mensch-Computer-Interaktion. Systemerklärung als kommunikatives Problem.** Niemeyer (1986) 275 S., *Man-machine communication; Design; Interface; User; Use; Communication theory; Human factor; Artificial intelligence* – Ref.Nr.: 87-02308
- Myers, J. J.; Kamemoto, D. Y.; Griffin, A. F.: **Artificial intelligence indexing. Creating knowledge bases of index terms ordered by semantic relations.** Long Beach, CA, US Dez.(1985) 33 S., *Artificial intelligence; Automatic indexing; Natural language processing; Knowledge representation; Knowledge base; Descriptor; Subject heading; Semantic network* – Ref.Nr.: 87-02311
- Bovey, J. D.; Brown, P. J.: **Interactive document display and its use in information retrieval.** *Journal of documentation*, 43(1987)2, S. 125-137, *Man-machine communication; Information retrieval; Data output; Computer graphics; Software; Trend* – Ref.Nr.: 87-02342
- Sparck Jones, K.: **Synonymy and semantic classification.** Edinburgh, GB (1986), 293 S., *Linguistics; Natural language processing; Computational linguistics; Artificial intelligence; Semantics; Semantic network; Synonym* – Ref.Nr.: 87-02355
- Kuhlen, R.; Hahn, U. (Mitarb.); Reimer, U. (Mitarb.): **Informationslinguistik. Theoretische, experimentelle, curriculare und prognostische Aspekte einer informationswissenschaftlichen Teildisziplin.** Niemeyer (1986), 266 S., *Information science; Computational linguistics; Information retrieval; Training in information science; Curriculum* – Ref.Nr.: 87-02439
- Kellermann, G.; Finke, E.; Hein, H. W.; Thomas, C.: **X-AiD: Eine wissensbasierte, anwendungs-unabhängige Mensch-Computer-Schnittstelle.** *GMD-Spiegel*, 17(1987)1, S. 23-29, *Man-machine communication; Software; Prototype; Knowledge base; Office organization* – Ref.Nr.: 87-02467
- Domenig, M.: **Entwurf eines dedizierten Datenbanksystems für Lexika. Problemanalyse und Software-Entwurf anhand eines Projektes für maschinelle Sprachübersetzung.** Niemeyer (1987) 253 S., *Automatic translation; Database system; Database organization; Dictionary; Computational linguistics; Software; Project; Prototype; Evaluation; Requirement; EC* – Ref.Nr.: 87-02493
- Lehmann, K. D. (Hrsg.); Strohl-Göbel, H. (Hrsg.); DGD: **The application of micro-computers in information, documentation and libraries.** *Proceedings of the second international conference, Baden-Baden, F.R.G, March, 1986 North-Holland (1987)*, 834 S., *Information and documentation; Librarianship; Archivistics; Microcomputer; Automation; Library automation; Archival automation; Hardware; Software; Computer graphics; Electronic publishing; Expert system; Training; Database; Information retrieval system; Economic efficiency; Decision support; Developing countries; Trend* – Ref.Nr.: 87-02536
- Harmon, P.: **Intelligent job aids. The use of small expert systems in office automation.** In: *Office Systems Research Association (OSRA), Houston, TX, US: A compendium of materials from the fifth*

- annual office systems research conference. March, 1986, (1986), S. 184-209, *Expert system; Office organization; Microcomputer; Trend* – Ref.Nr.: 87-02596
- Gilloux, P.; Servolle, B.: **The SOMIW workstation. Integration and speech management.** In: Commission of the European Communities (CEC), Directorate General XIII, Telecommunication, Information, Industries and Innovation, Luxembourg, LU: ESPRIT '86. Results and achievements. North-Holland (1987), S. 389-411, *Systems research; Distributed system; Operating system; Speech processing; Pattern recognition; Project ESPRIT; SOMIW* – Ref.Nr.: 87-02651
- Denoel, E.; Roelants, D.; Vauclair, M.: **Query translation for coupling prolog with a relational database management system.** In: Commission of the European Communities (CEC), Directorate General XIII, Telecommunication, Information, Industries and Innovation, Luxembourg, LU: ESPRIT '86. Results and achievements. North-Holland (1987) S. 595-605, *Systems research; Database system; Artificial intelligence; Software; Implementation; Integrated; Project* – Ref.Nr.: 87-02654
- Prangnell, R. D.: **MARKUP: an expert assistant for marking up bibliographic data.** Program, 21(1987)2, S. 174-182, *Automatic cataloging; Formal description; Formatting; Software; Expert system; MARKUP* – Ref.Nr.: 87-02700
- Cohen, R.: **Survey of computational linguistics courses.** Computational linguistics, 12(1986), S. 1-75, *Computational linguistics; Training; Training center; University education; Training course; Survey* – Ref.Nr.: 87-02701
- Evens, M.: **Directory of graduate programs in computational linguistics.** Computational linguistics, 12(1986), Graduate directory suppl., S. 1-52, *Computational linguistics; Training; Training center; University education; Training course; Curriculum; Computer* – Ref.Nr.: 87-02702
- Hoepelman, J. P.; Machate, J.: **Dialogue theory, theorem proving, data base questioning and natural language.** In: Commission of the European Communities (CEC), Information Technologies and Telecommunications Task Force, Bruxelles, BE: ESPRIT '85. Status report of continuing work. North-Holland (1986) S. 785-797, *Systems research; Natural language processing; Semantics; Man-machine communication; Model; Project* – Ref.Nr.: 87-02746
- Erlandsen, J.: **EUROHELP.PO, an intelligent help system for information systems.** In: Commission of the European Communities (CEC), Information Technologies and Telecommunications Task Force, Bruxelles, BE: ESPRIT '85. Status report of continuing work. North-Holland (1986) S. 809-824, *Artificial intelligence; Man-machine communication; Dialog service; Software; Prototype; Project ESPRIT; EuroHELP* – Ref.Nr.: 87-02748
- Knorz, G.: **Map on current research projects in the FRG related to the field of computational linguistics, July 1986.** Sprache und Datenverarbeitung, 10(1986)2, S. 5-12, *Computational linguistics; Research and development; University; Federal Republic of Germany* – Ref.Nr.: 87-02792
- Brietzmann, A.: **Semantische Analyse im Rahmen des Erlanger Spracherkennungssystems.** S. 13-21, *Semantic network; Case grammar; Pragmatics; Speech recognition; Federal Republic of Germany; EVAR* – Ref.Nr.: 87-02793
- Brückner, T.; Dickgiesser, S.; Kolvenbach, M.; Schmidt, R.; Teubert, W.; Wothke, K.: **Computational linguistics at the Institut für deutsche Sprache.** S. 22-29, *Computational linguistics; Lexicology; Dictionary compilation; Morphology; Lemmatization; Research and development; Federal Republic of Germany; Institut für deutsche Sprache; PLIDIS; LEDA; MOLEX; SYLT* – Ref.Nr.: 87-02794



Frankfurter Linguistische Forschungen

Nr. 3, Oktober 1987

Herausgeber: Helen Leuninger

THEORIE:

Frank Ernst Müller: Zur prosodischen Gliederung der Phrase *Ségmentée* im Französischen.

Peter Bertram, Jörg Keller, Erich Kiefer, Wolfgang Simanzik: Anmerkungen zur Metakognitionsforschung in Psychologie und künstlicher Intelligenz.

FORSCHUNGSPROJEKTE

AG Psycholinguistik und Aphasieforschung: ⊖ Theorie – Evidenz aus der Aphasie.

KLINISCHE LINGUISTIK:

Dorothea Evers-Volpp: Über Deblockierung – Erfahrungen mit einem Therapie-Konzept.

Michaela Zeh-Hau, Helen Leuninger: Berufsbild "Klinischer Linguist".

INTERVIEW

Matthias Hartig: Neue Wege in der Soziolinguistik.

NEUERSCHEINUNGEN

Brian Butterworth: Language Production.

Werner Radigk: Kognitive Entwicklung und zerebrale Dysfunktion.

Bestellungen sind an die Redaktion FLF, Institut für Deutsche Sprache und Literatur II, Johann Wolfgang Goethe-Universität, Gräferstraße 76, D-6000 Frankfurt a.M. 11 zu richten. Das Einzelheft kostet DM 2,50.

Redaktion FLF, Frankfurt

Sonderforschungsbereich 314
Künstliche Intelligenz – Wissensbasierte Systeme

KI – Labor am Lehrstuhl für Informatik IV

Leitung: Prof. Dr. W. Wahlster

Universität des Saarlandes

FB 10 Informatik IV

Im Stadtwald 15

D – 6600 Saarbrücken 11



KI – Labor
SFB 314

Neue Berichte und Memos

Einzelne Exemplare der folgenden Berichte und Memos können kostenlos bei Frau D. Borchers unter der o.g. Adresse angefordert werden:

Bericht Nr. 5

Kemke, C.: Entwurf eines aktiven, wissensbasierten Hilfesystems für SINIX.

Gekürzte Version in: LDV-Forum. Nr. 2, Dezember 1985, 43 – 60.

Abstract: Das System SC ist ein intelligentes Hilfesystem für das Betriebssystem SINIX. SC soll sowohl natürlichsprachliche Fragen des Benutzers über Konzepte und Kommandos des SINIX-Systems beantworten, als auch aktiv, ohne explizite Aufforderung durch den Benutzer Ratschläge bei der Durchführung von Aufgaben erteilen. SC verwendet hierzu sprachliches Wissen, Planungswissen, Wissen über die SINIX-Domäne und den Benutzer, das in Form separater Wissensquellen gespeichert ist. In vorliegendem Bericht werden nach einer kurzen Einführung über intelligente Hilfesysteme die Systeme UC, UCC und WIZARD vorgestellt, deren Domäne das Betriebssystem UNIX bzw. VMS ist. Daran schließt sich eine Darstellung der prinzipiellen Arbeitsweise und Struktur des SINIX Consultants an, in der die Aspekte Planung/Problemlösung und Wissensrepräsentation besondere Beachtung finden. Die Funktionen und Aufgaben der einzelnen Komponenten (nat.-sprachl. Schnittstelle, Problemlösungs-/Planungskomponente, Planerkennungskomponente) und Wissensbasen (sprachl. Wissen, Planungswissen, Domänenwissen, Benutzermodell) werden entsprechend dem derzeitigen Entwicklungsstand beschrieben. Abschließend werden wesentliche Merkmale der Implementierung und der momentane Projektstatus angeführt.

Bericht Nr. 6

Wahlster, W.: The Role of Natural Language in Advanced Knowledge-Based Systems

In: Winter, H. (Hrsg.): Artificial Intelligence and Man-Machine Systems. Heidelberg: Springer 1986, 62 – 83.

Abstract: Natural language processing is a prerequisite for advanced knowledge-based systems since the ability to acquire, retrieve, exploit, and present knowledge critically depends on natural language comprehension and production. Natural language concepts guide the the interpretation of what we see, hear, read or experience with other senses. In the first part of the paper, we illustrate the needed capabilities of cooperative dialog systems with a detailed example: the interaction between a customer and a clerk at an information desk in a train station. It is shown that natural language systems cannot just rely on knowledge about syntactical and semantical aspects of language but also have to exploit conceptual and inferential knowledge, and a user model. In the remainder, we analyze and evaluate three natural language systems which were introduced to the commercial market in 1985: Language Craft™ by Carnegie Group Inc., NLMenu by Texas Instruments Inc., and Q & A™ by Symantec Inc.. The detailed examination of these systems shows their capabilities and limitations. We conclude that the technology for limited natural-language access systems is available now, but in foreseeable future the capabilities of such systems in no way match human performance in face-to-face communication.

Bericht Nr. 7

Kobsa, A., Allgayer, J., Reddig-Siekmann, C., Reithinger, N., Schmauks, D., Harbusch, K., Wahlster, W.: Combining Deictic Gestures and Natural Language for Referent Identification

In: COLING '86. 11th International Conference on Computational Linguistics. Proceedings. Bonn 1986, 356 – 361.

Abstract: In virtually all current NL dialog systems, users can refer to objects by linguistic descriptions only. In human face-to-face conversation, however, participants also frequently use various sorts of deictic gestures. In this paper, we present the referent identification component of XTRA, a natural-language access system for expert systems. XTRA allows the user to combine NL input with pointing gestures on the terminal screen for referring to objects on the display. Information about the location and the type of this deictic gesture, as well as about the linguistic description of the referred object, the case frame and the dialog memory are utilized for identifying the object. The system is tolerant in respect to impreciseness of both the deictic and the natural language input. The user can thereby refer to objects more easily, avoid referential failures, and employ vague everyday terms instead of precise technical notions.

Bericht Nr. 8

Allgayer, J.: Eine Graphikkomponente zur Integration von Zeigehandlungen in natürlichsprachliche KI-Systeme.

In: GI - 16. Jahrestagung. Proceedings, Bd. 1. Berlin 1986, 284 - 298.

Abstract: Die Kombination von Sprache und Zeigen und damit das Schaffen einer natürlicheren Kommunikationsumgebung liefert in verschiedenen Bereichen eine Erhöhung der Akzeptanz von KI-Systemen. Das hier beschriebene System unterstützt als Teil einer natürlichsprachlichen Schnittstelle zu Expertensystemen den Aufbau einer Umgebung, in der Sprache und sprachbegleitende Zeigegesten kombiniert und kontextabhängig weiterverarbeitet werden. Innerhalb der auf dem Bildschirm visualisierten Umgebung kann eine Zeigegeste benutzt werden, die aus einem definierten Spektrum als für die gewünschte Äußerung geeignet ausgewählt wird. In vielen Fällen erlaubt erst die Kombination der Informationen aus der Analyse der Sprache und Gesten, das vom Benutzer intendierte Objekt zu extrahieren. Es besteht keine feste Zuordnung der Zeigepositionen innerhalb des Bildschirms zu den abgebildeten Objekten.

Bericht Nr. 9

Andre, E., G. Bosch, G. Herzog und T. Rist: Coping with the Intrinsic and Deictic Uses of Spatial Prepositions.

In: Jorrand, Ph. and V. Sgurev (ed.): Artificial Intelligence II. Proceedings of AIMS-86. Amsterdam: North-Holland 1987, 375 - 382.

Abstract: This paper deals with the definition of a computational semantics for spatial relations between objects in a scene. In particular we examine the effects of the observer's position on the computation of spatial relations, the use of path prepositions, and degrees of applicability for spatial relations. Representational prerequisites and the computational analysis of spatial relations are investigated in the German dialogue system CITYTOUR.

Bericht Nr. 10

Schmauks, D.: Form und Funktion von Zeigegesten, Ein interdisziplinärer Überblick.

Abstract: Zeigegesten, obwohl für mehrere Disziplinen interessant, sind bisher nie eingehender untersucht worden. Ihre theoretischen Grundlagen betreffen auch die KI, da neuerdings in bestimmten Diskursbereichen Zeigegesten auf einem Bildschirm simuliert werden können. Im folgenden werden Zeigegesten und ihre Interaktion mit Sprache unter verschiedenen Blickwinkeln untersucht. Kapitel 2 listet einige biologische Voraussetzungen der Verwendung von Zeigegesten auf. Dann werden Zeigegesten als Zeichen (Kapitel 3, Semiotik) bzw. als Mittel der nonverbalen Kommunikation (Kapitel 4, Psychologie) betrachtet. Im 5. Kapitel werden verschiedene Funktionen von Gesten im Hinblick auf Sprache unterschieden, ferner geht es um die syntaktische Einbettung obligatorischer Zeigegesten. Einige Parameter zur Klassifikation von Zeigegesten schlägt das 6. Kapitel vor. Das 7. Kapitel ist dem Beweis der These gewidmet, daß die Kombination von Zeigegesten und Sprache nicht nur besonders leistungsfähig, sondern auch häufig unumgänglich ist. Die beiden letzten Kapitel skizzieren einige verbleibende Probleme und Überlegungen zu ihrer Lösung.

Bericht Nr. 11

Kemke, C.: The SINIX-Consultant - Requirements, Design and Implementation of an Intelligent Help System for a UNIX Derivative.

Erscheint in: User Interfaces. Proceedings of the International Conference of the Gottlieb Duttweiler Institut. Rüschlikon/Zürich 1986.

Abstract: Inhalt dieses Berichtes ist die Beschreibung der Entwicklung des Systems SINIX Consultant

SC, eines intelligenten Hilfesystems für das Betriebssystem SINIX. SC soll einerseits natürlich-sprachliche Fragen über Konzepte und Kommandos des SINIX Systems beantworten. Andererseits soll es als aktives Hilfesystem das Verhalten des Benutzers beobachten und unaufgefordert Ratschläge für eine effizientere Arbeitsweise erteilen. SC ist somit ein Hilfesystem, das den Benutzer sowohl aktiv als auch passiv unterstützen soll. Um diese Anforderungen zu erfüllen, muß das System über die Verarbeitung natürlicher Sprache hinaus Tutoring- und Dialogführungsfähigkeiten besitzen, über ein Modell des Benutzerwissens verfügen und die aktuellen Pläne und Ziele des Benutzers erkennen können. Ein detailliertes Modell der Domäne, des SINIX Systems, ist notwendig, um die Problemlösungs- und Erklärungsaufgaben zu bewältigen, die für die Formulierung von adäquaten Antworten und Ratschlägen erforderlich sind.

Bericht Nr. 12

Allgayer, J. und C. Reddig: Processing Descriptions Containing Words and Gestures.
In: Rollinger, C.-R. und W. Horn (ed.): GWAI-86 und 2. Österreichische Artificial-Intelligence-Tagung. Proceedings. Berlin/Heidelberg: Springer 1986, 119 – 130.
Abstract: siehe Memo Nr. 11 (deutsche Version des Berichts)

Bericht Nr. 13

Reithinger, N.: Generating Referring Expressions and Pointing Gestures.
In: Kempen, G. (ed.): Natural Language Generation. Dordrecht: Nijhoff 1987, 71 – 81.
Abstract: Der Generator eines natürlichsprachlichen Systems muß, unter anderem, auch die Eingabe des Benutzers berücksichtigen. Die damit vom Generator erzeugte Ausgabe wird den Bedürfnissen des Benutzers besser angepaßt sein. Deshalb sollten der Generator und die anderen Teile des Systems Wissensbasen gemeinsam benutzen. In diesem Beitrag werden die ersten Überlegungen für den Entwurf der Generierungskomponente von XTRA – einem natürlichsprachlichen Zugangssystem zu Expertensystemen – dargestellt. Der Generator ist in das Gesamtsystem integriert und verwendet dessen Wissensbasen. Der Vorteil dieser Struktur wird am Beispiel der Generierung von referierenden Ausdrücken gezeigt. Als außersprachliche Mittel der Referenz können sowohl der Benutzer als auch das System Zeigegesten auf dem Bildschirm verwenden. Im letzten Teil des Beitrags wird aufgezeigt, wie Zeigegesten von dem Generierungssystem erzeugt werden können.

Bericht Nr. 14

Jansen-Winkel, R. M.: LEGAS - Inductive Learning of Grammatical Structures.
In: Hallam, J. and C. Mellish (ed.): Advances in Artificial Intelligence. Proceedings of the AISB Conference. Chichester: Wiley 1987, 169 – 181.

Abstract: This paper describes the theoretical approach to acquiring simple grammar rules from examples and its realisation in a program called LEGAS (LEarning GrAMmatical Structures). The resulting rule set is developed incrementally through gradual generalisation and specialisation. After each learning cycle the rule set achieves a status called "complete" and "consistent" which guides the flow of control.

Bericht Nr. 15

Werner, M.: RMSAI - Ein Reason Maintenance System für approximative Inferenzen.
Abstract: In dieser Arbeit wird eine Möglichkeit aufgezeigt, wie ein Reason Maintenance System in einem Inferenzmodell mit approximativen Inferenzen und Mehrfachableitungen die Konsistenz einer Aussagenmenge bezüglich einer Prämissen – und Regelmengen aufrecht erhalten kann. Hierzu wird ein Dependenznetz definiert, das alle bekannten Informationen über Aussagen und ihre Zusammenhänge (Justifications) repräsentiert, und ein Maintenance – Algorithmus vorgestellt, der Veränderungen bei einzelnen Knoten des Netzes zu allen davon betroffenen Nachfolgern propagiert. Schließlich wird ein Konsistenzbegriff für ein Dependenznetz mit evidenzbewerteten Knoten definiert und gezeigt, wie Inkonsistenzen im Dependenznetz aufgelöst werden können.

Bericht Nr. 16

Schmauks, D.: Natural and Simulated Pointing - An Interdisciplinary Survey.
In: Proceedings of the 3rd European ACL Conference, Kopenhagen, Danmark 1986, 179 – 185.
Abstract: Referent identification in human conversation is performed both by describing the objects in question and by pointing at them. Up till now, only the linguistic component could be simulated in dialog systems. But recently, technical innovations have made it possible to 'point' at the objects

on a display as well. The paper has two intentions. First, it investigates natural pointing in more detail and offers some possibilities to classify the great variety of pointing actions. Then, it tries to clarify the extent to which pointing by technical means (especially mouse-clicks) can be regarded as a simulation of natural pointing (or as a functional equivalent). Furthermore, some steps towards even more accurate simulation are briefly mentioned.

Bericht Nr. 17

Zimmermann, G., C. K. Sung, G. Bosch, J.R.J. Schirra: From Image Sequences to Natural Language: Descriptions of Moving Objects.

Abstract: We present our work concerning the connection of a vision system to a natural language system. That is, automatic processing transforms the original sequence of TV images into natural language descriptions concerning moving objects. It is the first time, that this transformation is achieved totally by computer. A vision system which has been developed in Karlsruhe is introduced in section 2. By means of analysing displacement vector fields, trajectories of object candidates are recognised. In section 3, the natural language system CITYTOUR is briefly presented. The verbalisation of spatial relations between static and moving objects can be studied with this system. Section 4 deals with the present state of the connection. The resulting data of the vision system are partially used and verbalised by CITYTOUR.

Bericht Nr. 18

Andre, E., Rist, T. und G. Herzog: Generierung natürlichsprachlicher Äußerungen zur simultanen Beschreibung von zeitveränderlichen Szenen.

In: Morik, K. (ed.): GWAI-87. 11th German Workshop on Artificial Intelligence. Proceedings. Berlin/Heidelberg: Springer 1987, 330 – 338.

Abstract: Die automatische Erzeugung simultaner Beschreibungen zeitveränderlicher Szenen bringt eine in bisherigen Generierungssystemen nicht berücksichtigte Problematik zum Vorschein. Da die Beschreibung nicht a posteriori, sondern simultan zum Ablauf der Szene erfolgt, ist der vollständige Verlauf der Szene zum Zeitpunkt der Sprachproduktion unbekannt. Bei der Koordination von Wahrnehmung und Sprachproduktion sind zeitliche Aspekte wie Dauer der Textgenerierung und Dekodierzeit des Hörers bzw. Lesers zu berücksichtigen. Der vorliegende Bericht untersucht die sich durch die Situation der Simultanbeschreibung ergebenden Konsequenzen in Bezug auf die Frage, welche Information über die Szene in welcher Form mitgeteilt werden soll. Anhand zweier Beispielszenen wird das Zusammenwirken von Ereigniserkennung und Sprachproduktion in dem von uns entwickelten System SOCCER erläutert.

Bericht Nr. 19

Rist, T., Herzog, G. und E. Andre: Ereignismodellierung zur inkrementellen High-level Bildfolgenanalyse.

In: Buchberger, E. und J. Retti (ed.): 3. Österreichische Artificial-Intelligence-Tagung. Proceedings. Berlin/Heidelberg: Springer 1987, 1 – 11.

Abstract: Dieser Bericht befaßt sich mit dem Problem der inkrementellen Erkennung von Ereignissen in zeitveränderlichen Szenen. Als Voraussetzung für die Anwendung einer inkrementellen Erkennungsstrategie wird das prototypische Auftreten eines Ereignisses durch ein Ablaufschema beschrieben. Ausgehend von einem diskreten Zeitmodell werden spezielle Prädikate definiert, mit deren Hilfe auch die zum Beobachtungszeitpunkt gerade auftretenden Ereignisse in einer formalen Zeitlogik repräsentierbar sind. Einerseits können diese Prädikate innerhalb des Ereignismodells zur Definition komplexerer Ablaufschemata herangezogen werden, andererseits erscheint uns die in propositionaler Form gegebene Information über gerade auftretende Ereignisse als eine geeignete Schnittstelle zwischen Ereigniserkennung und anderen Komponenten eines Systems zur natürlichsprachlichen Beschreibung zeitveränderlicher Realweltszenen.

Bericht Nr. 20

Beiche, H.-P.: LST-1 – Ein wissensbasiertes System zur Durchführung und Berechnung des Lohnsteuerjahresausgleichs.

In: Buchberger, E. und J. Retti (ed.): 3. Österreichische Artificial-Intelligence-Tagung. Proceedings. Berlin/Heidelberg: Springer 1987, 92 – 103.

Abstract: LST-1 ist ein wissensbasiertes System auf dem Gebiet "Lohnsteuerrecht". Es soll den Benutzer bei der Durchführung seines Lohnsteuerjahresausgleichs unterstützen, indem es durch ge-

eignete Konsistenzüberprüfungen ein falsches Ausfüllen der Formulare verhindert und ihn beim Geltendmachen von Absetzposten berät. LST-1 bietet dem Benutzer wahlweise eine freie Eingabe in das auf dem Bildschirm dargestellte Formular oder eine rechnergestützte Formularführung. Teilergebnisse können sofort berechnet werden, sobald die dafür benötigten Daten eingetragen wurden. Entwicklungs- und Ablaufumgebung des Systems ist BABYLON, eine von der GMD entwickelte Expertensystem-Shell. In den nun folgenden Kapiteln werden zunächst die für die Entwicklung von LST-1 relevanten Konzepte der XPS-Shell – sowie der dazu notwendigen Erweiterungen und Änderungen – beschrieben; darauf aufbauend folgt die detaillierte Darstellung des Systems.

Memo Nr. 8

Finkler, W. und G. Neumann: MORPHIX - Ein hochportabler Lemmatisierungsmodul für das Deutsche.

Abstract: Die morphologische Analyse stellt die erste Stufe bei der Verarbeitung geschriebener natürlicher Sprache dar. Die Wichtigkeit dieses Teilprozesses hängt von der zu analysierenden Sprache ab. Da die deutsche Sprache einer Sprachklasse mit weitgehend freier Wortstellung angehört, kann grammatikalische Information für ein Wort weniger durch seine Stellung im Satz als durch Analyse seiner Flexionsform gewonnen werden. In Lemmatisierungskomponenten werden flektierte Wortformen auf kanonische Formen zurückgeführt und ihre grammatikalische Information bestimmt, um die Größe der Lexika in natürlichsprachlichen Systemen zu reduzieren. In dieser Arbeit wird der portable Lemmatisierungsmodul MORPHIX vorgestellt, der in beliebige auf Lisp basierende natürlichsprachliche Systeme für das Deutsche integriert werden kann. Nach einer Analyse der Aufgabenstellung und der Einordnung von Lemmatisierungsverfahren in den Gesamtprozeß *Verstehen natürlicher Sprache* wird zuerst auf die Organisation des Lexikons sowie auf die Leistung von MORPHIX eingegangen. Dabei wird detailliert beschrieben, welche Wortformen regelgeleitet bearbeitet werden. Soll das System als unabhängiger Modul eingesetzt werden, so sind hohe Effizienzanforderungen zu stellen. Auf diese Anforderungen wird in einer ausführlichen Beschreibung der Funktionsweise des Algorithmus eingegangen. Eine zentrale Stellung in dieser morphologischen Analyse besitzt der GRIN – BAUM (GRAMMATIKALISCHE INFORMATION), in dem in effizienter Weise der Zusammenhang zwischen Flexionsendungen und zugehöriger grammatikalischer Information codiert ist. Kann eine zu untersuchende Wortform durch den Algorithmus nicht bearbeitet werden, so wird ein Klärungsdialog aktiviert. In diesem wird interaktiv mit dem Benutzer eine halbautomatische Lexikonerweiterung vorgenommen.

Memo Nr. 9

Portscheller, R.: AIDA - Rekursionsbehandlung, Konfliktauflösung und Regelcompilierung in einem Deduktiven Datenbanksystem.

Abstract: Im ersten Kapitel der vorliegenden Arbeit wird ein Überblick über die Entwicklung von herkömmlichen Relationalen DBS zu Deduktiven DBS bzw. Expertendatenbanksystemen gegeben. Neben der Darstellung grundsätzlicher Realisierungsansätze wird Wert auf eine exakte Definition Definiten und Indefiniten Deduktiven DBS gelegt. Das zweite Kapitel befaßt sich einerseits mit dem Verhältnis der "logischen" Programmiersprache PROLOG zu Definiten Deduktiven DBS; andererseits werden zwei als Frontends zu bestehenden Relationalen Datenbanksystemen realisierte Definiten Deduktive DBS besprochen. Das System AIDA wird im dritten Kapitel vorgestellt. Zur Optimierung existenzquantifizierter Anfragen wird ein a-priori Konfliktindex vorgeschlagen. Es werden überblickartig einige wichtige Arbeiten zur Behandlung rekursiver Regeln besprochen. Ein iterativer Fixpunktalgorithmus, der eine Propagierung von in der Anfrage enthaltenen Konstanten erlaubt und redundante Berechnungen vermeidet, wird vorgeschlagen. Weiterhin besteht in AIDA die Möglichkeit zur anfrageunabhängigen Regelcompilierung. Es wurde ein Algorithmus entwickelt, der die Aufnahme von Regeln in beliebiger Reihenfolge erlaubt, wobei nur solche bereits vorher bekannten Regeln noch einmal compiliert werden, die von dieser Änderung auch wirklich betroffen sind. Die Lösung der dabei auftretenden Konsistenzprobleme wird beschrieben. Das letzte Kapitel biete eine Zusammenfassung der Ergebnisse und zeigt einige mögliche Weiterentwicklungen des Systems auf.

Memo Nr. 10

Schirra, J.: MEGA-ACT - Eine Studie über explizite Kontrollstrukturen und Pseudoparallelität in Aktor-Systemen mit einer Beispielarchitektur in FRL.

Abstract: MEGA-ACT ist ein System, mit dessen Hilfe der Aufbau expliziter Kontrollstruktu-

ren und damit konstruierter deklarativer Programme in Aktor-Systemen untersucht wird. Ausgangspunkt der Untersuchung ist die Schwierigkeit, die von ihrem Wesen her massiv parallelen Aktor-Systeme mit "normalen" Computern mit nur einem Prozessor zu realisieren. Die inhärente Reflexivität des den Aktor-Systemen zugrundeliegenden Sprachgebrauchs, der sogenannten Aktor-Metapher, führt uns dazu, eine Sequentialisierung der simultanen Tätigkeiten der Aktoren vom System selbst vornehmen zu lassen. Das setzt voraus, daß die Programme, die diese Tätigkeiten vorschreiben, deklarativ dargestellt, d.h. dem System in ihrer Struktur zumindest teilweise zugänglich sind. Demzufolge wird als Kernstück dieser Arbeit ein Komplex aus fünf Aktoren vorgestellt, der ein allgemeines Schema für explizite Kontrollstrukturen bildet. Es handelt sich anders ausgedrückt um einen allgemeinen Baustein zur Konstruktion deklarativer Programme in Aktor-Systemen. Verschiedene Ausprägungen dieses sogenannten Programm-Moduls, die unterschiedliche Kontrollstrukturen realisieren, werden vorgestellt und schließlich an einem Beispiel, dem Parsen mit ATNs, in Anwendung gezeigt.

Memo Nr. 11

Allgayer, J. und C. Redding: Systemkonzeption zur Verarbeitung kombinierter sprachlicher und gestischer Referentenbeschreibungen.

Abstract: In der Kommunikation zwischen Menschen werden häufig Zeigegesten zur Vereinfachung und Beschleunigung des Dialogs benutzt. Mit diesem Bericht stellen wir einen Ausschnitt einer Systemkonzeption vor, die in der Mensch-Maschine-Kommunikation diese Kombination von natürlicher Sprache und Deixis zum Bezug auf Dinge, die der Bildschirm zeigt, ermöglicht. Primäre sprachliche Referenzträger sind Nominalphrasen; das System verarbeitet daher die sprachbegleitende Zeigegeste (die mit der "Maus" auf einer Lisp-Maschine realisiert ist) in Form eines syntaktischen Zeichens als Teil einer NP. Die morphologisch aufbereitete und syntaktisch analysierte Eingabe wird von der Satzsemantischen Analyse in eine sprachnahe, domainneutrale Funktionalsemantische Struktur abgebildet. Mit Hilfe von potentiell domainspezifischen Übersetzungsregeln evaluiert die Referenzsemantische Interpretation diese Struktur hinsichtlich der begleitenden Zeigegeste, der Relation zum Dialoggedächtnis und des assoziierten individualisierten Teils der aktuellen Wissensbasis. Dieser Prozess der Referentenidentifikation bestimmt (oder erzeugt) als Ergebnis das vom Sprecher intendierte Referenzobjekt.

Memo Nr. 12

Herzog, G.: Ein Werkzeug zur Visualisierung und Generierung von geometrischen Bildfolgenbeschreibungen.

Abstract: Bei der Entwicklung von Systemen zur natürlichsprachlichen Beschreibung von zeitveränderlichen Szenen stellt sich das Problem, daß bis heute noch kein bildverstehendes System in der Lage ist, eine nahezu vollständige geometrische Szenenbeschreibung für eine Realweltbildfolge zu erzeugen. Zur Beurteilung der Performanz des natürlichsprachlichen Zugangssystems durch den Benutzer und die Systementwickler ist es weiterhin notwendig, die visuellen Eingabedaten von ihrer numerischen in eine adäquate graphische Darstellung zu transformieren. Für die Domäne SOCCER des Projektes VITRA, in der Fussballspiele analysiert werden sollen, wurde daher das hier beschriebene Werkzeug entwickelt. Mit Hilfe dieses Programms kann zum einen eine Folge geometrischer Szenenbeschreibungen graphisch re-repräsentiert werden, zum anderen können interaktiv Bildfolgen vom Benutzer generiert werden. Die volle Ausnutzung der Programmierumgebung auf der SYMBOLICS LISPmaschine (Fenstersystem mit hochauflösender Graphik, MAUS-unterstütztes Menuesystem) ermöglichte dabei die Gestaltung einer höchst komfortablen Benutzeroberfläche.

Memo Nr. 13

Retz-Schmidt, G.: Deictic and Intrinsic Use of Spatial Prepositions: A Multidisciplinary Comparison.

In: Kak, A. and S.-S. Chen (ed.): Spatial Reasoning and Multisensor Fusion, Proceedings of the 1987 Workshop. Los Altos, CA: Morgan Kaufmann, 1987.

Abstract: In this paper, principles involving the deictic and intrinsic use of spatial prepositions are examined from the viewpoint of linguistic, psychological, and Artificial Intelligence approaches. After a brief introduction to the natural-language dialog system CITYTOUR, important concepts with respect to the deictic and intrinsic use of prepositions are defined. In the following section, those prepositions that permit deictic as well as intrinsic use are listed, and the way CITYTOUR copes with them is explained. Then, the identification of the front, back, left, and right regions of

a reference object is looked at in more detail. Finally, strategies concerning the deictic and intrinsic use of spatial concepts are pointed out.

Memo Nr. 14

Harbusch, K.: A First Snapshot of XTRAGRAM, A Unification Grammar for German Based on PATR.

Abstract: This paper describes XTRAGRAM, an almost purely syntactic grammar for parsing German sentences. The grammar is written in the PATR formalism and developed using the D-PATR workbench on a XEROX 1108 workstation. The discourse begins with an introduction to the PATR formalism in order to also allow people unfamiliar with unification to read the grammar. In the following section, XTRAGRAM is presented. First, all category and feature names are defined and all possible feature paths for the morphological categories are described. Then, the rules commended by their constructable feature paths, the words, the stems and the templates are defined. The linguistic coverage of the presented grammar is demonstrated by a variety of example DAG's. Finally, the paper closes with personal impressions received in working with the D-PATR workbench.

Memo Nr. 15

Fendler, M. und R. Wichlacz: SYCON - Symbolische Constraint-Propagierung auf Netzwerken, Entwurf und Implementierung

Abstract: Die Constraint-Propagierung auf Netzwerken hat sich bereits in mehreren Teilgebieten der KI als Verarbeitungsmechanismus bewährt. Es wird das Rahmensystem SYCON vorgestellt, mit dem beliebige symbolische Constraints definiert und dargestellt werden können. In den ersten vier Kapiteln der Arbeit werden die Definitionen von Constraints, lokaler Propagierung und Substitution, sowie die Funktionsweise der Propagierung erläutert. Die Kapitel 5 bis 7 enthalten die Beschreibung der Auflösung und Kompaktierung von Zyklen, der Arbeitsweise des Backtracking-Mechanismus und der Implementierung von SYCON.

The Joint Research Project WISBER: New Reports

The following new reports are available and can be ordered free of charge from the above address:

Dr. Johannes Arz, Universität des Saarlandes, FR. 10 Informatik IV
Im Stadtwald 15, D-6600 Saarbrücken 11

Electronic mail address: wisber%sbsvax.uucp@germany.csnet

Report No. 10

R. Block: Lexical Functional Grammar and Natural Language Generation, Universität Hamburg, Hamburg, 1986

Abstract: Lexical Functional Grammar (LFG) is one of the more promising candidates for a general linguistic theory to emerge from the theoretical upheavals of the seventies. It displays a high degree of formalization and lends itself particularly well to computer implementation. In addition, LFG seeks to shed light on the mental representation of language in humans.

This paper addresses itself to three points: 1. How plausible are LFG's claims to "psychological reality"? 2. To what extent is LFG superior or inferior to the Standard Theory of Transformational Grammar from which it is ultimately derived? 3. How realistic is LFG as a theory supporting language generation as opposed to language interpretation?

The autor concludes that, despite positive features in LFG, transformational operations on canonical structures is a more efficient and realistic method of generation natural language for man and machine.

Report No. 11

R. Hunze, H.-U. Block: A Two Step Reference Problem Solver, Siemens AG, München, 1987

Abstract: The article is based on the discourse representation theory (DRT) of H. Kamp and Chomsky's binding Theory. A mechanism is presented that builds a discourse representation structure (DRS) and determines all referents that are syntactically admissible and in agreement with the accessibility conditions of DRT.

Imposing semantic restrictions would lead to the minimization of the set of possible referents but this has not been investigated yet. Continuing earlier work, unification grammar is used to develop a formalism that allows for a flexible connection between syntax and semantics. This basic idea is to use the equation mechanism LFG provides to encode semantic information and model DRT explicitly. The system is written in InterLisp-D and runs on a Siemens 5815.

Report No. 12

H.-G. Siedka-Bauer: Repräsentation des Fachwissens einer komplexen Domäne in einer KL-ONE-artigen Repräsentationssprache, SCS, Hamburg, 1987

Abstract: An essential part of the knowledge base incorporated in the natural language investment consultation system WISBER is the terminological knowledge of the domain of investment.

QUIRK - a KL-ONE-like representation language - an outcome of the WISBER-project - was just to do this work.

The specific demands of the domain knowledge with respect to the representation and the possibilities as well as constraints of the representation language are the issues of this report.

EIN SYSTEM ZUR GENERIERUNG VON DEUTSCHEN TEXTEN AUS SEMANTISCHEN REPRÄSENTATIONEN

vorgelegt von: **Dietmar Rösner**

Berichter: **Prof. Dr. R. Gunzenhäuser, Prof. Dr. C. Rohrer**
Universität Stuttgart, September 1986

Zusammenfassung:

Kern des in dieser Arbeit vorgestellten Textgenerators SEMTEX ist ein vom Autor für das SEMSYN-Projekt implementierter Generator. Dieser erhielt als Eingabe semantische Strukturen - in Form einer erweiterten Kasusrahmennotation -, die aus Titeln japanischer wissenschaftlicher Arbeiten durch einen Parser des ATLAS/II-Systems (FUJITSU) gewonnen wurden. Die Kombination von Analyse durch ATLAS/II und Generierung durch SEMSYN stellt das erste maschinelle Übersetzungssystem vom Japanischen ins Deutsche dar.

Bei Design und Implementierung des Generators für SEMSYN wurde versucht, ein leicht erweiterbares und vielseitig einsetzbares System zu entwickeln. Rechnung getragen wurde diesen Vorgaben zum einen durch die bei der Implementierung gewählte objektorientierte Strategie, zum anderen dadurch, daß das Repertoire generierbarer

sprachlicher Formen nicht auf Titelsprache (d.h. Nominalgruppen unterschiedlicher Komplexität) beschränkt blieb, sondern alternativ auch verschiedene Arten von Sätzen produziert werden konnten.

Bei der Erweiterung zu SEMTEX werden nicht mehr lediglich isolierte Strukturen generiert, sondern mehrsätzliche zusammenhängende Texte. Ausgehend von Daten zum Arbeitsmarkt wird zunächst ein "Textplan" in Form einer Liste von semantischen Repräsentationen erstellt. Um solche Inhaltsangaben in akzeptable Zeitungstexte umwandeln zu können, mußte der Generator insbesondere um Mechanismen zur Einbeziehung kontextueller Information in seine Entscheidungen erweitert werden:

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

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

Dieter Seelbach:

TRANSFORMATIONSGRAMMATIKEN UND UNIFIKATIONSGRAMMATIKEN

Eine Einführung in die generative Linguistik und in die Grundlagen der Maschinellen Sprachverarbeitung.

Der Redaktion des LDV-Forum ist (in letzter Minute) die Gliederung einer Arbeit von Prof. Dr. Seelbach, Universität Mainz, zugänglich geworden, die für den Bereich Computerlinguistik von unmittelbarem Interesse ist. Nähere Informationen können beim Autor selbst eingeholt werden (Institut für vergleichende und allgemeine Sprachwissenschaft, Univ. Mainz, Saarstraße 1, D-6500 Mainz).

- 1 Einleitung (2 S.)
- 2 Transformationelle Grammatiken (21 S.)
 - 2.1 Der Harris-Typ der Transformationsgrammatik
 - 2.2 Die generative Transformationsgrammatik Chomskys (Standardtheorie)
 - 2.2 .1 Methode und Material
 - 2.2 .2 Arbeitsweise und Aufbau

- 2.2 .3 Andeutung der Entwicklung der generativ-transformationellen Linguistik im Anschluß an die Standardtheorie
- 2.3 Wichtige Arbeiten am Französischen vor transformationellem Hintergrund
 - 3 Unifikationsgrammatiken (56 S.)
 - 3.1 Die Definite Clause Grammar (DCG)
 - 3.2 Die Lexikalisch-Funktionale Grammatik (LFG)
 - 3.2 .1 Überblick und Grundbegriffe
 - 3.2 .2 Aufbau der F-Struktur
 - 3.2 .3 Deskriptive Kraft
 - 3.3 PROLOG als Parser
 - 3.4 Wichtige Arbeiten im Unifikationsrahmen
 - 4 Ausblick (2 S.)
 - 5 Bibliographie in Auswahl

Veranstaltungen

04.1.88 – 08.1.88, **CANBERRA (Australien)**

FTPLV-Weltkongreß

Information: *FIPLV Head Office, Seestr. 247, CH-8038 Zürich*

25.1.88 – 27.1.88, **STOCKHOLM (S)**

DATABAS 88

Information: *Peder Jonsson, KREAB, Sköldungagatan, S-114 27 Stockholm, Schweden*

3.2.88 – 5.2.88, **LONDON (GB)**

Corporate Information Management Conference (CIM '88), Konferenz über betriebliches Informationsmanagement (CIM '88)

Information: *CIM '88, Effective Techn. Marketing Ltd., 101 Bancroft, Hitching,*

9.2.88 – 12.2.88, **AUSTIN, TX (USA)**

Applied Natural Language Processing, 2nd Conference on ...

Veranstalter: *ACL*

Information: *ACL, D.E. Walker, Bell Communications Research, 445 South Street MRE 2A379, Morristown, NJ 07960, USA*

16.2.88 – 17.2.88, **GÖTTINGEN (D)**

Neue Ansätze in der computerlinguistischen Morphologie und Phonologie

Information: *Dr. U. Klenk, Seminar für Romanische Philologie der Univ. Göttingen, Nikolausberger Weg 23, D-3400 Göttingen*

27.2.88 – 5.3.88, **GÜNNE/MÖHNESEE (D)**

Künstliche Intelligenz, 6. Frühjahrsschule ... (KIFS-88)

Information: *Tagungssekr., Christine Harms, c/o GMD, Postfach 1240, D-5205 St. Augustin 1*

2.3.88 – 4.3.88, **WUPPERTAL**

Verständlichkeit, Jahrestagung 1988 der DGfS

Veranstalter: *DGfS*

Information: *Prof. Gisa Rauh, Bergische Universität, Gesamthochschule Wuppertal, Sprach- und Literaturwissenschaft, Postfach 10 01 27, 5600 Wuppertal 1*

3.3.88 – 4.3.88, **ZÜRICH (CH)**

Data Dictionaries und Entwicklungswerkzeuge für DB-Anwendungen

Information: *Schweizer Informatiker-Gesellschaft, Kurssekr., Inst. f. Informatik, ETH-Zentrum. CH-8092 Zürich, Schweiz*

4.3.88 – 18.3.88, **SAN DIEGO, CA (USA)**

Artificial Intelligence Applications, 4th IEEE Conference on ...

Veranstalter: *IEEE Computer Society*

Information: *Elaine Kant, Schlumberger-Doll Research, Old Quarry Rd., Ridgefield, CT 06877-4108, USA*

6.3.88 – 9.3.88, **MONTEREY, CA (USA)**

Theoretical Aspects of Reasoning about Knowledge, 2nd Conference on ...

Veranstalter: *AAAI; IBM*

Information: *Moshe Y. Vardi, IBM Research, Almaden Research Center K53-802, 650 Harry Rd., San Jose, CA 95120-6099, USA*

9.3.88 – 11.3.88, **SAARBRÜCKEN (D)**

Computerlinguistik und ihre theoretischen Grundlagen

, Jahrestagung 1988 der GLDV in Zusammenarbeit mit der DGfS (Sektion für Computerlinguistik) und der GI (FG Natürlichsprachliche Systeme)

Veranstalter: *GLDV; DGfS; GI;*

Information: *J. Haller, IAI / Eurotra-D, Martin-Luther Straße 14, D-6600 Saarbrücken*

14.3.88 – 18.3.88, **VENEDIG (I)**

International Conference Extending Data Base Technology (EDBT '88)

Information: *Michele Missikoff, IAIS-CNR, Viale Manzoni 30, I-00185 Rom, Italien*

16.3.88 – 23.3.88, **HANNOVER (D)**

Hannover-Messe, Welt-Centrum Büro-Information – Telekommunikation (CeBIT '88)

Information: *Deutsche Messe- und Ausstellungs-AG, Messegelände, D-3000 Hannover 82*

17.3.88 – 19.3.88, **DARMSTADT (D)**

Klassifikation und Ordnung, 12. Jahrestagung der Gesellschaft für Klassifikation

Information: *Prof. R. Wille, TH Darmstadt, FB Mathematik, Schloßgartenstr. 7, D-6100 Darmstadt*

21.3.88 – 24.3.88, **CAMBRIDGE, MA (USA)**

User Oriented Content-Based Text and Image Handling (RIO 88)

Veranstalter: *CNRS, CNET, INRIA, ENSM de Paris, CID*

Information: *RIO '88, CID, 36 rue Ballu, F-75009 Paris*

23.3.88 – 26.3.88, **DUISBURG (D)**

Linguistic Approaches to Artificial Intelligence, International Symposium on ... (LAUD Symp.)

Veranstalter: *Linguistic Agency, Univ. of Duisb., LAUD*

Information: *Prof. R. Dirven, Universität, Gesamthochschule Duisburg, Postfach 101629, D-4100 Duisburg 1*

5.4.88 – 8.4.88, **FLORENZ (I)**

Western European Studies and North American Research Libraries, 2nd Conference on ...
Information: *Assunta Pisani, ACRL-WESS Conf. Feasibility Committee, c/o Harvard College Library, Cambridge, MA 02138, USA*

5.4.88 – 8.4.88, **WIEN (A)**

Artificial Intelligence, In: 9th European Meeting on Cybernetics and Systems Research
Veranstalter: *Austrian Society for Cybernetic Studies*
Information: *Robert Trappl, Dept. of Medical Cybernetics and AI, Universität Wien, Freyung 6/2, A-1010 Wien, Österreich*

11.4.88 – 12.4.88, **FRANKFURT/M**

Software für Datenbankaufbau, Retrieval und Thesaurusmanagement, 6. KTK-Workshop der DGD
Information: *DGD e.V., Westendstr. 19, D-6000 Frankfurt 1*

19.4.88 – 22.4.88, **NOTTINGHAM (GB)**

60th ASLIB Conference
Information: *University of Nottingham, East Midlands Conference Centre*

20.4.88 – 22.4.88, **NIZZA (F)**

Electronic Publishing, Document Manipulation and Typography, International Conference on ... (EP 88)
Information: *INRIA, Service des Relations Exterieures, Domaine de Voluceau, Rocquencourt B.P. 105, F-78153 Le Chesnay Cedex, Frankreich*

25.4.88 – 27.4.88, **TYSONS CORNER, Virg. (USA)**

Expert Database System, 2nd International Conference on ...
Veranstalter: *G. Maon University, AAAI, ACM, IEEE*
Information: *Prof. L. Kerschberg, George Mason University, Dept. of Information Systems Eng., 4400 Univ. Drive, Fairfax, Virginia 22030, USA*

3.5.88 – 5.5.88, **FRANKFURT/M (D)**

10. Frühjahrstagung der Online-Benutzergruppe der DGD (DGD-OLBG)
Information: *Deutsche Gesellschaft für Dokumentation, Westendstraße 19, D-6000 Frankfurt/M 1*

5.5.88 – 8.5.88, **PARIS (F)**

Symposium über Semiotik und Epistemologie
Information: *Prof. Dr. R. Posner, Institut für Linguistik, TU Berlin, Sekr. TEL 6, Ernst Reuther Platz 7, D-1000 Berlin 10*

15.5.88 – 18.5.88, **ANN ARBOR, MICHIGAN, (USA)**

Artificial Intelligence: Expert Systems and other Applications, 17th ASIS Mid-Year Meeting
Information: *L.C. Smith, ASIS 88 Grad. School of*

Lib. & Inf.Sci., Univ. of Illinois, 410 D. Kinley Hall, 1407 W. Gregory, IL 61801, USA

16.5.88 – 20.5.88, **ATHEN (G)**

Information Technology for Organisational Systems, 1st European Conference on ... (EURINFO 88)
Information: *Conf. Office EURINFO '88, c/o GINIS, Vacances Ltd., 23-25 Ermou Str., G-10563 Athen, Griechenland*

23.5.88 – 26.5.88, **ARGONNE, II (USA)**

Automated Deduction, 9th International Conference on ... (CADE-9)
Information: *CADE-9, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA*

24.5.88 – 28.5.88, **BERLIN (West)**

Deutscher Bibliothekskongreß
Information: *Technische Universität Berlin, Straße des 17. Juni, D-1000 Berlin 12*

1.6.88 – 3.6.88, **AVIGNON (F)**

Expert Systems and their Applications, 7th International Workshop on
Veranstalter: *European Coordination Committee for AI*
Information: *Avignon 88, EC2, 269-287, rue de la Garenne, F-9200 Nanterre*

6.6.88 – 10.6.88, **EDMONTON, Alberta (CA)**

Canadian Artificial Intelligence Conference (CSCSI-88)
Veranstalter: *Canadian Soc. f. Comput. Studies of Int.*
Information: *Nick Cercone, School of Computing Science, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada*

7.6.88 – 10.6.88, **BUFFELO, N.Y. (USA)**

26th Annual meeting of the Association for Computational Linguistics (ACL 88)
Information: *J.R. Hobbs, ACL 88 Programm Chair, SRI Int., 333 Ravenswood Avenue, Menlo Parc, CA 94025, USA*

12.6.88 – 15.6.88, **GRENOBLE (F)**

Research and Development in Information Retrieval, 11th International Conference on ... (SIGIR '88)
Veranstalter: *acm*
Information: *Yves Chiamella, Laboratoire Genie Informatique IMAG, BP 68 - 38402 Saint Martin D'Heres Cedex, France*

12.6.88 – 15.6.88, **ANN ARBOR, MI (USA)**

Machine Learning, 5th International Conference on ...
Information: *#03J. E. Laird, University of Michigan, 904 Monroe Street, Ann Arbor, MI 48107-1234, USA*

VERANSTALTUNGEN

23.6.88 – 26.6.88, **AMSTERDAM (NL)**

Artificial Intelligence in Management Theory, Workshop on ...

Information: *Univ. of Amsterdam, University Center, Oost Indisch Huis, OudeHoogstraat 24, NL-1012 CE Amsterdam*

29.6.88 – 30.6.88, **WESTMINSTER (GB)**

International Conference on Information Management (INCONIM)

Information: *Sherry Jespersen, Information House, 26-27, Boswell St., London WC1N 1JZ, Großbritannien*

4.7.88 – 8.7.88, **CANTON (China)**

The Role of Artificial Intelligence in Databases and Inf. Systems

Veranstalter: *IFIP*

Information: *Robert Meersmann, Infolab, K.U. Brabant, Postbus 90153, NL-5000Le Tilburg, Netherlands*

5.7.88 – 8.7.88, **EDINGBURGH, (Schottland)**

Logic in Computer Science, 3rd Annual Symposium on ... (LICS)

Veranstalter: *IEEE Comp. Soc.; Univ. of Edingb.; ...*
Information: *Ashok k. Chandra, IBM T.J. Watson Research Center, P.O.B. 218, Yorktown Heights, N.Y. 10598, USA*

8.7.88 – 22.7.88, **URBINO (I)**

Semiotisches Sommerinstitut

Information: *Christina Catani, Centro Internazionale di Semiotica e de Linguistica, Piazza del Rinalcmento 7, I-61029 Urbino, Italien*

18.7.88 – 22.7.88, **PARIS (F)**

Second Generation Expert System, 12th IMACS World Congress '88 (IMACS '88)

Information: *IMACS '88, Laboratoires de Marcousis, Frankreich*

24.7.88 – 29.7.88, **LAUSANNE (CH)**

Computers in Education, European Conference on ... (ECCE '88)

Information: *Prof. Bernard Levrat, ECCE 1988, Centre Universitaire d'Informatique, 12, rue du Lac, CH-1207 Genf*

24.7.88 – 31.7.88, **ZAGREB (Israel)**

Symposium über Ikonismus der Syntax, in: 12 Internationaler Kongreß für antropolog. u. ethnolog. Wiss.

Information: *Prof. Marze E. Lansberg, Shik, ona-Straße 1, Bat-Galim, Haifa 35014, Israel*

1.8.88 – 5.8.88, **MÜNCHEN (D)**

European Conference on Artificial Intelligence (ECAI-88)

Veranstalter: *ECCAI; GI; TU München*

Information: *Prof. B. Radig, TU München, Institut*

für Informatik, Postfach 202420, 8000 München 2

1.8.88 – 2.8.88, **MÜNCHEN (D)**

Natural Language Understanding

Veranstalter: *Deutsche Informatik Akademie*

Information: *GI Deutsche Informatik Akademie, Geschäftsstelle, Godesberger Allee 99, D-5300 Bonn 2*

15.8.88 – 17.8.88, **LINZ (AU)**

Office Information Systems: "The Design Process", Working Conference on

Veranstalter: *IFIP-WG 8.4*

Information: *Prof. Dr. W. Wagner, Universität Linz, Institut für Informatik, A-404 Linz*

17.8.88 – 21.8.88, **BRIGHTON (GB)**

18th World Congress of Philosophy

Information: *Conference Service, World Congress of Philosophy, 3-5 Bute Street, GB London SW7 3EY*

18.8.88 – 19.8.88, **BUDAPEST (U)**

New Directions in Machine Translation, International Conference

Veranstalter: *BSO, NJSZT*

Information: *John v. Neumann Soc. for Computing Science, Conference Secr., Budapest, Báthori u. 16, 1054*

22.8.88 – 27.8.88, **BUDAPEST (H)**

Computational Linguistics, 12th International Conference on ... (COLING '88)

Information: *Coling 88 Secr., c/o MTE SZ Congress Bureau, Kossuth ter 6-8, H-1055 Budapest, Ungarn*

22.8.88 – 27.8.88, **MÜNCHEN (D)**

Raum und Grenzen, 12. Kongreß der Int. Gesellschaft für Vergleichende Lit.-Wiss.

Information: *Prof. Roger Bauer, Abteilung für Vergleichende Literaturwissenschaft, Universität München, Schellingstr. 3, D-8000 München 40*

28.8.88 – 1.9.88, **HELSINKI (SF)**

Informarion-Knowledge Evolution, 1988 FID Congress

Information: *Finnish Soc. of Inf. Services, Ms. R. Launo, P.O.B. 1025, SF-00101 Helsinki, Finland*

29.8.88 – 31.8.88, **WIEN (A)**

Wiener Workshop Wissensbasierte Sprachverarbeitung (WWWS 1988)

Veranstalter: *ÖGAI; GI / FG Natürlichsprachliche Systeme*

Information: *Dr. H. Trost, Inst. für Medizinische Kybernetik und Artificial Intelligence, Universität Wien, Freyung 6 / 2, A-1010 Wien, Österreich*

31.8.88 – 2.9.88, **BRÜGGE (B)**

Database Theory, International Conference on ... (ICDT 1988)

Information: *Jan Paredaens, Computer Science Dept., Universitairs Instelling Antwerpen, Universiteitsplein 1, B-2610 Wilrijk*

6.9.88 – 10.9.88, **BOSTON, MA (USA)**

International neural network Society 1988 Annual Meeting

Information: *UNIGLOBE, Neural Networks 1988, 40 Washington Street, Wellesley Hills, MA 02181, USA*

19.9.88 – 23.9.88, **ERINGERFELD (D)**

German Workshop on Artificial Intelligence (GWAI '88)

Information: *Prof. Wolfgang Hoepfner, EWH Koblenz, Rheinau 2-3, D-5400 Koblenz*

20.9.88 – 23.9.88, **GÜNZBURG (D)**

Interdisziplinäres Symposium über Text und Bild, Bild und Text

Information: *Prof. Georg Jäger, Institut für Deutsche Philologie, Universität München, Schellingstr. 3, D-8000 München 40*

25.9.88 – 28.9.88, **HANNOVER (D)**

Expert Systems and Decision Support in Medicine

Veranstalter: *EFMI; GMDS*

Information: *Conf. Sekr., Medical School Hannover, Institut für Medizinische Informatik, Postfach 610180, 3000 Hannover 61*

27.9.88 – 30.9.88, **EBERNBURG (D)**

Object-Oriented database Systems, 2nd International Workshop on

Veranstalter: *GI-FG 2.5.1, FZI Karlsruhe, ACM SIGMOD*

Information: *Dr. K. Dittrich, Forschungszentrum Informatik an der Univ. Karlsruhe, (FZI, haid-und-Neu-Straße 10-14, D-7500 Karlsruhe 1*

Information: *Dr. D. Hohgreffe, Universität Hamburg, FB Informatik, Rothembachchaussee 67-69, D-2000 Hamburg 13*

28.11.88 – 2.10.88, **TOKIO (J)**

Fifth Generation Computer Systems 1988, International Conference on (FGCS '88)

Veranstalter: *ICOT*

Information: *Prof. Dr. H. Schweppe, FU Berlin, Institut für Informatik, Netorstraße 8-9, D-1000 Berlin 31*

28.9.88 – 30.9.88, **AACHEN (D)**

Deutscher Dokumentartag 1988

Veranstalter: *Deutsche Gesellschaft für Dokumentation*

Information: *DGD e.V., Westendstr. 19, D-6000 Frankfurt 1*

5.10.88 – 7.10.88, **CHARLES, ILL (USA)**

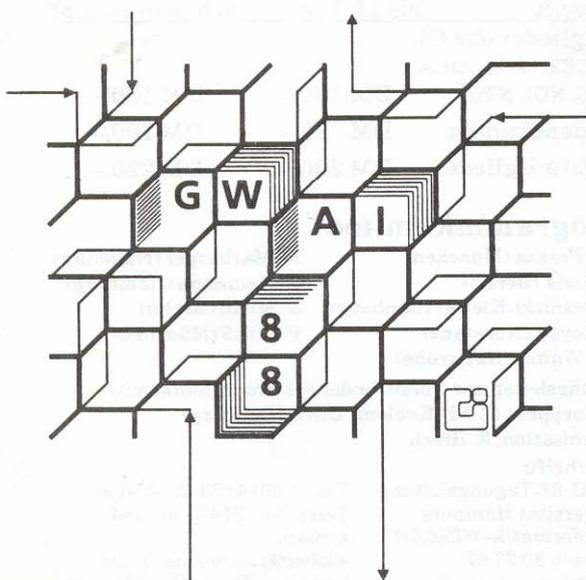
Spatial Reasoning and Multi-Sensor Fusion, Workshop on ...

Information: *Si-Chin Chen, Dep. of Computer Science, University of North Carolina, Charlotte, NC 28223*

17.10.88 – 19.10.88, **BERLIN (West)**

Vernetzte und komplexe Informatik-Systeme, 18. GI-Jahrestagung

Veranstalter: *GI, Univ. Hamburg*



GWAI 88

**12. Fachtagung
Künstliche Intelligenz**

**19.-23. September 1988
Schloß Eringerfeld**

Die Tagung wird vom Fachausschuß 1.2
"Künstliche Intelligenz
und Mustererkennung"
im Fachbereich 1
der Gesellschaft für Informatik e.V.
veranstaltet

Call for Papers

Beiträge sind aus allen Bereichen der Künstlichen Intelligenz erwünscht, z.B.:

- Wissensrepräsentation
- Inferenzen
- Deduktion
- Natürlichsprachliche Systeme
- Bildverstehen
- Robotik
- Maschinelles Lernen
- Expertensysteme
- Kognition
- KI-Programmierung

Der Tradition der letzten Fachtagungen folgend ist während der GWAI-88 ein Tag für spezielle Sektionen vorgesehen:

- Bildverstehen: Niedere Bilddeutung in der KI (S. Stiehl, TU Berlin)
- Integration von Analyse und Synthese in KI-Systeme (K. Haenelt, R. Th. King, GMD Darmstadt)
- Grenzen der Künstlichen Intelligenz (E. Lehmann, Univ. Stuttgart)

Auch zu diesen Sektionen können Beiträge eingesandt werden.

Die Veröffentlichung aller Vorträge in den Informatik-Fachberichten des Springer-Verlags ist vorgesehen.

Nicht pünktlich eingehende Endfassungen akzeptierter Beiträge können nicht in den Tagungsband aufgenommen werden.

Tagungssprachen sind Deutsch und Englisch. Systemvorführungen und Projektvorstellungen werden begrüßt und sollten möglichst frühzeitig der Tagungsleitung angekündigt werden.

Termine

Einreichen des vollständigen Beitrags in vier Exemplaren beim Tagungsleiter	bis 1.4.1988
Benachrichtigung über Annahme oder Ablehnung des Beitrags	ab 1.6.1988
Anmeldung von Systemvorführungen und Projektvorstellungen bei der Tagungsleitung	bis 15.7.1988
Abgabe der endgültigen Fassung akzeptierter Papiere	bis 1.7.1988

Form der Beiträge

Auf der ersten Seite eines eingereichten Beitrags ist anzugeben: der Titel des Papiers, sein(e) Autor((in (n)) (en)), ob es ein Lang- oder Kurzpapier sein soll und zu welchem Themenbereich es eingereicht wird. Beiträge für spezielle Sektionen sollten als solche kenntlich gemacht sein.

Das Manuskript muß auf DIN A 4 Papier mit jeweils 2 cm Rand oben und unten sowie je 1,5 cm Rand an den Seiten, mit zweizeiligem Zeilenabstand und 12 cpi Schriftgröße (Elite) geschrieben werden (Springer Format).

Langpapiere sind höchstens 10 Seiten, Kurzpapiere höchstens 5 Seiten lang. Einem akzeptierten Langpapier entspricht ein Fachvortrag von 45 Minuten Dauer incl. Diskussion. Einem akzeptierten Kurzpapier entspricht ein Vortrag von 30 Minuten Dauer incl. Diskussion.

Begutachtung

Pünktlich bis zum 1. April 1988 bei der Tagungsleitung eingereichte Beiträge werden an die für den betreffenden Themenbereich zuständigen Mitglieder des Programmkomitees weitergeleitet, die sie begutachten und zusätzlich zwei weitere Gutachten einholen. Papiere, die zu einer speziellen Sektion eingereicht werden, begutachtet der Sektionsleiter, ein von ihm gewählter zusätzlicher Gutachter sowie ein Mitglied des Programmkomitees.

Die Endverantwortung trägt der Vorsitzende des Programmkomitees. Die Einarbeitung von Anmerkungen der Gutachter(innen) in die Endfassung eines Papiers ist Voraussetzung dafür, daß es in den Tagungsband aufgenommen wird.

Anmeldung

Die Anmeldung zur Tagung erfolgt durch Einsenden des Anmeldeformulars und durch Bezahlung des Tagungsbeitrages auf das Konto

Sonderkonto G (Prof. W. Hoepfner)

Konto-Nr. 469 77-209,

Postgiroamt Hamburg, BLZ 200 100 20

Tagungsbeitrag

Bezahlt bis 15.7.88 nach dem 15.7.88

Mitglieder der GI (AFCET, AISB, AICA, BCS, NGI, NTG)	DM 160,-	DM 180,-
Student/inn/en	DM 80,-	DM 100,-
Nichtmitglieder	DM 200,-	DM 220,-

Programmkomitee

Chr. Freksa (München)	H. Marburger (Hamburg)
K. Morik (Berlin)	B. Neumann (Hamburg)
B. Owsnicki-Klewe (Hamburg)	S. Stiehl (Berlin)
H. Stoyan (Konstanz)	P. Struß (München)
Chr. Walter (Karlsruhe)	

Tagungsleiter und Vorsitzender des Programmkomitees:

W. Hoepfner (EWH Koblenz, Univ. Hamburg)

Organisation: R. Brech

Anschrift:

GWAI-88-Tagungsleitung

Universität Hamburg

FB Informatik-WISBER

Postfach 30 27 62

D-2000 Hamburg 36

Tel.: (040) 4123-2595/3699

Telex-Nr.: 21473 unihhd

E-mail:

wisber%rz.informatik.uni-

hamburg.dbp.de@unido.uucp

Arbeitskreise

ZUR ERINNERUNG ...

Bitte stellen Sie Zurückhaltung, Säumigkeit oder Scham beiseite ...

... und senden Sie Ihren Fragebogen zur Berufsfelderkundung *Computerlinguistik* im vorbereiteten Rückantwort-Umschlag ohne Absender nach Koblenz zurück! Ausgefüllt hätten wir ihn am liebsten.

Ihre *BCL-Projektgruppe*

(*"Erkundung von Berufs- und Tätigkeitsfeldern von Computerlinguisten/innen"*)

FRAGEBOGEN TEXTANALYSE – EINIGE ERGEBNISSE –

Der Fragebogen Textanalyse wurde im Herbst 1986 an alle Mitglieder des Arbeitskreises Textanalyse in der GLDV verteilt. Das Ziel der Umfrage war es (i) eine Diskussionsgrundlage für den AK bereitzustellen und (ii) Hinweise zur praktischen Arbeit des AK zu erhalten. Da insbesondere die Diskussionsgrundlage allgemein interessieren könnte, soll sie hier kurz referiert werden.

Das wichtigste (vielleicht etwas überraschende) Ergebnis war, daß Konzepten der KI eine erhebliche Wichtigkeit insbesondere hinsichtlich der Theorie der Textanalyse beigemessen wurde. Dies gilt für die gegebenen Definitionen der TA, die zusammengefaßt werden können als "Abbildung von Text(mengen) auf eine vordefinierte Repräsentation(ssprache)" und für den prozentualen Anteil anderer Disziplinen an der Theorie der Textanalyse (KI: 21 %, Linguistik ohne Quantitative und Textlinguistik 20 %). In diesem Zusammenhang ist es auch bemerkenswert, daß "Textverstehen" als das mit der TA ähnlichste Gebiet angesehen wird. Denn bislang scheint dieser Begriff doch auf die Arbeiten der Yale-Gruppe (einschließlich Vorläufer und Nachfolger) beschränkt gewesen zu sein. Insgesamt kann man konstatieren, daß die TA in enger Verbindung zu Ansätzen des Wissenserwerbs (nach denen leider nicht gefragt worden war) gesehen wird.

Möglicherweise ist jedoch auch das nur eine Facette der Forschungslandschaft. Diese nämlich wird von den Beantworter(inne)n sehr unterschiedlich gesehen. Satzübergreifende, gegebenenfalls textlinguistisch orientierte Verfahren, die Interaktion mit fachlichen Wissensquellen, formale Grammatiken, Parsingverfahren, Systemtests – alles das ist wichtig in der Textanalyse. Allgemein wird – übereinstimmend mit allgemeinen LDV-Tendenzen, so scheint mir – die Grundlagenforschung für eher vorwiegend gehalten.

Insgesamt, das jedenfalls hat der Fragebogen gezeigt, gibt es für den AK Textanalyse genug an be-

grifflicher Klärungsarbeit zu tun. Wer an diesem AK teilnehmen möchte, wende sich bitte an den Verfasser. Die nächste Arbeitskreissitzung wird, so die derzeitige Planung, am Freitag (nach-)mittag im Rahmen der GLDV-Tagung Saarbrücken 1988 stattfinden.

Sebastian Goeser, SFB 129 TP B2, Universität Ulm, Am Hochsträß 8, D-7900 Ulm, AK Textanalyse

AKADEMISCHE WEITERBILDUNG AM COMPUTER FÜR GEISTESWISSENSCHAFTLER ODER PHILOLOGEN

An den Universitäten *Bochum, Göttingen* und *Würzburg* (s. LDV-Forum 5(1987)1: S. 51) gibt es seit kurzem Weiterbildungs-Studiengänge, die Absolventen der Geisteswissenschaften befähigen sollen, den Computer zur philologischen Textverarbeitung bzw. zur Anwendung auf linguistische Fragestellungen einzusetzen. Auf diese Weise sollen den Geisteswissenschaftlern durch eine zusätzliche akademische Ausbildung neue Berufsfelder erschlossen werden.

Da man dabei in Bochum, Göttingen und Würzburg von unterschiedlichen Konzepten ausgeht, werden alle drei Studiengänge einzeln systematisch vorgestellt [1].

Die Darstellung enthält jeweils

- die Universitätsadresse, bei der man genauere Informationen erfragen kann,
- Informationen über den Studiengang, geordnet nach
 - (1) Bezeichnung des Studiengangs
 - (2) am Studium beteiligte Fächer
 - (3) vorgesehene Studienzeit
 - (4) regelmäßiger Studienbeginn
 - (5) Studienvoraussetzungen
 - (6) Zeitpunkt der Einrichtung des Studiengangs
 - (7) Umfang und Inhalt des Studium
 - (8) Abschlußprüfung und Prüfungsleistungen
 - (9) Studienziele

Universität BOCHUM

Fakultät für Philologie, Dekanat, Universitätsstraße 150, Gebäude GB 5. Stock, D-4630 Bochum 1

- (1) Zusatzstudiengang Informatik für die Geisteswissenschaften
- (2) Beteiligte Fächer: Philologen (einschl. Allgemeine Sprachwissenschaft), Sozialwissenschaft, Geographie, Evangelische und Katholische Theologie, Philosophie, Pädagogik, Geschichtswissen-

ARBEITSKREISE

schaft, Mathematik, Elektrotechnik, Wirtschaftswissenschaft, Jura.

- (3) Regelstudienzeit 3 Semester
- (4) Studienbeginn: Wintersemester
- (5) Studienvoraussetzung: erfolgreicher Abschluß eines Studiums in einem geisteswissenschaftlichen Fach, einschl. Geographie
- (6) Die Prüfungsordnung tritt mit dem Wintersemester 1987/1988 in Kraft.

(7) Der Gesamtumfang des Studiums im Pflicht-, Wahlpflicht- und Wahlbereich umfaßt 60 SWS.

Studieninhalte sind:

- mathematische Grundlagen
- Programmierung
- Grundlagen der Informatik (Hardware, Software, Datenkunde, Problemanalyse, Systemanalyse, Anwendersoftware)
- fachbezogene Problemstellungen und Anwendungen

(8) Die zweiteilige Abschlußprüfung besteht aus einer Hausarbeit (Frist: 4 Wochen) und einem 45-minütigen Kolloquium, für das der Kandidat 3 Themenkreise vorschlagen kann.

Der erfolgreiche Abschluß der Prüfung wird durch ein Zeugnis "Zusatzstudium Informatik für die Geisteswissenschaften" bestätigt

(9) Studienziele: Die Absolventen des Zusatzstudienganges sollen sich auf dem Gebiet der Informatik Fachkenntnisse und Fähigkeiten aneignen, die sie befähigen, wissenschaftliche Methoden und Erkenntnisse der Informatik mit der vorausgegangenen fachlichen Ausbildung zu verbinden und sie in diesem Bereich anzuwenden.

Universität GÖTTINGEN

Fachbereich Historisch-Philologische Wissenschaften, Wilhelmshaus 1, D-3400 Göttingen

(1) Ergänzungsstudiengang Linguistische Datenverarbeitung

(2) Beteiligte Fächer: Neuere Englische Sprache, Allgemeine und Romanische Sprachwissenschaft, Systemorientierte Informatik im Fachbereich Mathematik

(3) Regelstudienzeit 4 Semester

(4) Studienbeginn: Wintersemester

(5) Studienvoraussetzung: Zulassungsvoraussetzung für das Studium ist der erfolgreiche Abschluß der Magisterprüfung in einem philologischen Fach an einer wissenschaftlichen Hochschule oder in einem entsprechenden wissenschaftlichen Fach an einer Gesamthochschule.

Jährlich werden 20 Bewerber/innen zugelassen, die ihren Zulassungsantrag für das Wintersemester jeweils bis zum 30. Juli eines Jahres bei der Uni-

versität eingereicht haben müssen. (Da die Universität über die Form des Antrags und beizufügende Anlagen bestimmt, empfiehlt sich eine rechtzeitige Bewerbung, etwa im ersten Quartal eines Jahres). Ein etwaiger Zulassungsbescheid muß bis zu einem genannten Termin vom Bewerber mit einer Annahmeerklärung beantwortet werden ansonsten wird die Zulassung unwirksam.

(6) Die Studien- und Prüfungsordnung ist mit Wirkung zum Wintersemester 1986/1987 in Kraft getreten.

(7) Aufbau des Studiums: Der erste Studienabschnitt (mindestens 36 SWS) umfaßt 2 Proseminare mit Übungen zur Einführung in die LDV, 2 vierstündige Übungen zur Nichtnumerischen Programmierung, eine zweistündige Einführung in die Logik und 12 SWS Einführung in die Informatik und 6 SWS Kleinrechnersysteme.

Im zweiten Studienabschnitt (mindestens 20 SWS) folgen Wahlpflichtveranstaltungen zur systemorientierten Informatik und zu Datenbanksystemen, darüber hinaus Seminare zur Sprachstatistik und Seminare, die für die Studiengänge *Neuere Englische Sprache* und *Allgemeine und Romanische Sprachwissenschaft* angeboten werden. Letztere sind Hauptseminare zu Themen wie

- Analyse- und Syntheseverfahren im Bereich natürlicher Sprachen
- Automatische Übersetzung
- Automatische Lemmatisierung
- Automatische Textverarbeitung

Diese Seminare werden jeweils als zum Ergänzungsstudiengang zugehörig gekennzeichnet.

(8) Die Abschlußprüfung besteht aus einer Hausarbeit, die spätestens 4 Monate nach der Zulassung zur Prüfung abzugeben ist und aus einer einstündigen mündlichen Prüfung besteht. Über die erfolgreiche Prüfung wird ein Zertifikat ausgestellt.

(9) Studienziele: Während im ersten Studienabschnitt Grundkenntnisse in LDV, theoretischer Linguistik, formaler Logik und Informatik erworben werden, befaßt sich der zweite Studienabschnitt mit der Umsetzung dieser Kenntnisse in Anwendungsbereiche. dabei kann die/der Studierende einen Schwerpunkt auswählen.

Das Ergänzungsstudium führt Philologen in die Methoden der elektronischen Datenverarbeitung ein und vermittelt ihnen Erfahrungen durch die Bearbeitung von Themen aus der LDV.

Universität WÜRZBURG

Institut für Deutsche Philologie, Linguistische Informations- und Textverarbeitung, Am Hubland, D-8700 Würzburg.

(1) Aufbaustudiengang Linguistische Informations- und Textverarbeitung

(2) Beteiligte Fächer: Deutsche Philologie, Englische Philologie, Angewandte Mathematik (mit Informatik)

(3) Regelstudienzeit 2 Semester

(4) Studienbeginn: Wintersemester (während der zweijährigen Erprobungsphase; später in jedem Semester)

(5) Studienvoraussetzung: Zulassungsvoraussetzung für das Studium ist der erfolgreiche Abschluß in einem philologischen Fach mit sprachwissenschaftlichem Schwerpunkt (an einer Hochschule oder mit staatlicher Abschlußprüfung). Absolventen anderer Fächer können in begründeten Ausnahmefällen zugelassen werden.

(6) Der Aufbaustudiengang wird mit dem Wintersemester 1985/1986 angeboten.

(7) Aufbau des Studiums: Es umfaßt mindestens 48 SWS. Studieninhalte sind:

- EDV-Englisch
- betriebswirtschaftliche Grundkenntnisse
- Großrechner I und II (mit Praktika)
- Mikrorechner I und II (mit Praktika)
- Dokumentation
- Textverarbeitungsprogrammierung (mit Übungen)
- Programmiertechnik (zwei Seminare)

Die Seminare zur Programmiertechnik stammen aus dem Lehrangebot des Instituts für Angewandte Mathematik und Statistik. Ferner gehören auch Informatik-Lehrveranstaltungen zum Lehrangebot.

(8) Die Abschlußprüfung besteht aus einer 40-minütigen mündlichen Prüfung auf den Gebieten *Linguistische Textverarbeitung* und *Dokumentation*, und zwar zu etwa gleichen Teilen mit Großrechneranwendung und Kleinrechneranwendung.

Der erfolgreiche Abschluß wird durch ein Zertifikat bestätigt.

(9) Studienziele: Das Aufbaustudium soll Philologen befähigen, im Beruf solche Tätigkeiten auszuüben, für die sie zwar gemäß ihrer philologischen Ausbildung geeignet sind, die ihnen aber ohne Erfahrungen mit der elektronischen Datenverarbeitung verschlossen bleiben.

Das Studium soll deshalb die Studierenden mit Grundbegriffen und Verfahren der EDV im Bereich der Linguistischen Informations- und Textverarbeitung vertraut machen und sie dazu anleiten, eigenständige Arbeiten an Groß- und Kleinrechnersystemen durchzuführen.

Anmerkung

- [1] Die Darstellungen der Studiengänge wurden mit jeweils Zuständigen an den Universitäten abgesprochen und waren ursprünglich zur Veröffentlichung in den "Blättern zur

Berufskunde der Bundesanstalt für Arbeit: "Sprachwissenschaftler/in, Computerlinguistin, Phonetiker/in" vorgesehen. da dort aber nur ein kurzer Hinweis zur Weiterbildung erscheinen wird, geben wir unsere Informationen hier ungekürzt wieder.

AK "Ausbildung und Berufsperspektiven"

KONTAKTADRESSEN DER GLDV-ARBEITSKREISE

AK "Textanalyse"

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

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*

AK "Maschinelle Übersetzung"

Dr. D. Rösner, *GMD F4, Projekt GENESIS, Dolivostraße 15, D-6100 Darmstadt*

AK "Ausbildung und Berufsperspektiven"

M. Lutz-Hensel, *Institut für Kommunikationsforschung und Phonetik, Poppelsdorfer Allee 47, D-5300 Bonn 1*

AK "Lexikographie"

J. Brustkern, *Institut für Angewandte Kommunikationsforschung und Phonetik, Poppelsdorfer Allee 47, D-5300 Bonn 1*

Mitteilungen aus der GLDV

SYMPOSIUM: COMPUTERLINGUISTIK – UND IHRE THEORETISCHEN GRUNDLAGEN

Jahrestagung der GLDV, veranstaltet zusammen mit der DGfS (Sektion Computerlinguistik) und der GI (Fachgruppe Natürlichsprachliche Systeme)

Vom Mittwoch, dem 9. bis zum Freitag, dem 11. März 1988 wird in Saarbrücken ein Symposium mit dem Thema *Computerlinguistik und ihre theoretischen Grundlagen* stattfinden. Veranstalter sind die *Gesellschaft für Linguistische Datenverarbeitung (GLDV)*, die *Sektion für Computerlinguistik* in der *Deutschen Gesellschaft für Sprachwissenschaft (DGfS)* und die *Fachgruppe Natürlichsprachliche Systeme* in der *Gesellschaft für Informatik (GI)*.

Die Veranstaltung wird organisiert vom *Institut für Angewandte Informationswissenschaft (IAI)/EUROTRA-D* in Saarbrücken. Das wissenschaftliche Programm wird sich auf den Mittwoch sowie die Vormittage des Donnerstag und Freitag verteilen. Der späte Nachmittag ist am Donnerstag für die Jahresversammlung der GLDV-Mitglieder reserviert, am Freitagnachmittag werden die Sitzungen der Arbeitsgruppen der GLDV stattfinden. Bisher haben die Arbeitskreise *Lexikographie* und *Textanalyse* Aktivitäten angekündigt.

Auf der Tagung sollen insbesondere Probleme behandelt werden, die die Grundlagen der Computersimulation von sprachlichen Prozessen und der Mensch-Maschine-Interaktion betreffen. Im *Call for*

Papers wurden Beiträge angefordert, die die Anwendungsrelevanz theoretischer Modelle oder den Theoriebedarf anhand anwendungsbezogener Vorhaben beleuchten. Als Beispiele waren genannt:

- Grammatiktheoretische Grundlagen
- Modellierung des sprachlichen Wissens auch unter Heranziehung nicht-spezifisch linguistischer Ansätze
- Theorien und Modelle natürlichsprachlicher Mensch-Computer Interaktion und ihre Bewertung
- Einfluß der computerlinguistischen Konzepte auf die allgemein linguistische Theoriebildung
- Probleme des Theorietransfers zwischen der Computerlinguistik und anderen Wissenschaftsgebieten
- Komplexitätsaspekte der Modellierung
- Theoretische Grundlagen für die Erstellung linguistischer Werkzeuge

Eingeladene Vorträge werden von *Robert C. Moore* (SRI, Menglo Parc) und von *Hans Uszkoreit* (IBM, Stuttgart) gehalten.

Die eingereichten Vorträge werden von einem Programmkomitee, bestehend aus I. Batori (Koblenz), F. Günthner (Tübingen), Ch. Habel (Hamburg), U. Hahn (Passau), S. Kanngießer (Osnabrück), J. Kilbury (Trier), M. Pinkal (Hamburg) und W. Wahlster (Saarbrücken) begutachtet. Informationen können von Prof. Dr. I. S. Batori, EWH Koblenz, Schwerpunkt Linguistik/Studiengang Informatik, Rheinau 3 – 4, D-5400 Koblenz angefordert werden.

d. R.

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