

Band 4 · Nummer 1 · Juni 1986

ISSN 0172 - 9926

LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV

Sprachorientierte
KI - Forschung

Universitäts-
bibliothek

1-3. 07. 86



LDV-FORUM 4(1986)1

Forum der Gesellschaft
für Linguistische
Datenverarbeitung (GLDV)

Die mit Namen gekennzeichneten Beiträge
vertreten ausschließlich die Meinung
der Autoren

Redaktion:

Dr. Gerhard Knorz
TH Darmstadt
Fachbereich Informatik
Karolinenplatz 5
(Tel.: 06151/16-3952, -2859)

Titelgestaltung: A. Trottler

Erscheinungsweise

halbjährlich: 30. Juni und
30. Dezember

Bezugsbedingungen

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

Anzeigen

Media-Information kann bei der
Redaktion angefordert werden.

Beiträge

Autorenrichtlinien für die
Abfassung und die Übermittlung
von Manuskripten sind bei der
Redaktion anzufordern. Die Länge
von Fachbeiträgen sollte 8 bis
12 Seiten nicht überschreiten.

EDITORIAL

Roter Plüsch, Messing- und Kristall-Lüster, Apfelstrudel, Melange und ein Stoß Papier auf einem kleinen Marmortisch bei gedämpfter Kaffeehaus-Atmosphäre im Herzen Wiens. Es setzt sich der Gedanke in mir fest, daß es offensichtlich zum Autorenarbeitsplatz auch heute noch ernstzunehmende und liebenswerte Alternativen gibt. Der oberste Papierbogen ist beschrieben und beginnt mit ...

... "Sprachorientierte KI-Forschung" - Ein Schwerpunktthema, das sich, wegen der Vielzahl der Beiträge, zu einer Art Triptichon entwickelt hat. Nach den "Thesen" im letzten Heft 2/85 sind in der vorliegenden Ausgabe, neben Beiträgen zur (akustischen) Spracherkennung und zur Sprachgenerierung, einschlägige Projekte in Hamburg, Konstanz und Saarbrücken vorgestellt. In der nächsten Ausgabe soll das Schwerpunktthema abgeschlossen werden. Dazu sollten sich die Leser des LDV-FORUM aufgefordert fühlen, die offensichtlich lebhaftere Diskussion der veröffentlichten Stellungnahmen in kurze und prägnante schriftliche Kommentare umzusetzen. Ansonsten steht der Vorschlag zur Diskussion, eine (kleine) Datenbank mit Produkten der sprachorientierten KI aufzubauen. Meinungen, Anregungen und Beiträge sind dazu sehr erwünscht. Als nächstes Schwerpunktthema ist für die erste Ausgabe 1987 "Maschinelle "Übersetzung" vorgesehen.

Ich halte es im Hinblick auf die Offenheit des LDV-FORUM für sehr wichtig, daß neben dem Schwerpunkt genügend Raum für weitere Fachbeiträge bleibt. Der Redaktionsschluß-Termin bezieht sich allerdings insofern nicht auf Fachbeiträge, als in diesem Bereich die reaktionelle Planung langfristiger erfolgen muß. Manuskripte sollten also so frühzeitig wie möglich eingereicht werden oder in besonderen Fällen angekündigt werden. Es ist zu erwarten, daß Vorstand und Beirat der GLDV in Kürze die Frage der Begutachtung fachlicher Beiträge, auch in organisatorischer Hinsicht, abschließend behandeln.

Beiträge für die Rubriken "Mitteilungen", "Tagungsberichte" u.s.w. werden selbstverständlich im Regelfall ohne "redaktionelles Filter" veröffentlicht. Die Redaktion muß sich jedoch das Recht vorbehalten, im Einzelfall und nach Rücksprache mit dem Autor die Vorlage zu kürzen oder die Veröffentlichung zu verschieben. Je umfangreicher das Manuskript ist, o zeitiger sollte es eingereicht oder zumindest der Redaktion angekündigt sein.

Im nächsten Heft wird eine neue Rubrik "Rezensionen" eingerichtet werden. Insbesondere liegen der Redaktion Rezensionsexemplare für die folgenden Publikationen vor:

- Batori, I.; Weber, H.J. (Hrsg.): Neue Ansätze in Maschinellem Sprachübersetzung: Wissensrepräsentation und Textbezug. Niemeyer 1986.

- Panyr, J.: Automatische Klassifikation und Information Retrieval. Niemeyer 1986.



Editorial

Bankverbindung

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

Herstellung

Verlagsdruckerei Hoppenstedt
Auage: 300 Exemplare

Redaktionsschluß für Heft 2/86

15. Oktober 1986

Anschrift der GLDV

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

Mitgliederbeiträge

für Studierende	DM 10,-
für natürliche Personen	DM 50,-
für wissenschaftliche Institute	DM 100,-
für gewerbliche Unternehmen, Behörden und andere juristische Personen	DM 250,-

Zur Zeit bestehende Arbeitskreise

AK Ausbildung und Berufsperspektiven
AK LDV und Nachbarn
AK Lexikographie
AK Maschinelle Übersetzung
AK Spracherkennung und -produktion
AK Textanalyse

Studienführer LDV

Lutz Henkel, M. Studienführer: Linguistische Datenverarbeitung (LDV) für die wissenschaftlichen Hochschulen der Bundesrepublik Deutschland 1985.
Zu beziehen über:
Prof. Dr. Jürgen Krause
Universität Regensburg
Linguistische Informationswissenschaft
Postfach 397
8400 Regensburg

Neues Design-Konzept für das LDV-Forum

Als wir (Ilja Paulini Toth, Andrea Trottler und Fred Zimmermann) uns im 6. Sem Grafik-Design entschlossen (auf Anfrage von Herrn Knorz), ein neues Design-Konzept für das LDV-Forum zu erarbeiten, so war dies unser erster Kontakt zur Zeitschriftengestaltung. Was aber jetzt noch hinzukam, war die Komponente der Elektronischen Datenverarbeitung. Da aber als Ausgabegeräte später Fotosatzmaschinen benutzt werden sollen, erarbeiteten wir ein Konzept, das sich im wesentlichen kaum von einer herkömmlichen Verfahrensweise unterscheidet. Das neue LDV-Forum sollte folgende Voraussetzungen erfüllen:

Nicht größer als A4, schwarz/weiß, modernes Erscheinungsbild, mehr Bildinformation, klare Gliederung, Seriencharakter.

Zunächst konstruierten wir ein Raster, auf dem in Zukunft die ganze LDV-Forum-Gestaltung aufbaut. Es besteht aus dem Format, dem Satzspiegel, 2-3 spaltige Einteilung (in mm) und einer Höheneinteilung von jeweils 6 Zeilen 10p Schrift plus 2p Zeilenabstand. (p=Punkt, nach Typografischem Maßsystem).

Um die Übertragung auf ein Textsystem zu vereinfachen, haben wir uns bei der Rubrikenkopf- und Seitengestaltung auf zwei verschiedene Rasterbalken beschränkt.

Als Schriftart verwenden wir eine 10p serifenlose Schrift um einen etwas moderneren Eindruck zu vermitteln. Um den einzelnen Heften in Zukunft ein Zusammenhalt zu geben, erscheint nun

- Eisenberg, P. (Hrsg.): Semantik und künstliche Intelligenz. De Gruyter, 1977

- Capurro, R.: Hermeneutik der Fachinformation. Verlag Karl Alber, 1986

Interessenten melden sich bitte bei der Redaktion. In diesem Zusammenhang sei auch auf die neue Olms-Reihe "Linguistische Datenverarbeitung" hingewiesen, die in Zusammenarbeit mit der GLDV eingerichtet worden ist (näheres unter der Rubrik "Publikationen").

Es ist davon auszugehen, daß bereits das äußere Format dem Leser signalisiert hat, daß die Umstellung auf das angekündigte neue Layout inzwischen in wesentlichen Teilen vollzogen ist. Die an dem Entwurf maßgeblich beteiligten Studenten der Fachhochschule Darmstadt (Fachbereich Gestaltung) kommen selbst zu Wort, um die Konzeption zu erläutern und der Realisierungsgrad zu kommentieren. Die Schwachstelle bei den Herstellungsarbeiten ist gegenwärtig noch der Datenaustausch zwischen Autor und Redaktion. Ich gehe davon aus, daß in den nächsten Monaten eine zufriedenstellende Situation geschaffen werden kann.

Für die Zukunft wünsche ich mir eine kontinuierlich weiter steigende Attraktivität des LDV-FORUM bei gleichzeitig deutlich sinkender Arbeitsbelastung der Redaktion (vielleicht könnte dann sogar das korrekturelem eingeführt werden?!?). - Ein typischer Wiener Kaffeehausraum, fernab der Darmstädter Realität?

G.K

P.S.: Das LDV-FORUM in seinem gegenwärtigen Umfang wäre trotz eines sehr entgegenkommenden Angebots der Verlagsdruckerei Hoppenstedt nicht finanzierbar ohne den wesentlichen Beitrag der Anzeigen. Ich möchte den Firmen dafür herzlich danken und glaube, daß sich die Beachtung der Anzeigen auch für den Leser lohnen wird.

Last not least, Redaktionsschluß für das nächste Heft 2/1986 wird am 15. Oktober sein.

auf jeder Titelseite ein Buchstabe des Alphabets, wobei die Gestaltung einen Zusammenhang zur Linguistischen Datenverarbeitung herstellen soll. Dadurch ist ein einzelnes Heft zusätzlich noch ein Teil der ganzen Summe, nämlich nach 13 Jahren ein Teil des Alphabets.

Doch nun, nach dem wir ein einigermaßen brauchbares Konzept hatten, fing die Arbeit erst richtig an. Denn obwohl das LDV-Forum noch lange nicht über Fotosatzmaschine läuft, wollten wir doch wenigstens schon einen Teil unseres Konzeptes verwirklichen. Wir haben versucht die Schriftgrößen- und arten einigermaßen anzugleichen, wobei wir uns für das Verständnis der Autoren für unser Richtlinienblatt nochmals bedanken. Doch Schreibmaschinen, Matrix-Drucker und Laser-Drucker haben leider noch nicht die Qualität einer Fotosatzmaschine. Wir hoffen aber, daß man trotz der alten Fertigungsweise schon eine kleine Verbesserung sieht.

Fred Zimmermann

Themenschwerpunkt

GENERIERUNG VON ERKLÄRUNGEN AUS WISSENSREPRÄSENTATIONEN

Hannelore Rösner

Nixdorf Computer AG

Pontanusstraße 55

D-4790 Paderborn

Erklärungskomponenten von Expertensystemen beanspruchen. Computerleistung transparenter zu machen. Diesem Anspruch werden sie erst dann gerecht, wenn sie der Erwartung des Benutzers bezüglich einer Erklärung sowohl inhaltlich wie von der sprachlichen Form genügen. Im vorliegenden Beitrag wird auf der Basis eines verfügbaren Prototyps über Arbeiten an der Erklärungskomponente der Nixdorf Expertensystem Shell TWAICE berichtet, die auf eine erweiterte Erklärungsmächtigkeit und eine natürlichsprachliche Ausgabe zielen. Es wird eine Erklärungstypologie entwickelt, die Entscheidung für einen Grammatiktyp auf der Basis des gegenwärtigen Standes der Grammatiktheorie diskutiert und es werden die Probleme der Überführung formaler in natürlichsprachige Wissensrepräsentation speziell im Zusammenhang mit Fachsprachen behandelt. Randbedingungen, Architektur und Komponenten der Sprachgenerators werden beschrieben. Beispiele illustrieren die Generierung von Erklärungen.

1 Einleitung: Problemstellung

Neue und zukunftsstrahlende Technologie im Bereich der Softwareentwicklung sind derzeit Expertensysteme. Expertensysteme unterscheiden sich von herkömmlichen Programmen in erster Linie dadurch, dass sie a) Wissen in Form von Fakten und Regeln verarbeiten und b) nicht-deterministisch ablaufen. Wegen ihres nicht-deterministischen Ablaufs wird das Vorhandensein einer Erklärungskomponente als eines ihrer wesentlichen Bestandteile angesehen: die Arbeit des Programms soll nachvollziehbar sein, das Programm soll seine Schlussfolgerungen "erklären" können. Dies erfolgt bisher zumeist als

- Ausgabe der angewendeten Regeln in der Form, in der sie vom KE erstellt wurden,
- Angabe der Wissensquelle, mittels derer ein bestimmter Fakt geschlossen wurde.

Je nach Benutzerfreundlichkeit wird auf die Benutzerfrage Bezug genommen und zusätzlich ein sprachlicher Verbindungs-/Zusatztext formuliert. In der Regel sind solche sprachlichen Verbindungstexte Textkonserven mit Variablen (sogenannte Luekkentexte), die zur Laufzeit des Programms gefüllt werden.

Es wird argumentiert, dass solches Transparentmachen des Programm-Schlussfolgerungsweges insgesamt als ein Schritt auf die erhöhte Benutzerfreundlichkeit von Programmen zu sehen ist. Dies trifft jedoch bisher allenfalls fuer Experten zu:

- Die formalen Regeln sind fuer den nicht geschulten Benutzer zumeist schwer lesbar bzw. nur teilverstaendlich und ihre Verkettung dadurch kaum nachvollziehbar.

- Falls zusätzliche verbindende Texte formuliert werden, klingen diese zumeist sehr mechanisch: Wegen des

nicht-deterministischen Ablaufs aendern sich die Erklärungskontexte staendig, d.h. der Luekkentextrahmen muss so allgemein sein, dass er alle moeglichen Kontexte sprachlich korrekt abdecken kann. Waehrend im Englischen, das durch geringe Flexion und relativ starre Wortstellung gekennzeichnet ist, mit diesem Verfahren noch recht brauchbare Ergebnisse zu erzielen sind, werden im Deutschen mit zunehmender Groesse der Wissensbank die Texte sehr monoton und unnatuerlich, wenn nicht sogar inkorrekt.

- Erklärungen zu dem in der Wissensbank enthaltenen Wissen (z.B. Begriffe, Abkuerzungen u.a.m.) werden nicht gegeben.

Die folgende Abbildung aus einer Konsultation zur Konfiguration eines Rechners moege dies veranschaulichen:

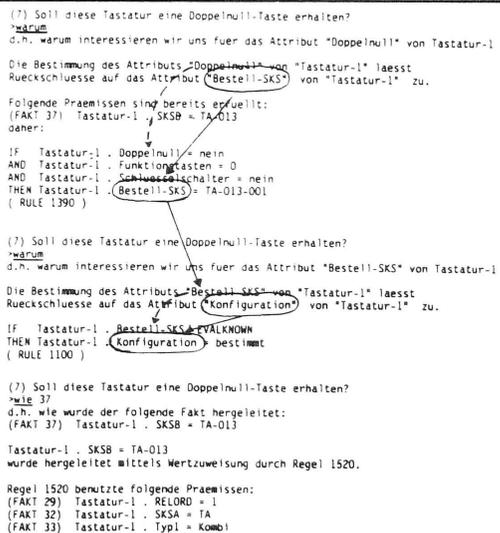


Abbildung 1: Mehrfacher Aufruf der TWAICE-Erklärungskomponente (warum: Regelverkneufungen; wie: Faktenherleitung)

Themenschwerpunkt

Im folgenden wird ueber den Stand der Arbeiten zu einer erweiterten Erklarungskomponente der Nixdorf Expertensystem Shell TWALICE berichtet. Ziel der Erweiterungen ist

- eine Erweiterung der Erklarungsmaechtigkeit,
- die naturlichsprachliche Ausgabe von Erklarungen.

Technische Details beziehen sich - sofern nicht anders angegeben - auf einen bereits erstellten Prototypen.

2 Erklarungsmaechtigkeit von Expertensystemen

2.1 Stand der wissenschaftlichen Forschung

2.1.1 Struktur von Erklarungen

Linguistisch gesehen sind Erklarungen

- a) dem Gegenstandsbereich der Pragmatik zuzuordnen, die Sprache als interaktives Handeln begreift,
- b) mit den Methoden der Textlinguistik und der Diskursanalyse zu beschreiben, da eine Erklarung meistens aus mehr als einem Satz besteht (jeder einzelne Satz also auch immer eine satzuebergreifende Funktion/ Bedeutung hat) und u.U. interaktive Komponenten hat.

Voraussetzung fuer eine Generierung von Erklarungen ist die Kenntnis ihrer Struktur. Hierunter verstehen wir

- a) auf der Makroebene:
die TYPOLOGIE VON ERKLAERUNGEN: sie gibt an, welche Typen von Erklarungen zu unterscheiden sind (z.B. Erklarung eines Begriffs, einer Schlussfolgerung usw.)
- b) auf der Mikroebene:
eine BESCHREIBUNG DER SPRACHLICHEN STRUKTUR jedes Typs als
 - sprachliche Struktur von Saetzen,
 - Anordnung dieser sprachlichen Strukturen in der Zeit und Beschreibung ihrer Verknuepfungen,
 - eine Beschreibung der zusaetzlich eingesetzten nichtsprachlichen Darstellungen (z.B. Graphik) soweit verwendet und ihre Zuordnung zu den sprachlichen Strukturen.

Um fuer praktische Belange brauchbar zu sein, muss eine solche Beschreibung darueberhinaus auf ein Benutzermodell Bezug nehmen.

Die wissenschaftliche Taetigkeit auf dem Gebiet der Erklarung ist allerdings bisher gering und liefert kaum Grundlagen fuer unser Vorhaben:

Detaillierte Untersuchungen finden sich nur im Bereich normativer wissenschaftlicher Erklarung, i.e. die deduktiv-nomologische und die statistische Erklarung (z.B. POPPER 1959, HEMPEL 1965, FORGE 1978, GLYMOUR 1978, SCPIVEN 1958).

Dieser Typ von Erklarung wird durch die klassische Erklarungskomponente eines Expertensystems nachgebildet. Die logische Struktur dieser Erklarung ist gut untersucht und dokumentiert. Ihre sprachliche Struktur wird dagegen nur implizit behan-

delt und ist gaenzlich auf den Wissenschaftler abgestellt, damit liefert sie keine Anhaltspunkte fuer die sprachliche Formulierung von Erklarungen an Laien-Benutzer eines Expertensystems.

Im paedagogisch-psychologischen Bereich, wo wir am ehesten eine Klassifikation verschiedener Erklarungsformen und ihrer sprachlichen Erscheinung erwartet haetten, sind wir entgegen unserer Erwartung nicht fuerdig geworden. Auch bei POLYA 1971, der zumindest sehr viele Beispiele fuer den mathematischen Unterricht gibt, aus denen sich Strukturen von (paedagogischen) Erklarungen gewinnen liessen (eine Auswertung wurde bisher nicht vorgenommen, da noch nicht eingeschaezt werden kann, welchen Verallgemeinerungsgrad diese Beispiele fuer EXPERTEN-Erklarungen haben) findet sich keinerlei systematische Darstellung von Erklarungsstrukturen.

Im angelsaechsischen Bereich von Expertensystemen und tutoriellen Systemen wird der Begriff der Erklarung ueberwiegend dem intuitiven Vorverstaendnis ueberlassen (vgl. dazu auch den Kommentar von WAHLSTER 1981, S. 11 ff. Dies ist umso ueberraschender als haeufig das Thema "Erklarung" das zentrale Thema der Arbeit ist (s. z.B. SLEEMAN/BROWN 1982, FININ 1983, CLANCEY 1983, GOGUEN/WEINER/LINDE 1983, SWARTOUT 1984, JACKSON/LEFRERE 1984, ANDERSON/REISER 1985). Eine Ausnahme macht WAHLSTER 1981, S. 11, der fuer fuenf Bedeutungsvarianten des Begriffs Erklarung aufuehrt:

- a) Erklarung von Tatsachen, Behauptungen, Ereignissen und Prozessen
- b) Erklaren des Funktionierens eines komplexen Gebildes
- c) Erklarung als Handlungsanleitung
- d) Erklarung von Begriffen
- e) Erklarung als durch die offizielle Funktion des Sprechers gestuetzte Festlegung der Feststellung.

Seine weiteren Ausfuehrungen beschaenkt er auf Typ a), auf den er die Erklarungskomponente eines Expertensystems beschaenkt wissen moechte. Eine Uebersicht ueber von Experten verwendete Strukturere dieses Erklarungstyps findet sich jedoch auch bei ihm nicht. Seine Beispiele setzen implizit voraus, dass die sprachliche Form dieses Erklarungstyps bekannt ist.

Im Bereich der Linguistik ist der Gegenstand "Erklarung" kein Forschungsthema. Es wurde nur eine Arbeit gefunden, die sich zentral damit beschaefigt (LANC 1976). Diese widmet sich wiederum nur dem Teilgebiet der sprachlichen Struktur

der deduktiv-nomologischen Erklärung bei Kindern im Grundschulalter. Die weiteren Arbeiten behandeln Erklärungen nur am Rande als eine Möglichkeit des Antwortens auf Fragen und setzen wiederum intuitives Vorverständnis voraus (z.B. KINDT 1976, HINDELANG 1980, McKEOWN 1982, LEHNERT 1978).

Empirische Untersuchungen gesprochener oder geschriebener Sprache mit dem Ziel der Identifikation und Klassifikation von Erklärungen liessen sich nicht finden. Besonderer Mangel fuer unsere Arbeiten ist darueberhinaus, dass nichts Systematisches darueber bekannt ist, wie EXPERTEN erkaeren: a) welche sprachlichen Merkmale die Erklarung auszeichnen und b) welche zusaetzlichen Mittel eingesetzt werden (z. B. Tabellen, Graphik, Histogramme).

2.1.2 Aufgabengebiet einer Erklarungskomponente von Expertensystemen

In der Frage, was die Erklarungskomponente eines Expertensystems leisten soll, herrscht derzeit noch Uneinigkeit (wofuer das Defizit der wissenschaftlichen Forschung auf dem Gebiet der Erklarungen sicherlich eine Ursache ist). Sie wird derzeit (grob gesehen) aus zwei Richtungen beantwortet:

- 1) Eine Richtung (insbesondere WEINER 1979, WAHLSTER 1981) moechte die Erklarungskomponente eines Expertensystems strikt einschaerken auf die "Erklarung von Tatsachen, Behauptungen, Ereignissen und Prozessen", d. h. auf das Transparentmachen einer logischen Schlussfolgerungskette. Eine derart eingeschaerkte Erklarungskomponente ist State-of-the-Art heutiger Expertensysteme.
- 2) Die andere Richtung (vertreten vor allem durch HASLING/CLANCEY/RENNELS 1984, SWARTOUT 1983, SWARTOUT 1984, WILENSKY/ARENS/CHIN 1984, sowie von Benutzerseite vertreten durch HAND 1985) moechten Erklarungen eher tutorieller Art wie z. B. Erklarungen von Begriffen und Erklarungen, wie etwas funktioniert oder wie etwas gemacht wird integriert sehen, wobei allerdings keiner dieser Autoren den Skopus dessen, was er unter Erklarung verstanden wissen moechte, hinreichend praezisiert.

Wir meinen, dass WAHLSTER 1981 zwar Recht hat, wenn er die Erklarungskomponente vom Begriff 'Frage-Antwort-System' noch unterscheiden koennen moechte, doch bedarf es unseres Erachtens dazu nicht der von ihm vorgenommenen rigorosen Beschraenkung.

Wir gehen davon aus, dass die Erklarungskomponente eines Expertensystems der

Transparenz des Systems fuer den Benutzer dienen soll. Sie sollte unseres Erachtens daher alle Typen von Erklarungen umfassen, die dieser Transparenz dienen.

WAHLSTERS oben angefuehrte fuenf Bedeutungsvarianten des Begriffs Erklarung lassen sich unseres Erachtens so zusammenfassen, dass der Benutzerwunsch nach Erklarung staerker hervortritt:

informierende Erklarungen:

- dynamischer Art:
Erklarungen von Tatsachen, Behauptungen, Ereignissen und Prozessen (z. B.: A erkaert, warum das Symptom X fuer die Diagnose von Bedeutung ist)
- statischer Art:
Erklarungen von Begriffen, Funktionen und Handlungskonzepten (z. B.: A erkaert, was ein Symptom ist, wie ein Blutdruckmesser arbeitet, wie ein Test durchzufuehren ist)

feststellende Erklarungen:

Erklarung als durch die offiziell Funktion des Sprechers gestuetzte Festlegung oder Feststellung (z.B: A erkaert, dass die Sitzung eroeffnet ist).

Unserer Auffassung nach fallen alle informierenden Erklarungen in den

Aufgabenbereich einer Erklarungskomponente.

2.2 Entwurf einer Erklarungstypologie

Da uns WAHLSTERS Typologie fuer die konkrete Arbeit als zu rudimentaer erschien, haben wir eine differenziertere Typologie entwickelt. Wir betrachten sie z.Zt. als den Teil einer Erklarungstypologie, die den Bereich der Experten-Erklarungen abdecken soll (d.h. wir interessieren uns z.Zt. nicht dafuer, wie vollstaendig sie z.B. in bezug auf Erklarungen im Bereich schulischen Unterrichtens ist), und auch hier koennen wir z.Zt. noch keine Aussagen ueber ihre Maechtigkeit bzw. Vollstaendigkeit machen.

Wir unterscheiden folgende Erklarungen:

- a) Erklarungen zu statischem Wissen:
Unter statischem Wissen verstehen wir die Menge der einzelnen Wissensstuecke BEVOR sie im konkreten Fall angewendet werden. In einem Expertensystem ist dies das in der Wissensbank formalisierte Wissen. Wir identifizieren bisher Wissen ueber:

Abkuerzungen:

Wissen ueber Abkuerzungen beinhaltet die ausfuehrliche Form eines Begriffes und eine oder mehrere Abkuerzungen dafuer.

Themenschwerpunkt

Begriffe:

Unter Wissen ueber Begriffe verstehen wir

- die Definition eines Begriffes,
- die dem Experten wichtige Ueberblick-Information (strukturierte Form und gestufte Informationstiefe) zum Begriff,
- Angaben darueber, wo man weitere Information zu diesem Begriff finden kann.

Beziehungen zwischen Begriffen:

Mit Wissen ueber die Beziehungen zwischen Begriffen meinen wir in welcher Relation ein Begriff zu anderen steht (z.B. X ist ein Wert von Y; Z kann sein ein Teil von U oder V).

Verwendung von Begriffen:

Mit Wissen ueber die Verwendung von Begriffen meinen wir den Gebrauch des Begriffes durch den Experten (bei der Verwendung formaler Regelsprache zur Wissensrepraesentation die Verwendung des Begriffes in der Regel).

Regeln:

Mit Wissen ueber Regeln ist die inhaltliche Bedeutung einer Regel gemeint, z.B.: 'Regel 2 dient dazu, optional einen Ergebniszwischenausdruck vorzunehmen').

Beziehungen zwischen Regeln:

Wissen ueber Beziehungen zwischen Regeln beinhaltet die inhaltliche Bedeutung mehrerer Regeln, z.B.: 'Regeln 2-5 dienen dazu, die Ergebnisausgabe zu steuern. Die jetzige Reihenfolge bewirkt ...').

Verwendung von Regeln:

Wissen ueber die Verwendung von Regeln meint das Wissen ueber Anwendungskontexte, d.h. Wissen darueber, in welchen Kontexten eine Regel ueberhaupt angewendet wird.

Strategien:

Mit Wissen ueber Strategien meinen wir das Wissen des Experten ueber die Teilgebiete der Wissensdomaene und ihre Abhaengigkeiten. So wissen Experten z.B., in welcher Reihenfolge welche Teilgebiete zu erfragen/erschliessen sind, damit sie moeglichst schnell zum Ziel kommen und nicht in Sackgassen laufen. Solch strategisches Wissen kann sein:

- konsultatives Wissen: Wissen ueber die Beratungssituation,
- fachbezogenes Wissen: Wissen ueber domaininterne Abhaengigkeiten.

Beziehungen zwischen Strategien:

Wissen ueber Beziehungen zwischen Strategien meint Wissen darueber, ob innerhalb der Beratungssituation eine Strategie angewendet wird oder falls es mehrere ist, ob diese unabhaengig voneinander sind bzw. in welcher Reihenfolge sie angewendet werden und ob sie einer uebergreifenden Strategie zugeordnet werden koennen.

Verwendung von Strategien:

Wissen ueber die Verwendung von Strategien beinhaltet das Wissen darueber, warum eine bestimmte Strategie angewendet wird und nicht eine andere.

b) Erklarungen zu dynamischem Wissen:

Mit dynamischem Wissen meinen wir das in einer Beratungssituation anfallende Wissen sen ueber den individuellen Beratungsfall Zum dynamischen Wissen gehoert Wissen ueber:

Die aktuelle Regelverknuepfung:

Hierunter verstehen wir die Kette der bis zur warum-Frage angewendeten Regeln.

Nicht angewandte Regeln:

Nicht anwendbare Regeln:

- Hierunter verstehen wir die bis zur warum-Frage fehlgeschlagenen Regeln,
- (noch) nicht angewandte Regeln: Hierunter verstehen wir die bis zum Zeitpunkt der warum-Frage noch nicht bearbeiteten Regeln.

Abgeleitete Fakten:

Hierunter verstehen wir die Menge der bis zum Zeitpunkt der wie-Frage abgeleiteten Fakten.

Nicht abgeleitete Fakten:

- Nicht ableitbare Fakten: Hierunter verstehen wir die Menge der bis zum Zeitpunkt der wie-Frage ausgeschlossenen Fakten,
- (noch) nicht abgeleitete Fakten: Hierunter verstehen wir die Menge der bis zum Zeitpunkt der wie-Frage noch abzuleitenden Fakten.

Aktuelle Strategie:

Hierunter verstehen wir die Verknuepfung der bis zum Zeitpunkt der wozu-Frage angewandten Strategien.

Nicht angewandte Strategien:

- Nicht anwendbare Strategien: Hierunter verstehen wir die Verknuepfung der bis zum Zeitpunkt der wozu-Frage angewandten Strategien,
- (noch) nicht angewandte Strategien:

Hierunter verstehen wir die Menge der bis zum Zeitpunkt der wozu-Frage noch nicht beruehrten Strategien.

Dabei ist dieses Wissen in bisherigen Expertensystemen haeufig nur zum Teil explizit formuliert: So ist der Aufbau der Wissensbank kondensierte Expertenstrategie, die Strategie/n selbst ist/sind jedoch nur mehr implizit vorhanden.

Semantisch notwendige Information zum Sachverhalt Erklarung durch ein Expertensystem wird beschrieben durch:

- Wissen ueber den geforderten Erklarungstyp,
- Wissen ueber den jeweiligen Benutzer,
- das formale Wissen zur Laufzeit:
 - fuer Erklarungen zu statischem Wissen: die geladene Wissensbank,
 - fuer Erklarungen zu dynamischem Wissen: der aktuelle Stack.

3 Grammatikmodelle fuer die Generierung natuerlicher Sprache

3.1 Stand der wissenschaftlichen Forschung

Unter Generierung von Sprache verstehen wir eine Menge von Verfahren, die auf der Grundlage einer (formalen) semantischen Information zu einem Sachverhalt mit Hilfe einer Grammatik korrekte Saetze dieser Sprache ueber diesen Sachverhalt erzeugt.

Der derzeitige Stand der Grammatiktheorie ist gepraeagt durch eine Fuelle konkurrierender Ansaetze, die eine praktische Entscheidungsfindung sehr erschweren. Die folgenden Ausfuehrungen versuchen zunaechst, den Hintergrund der derzeitigen Situation aus unserer Sicht zu beleuchten, um hieraus Konsequenzen fuer die Entscheidungsfindung abzuleiten.

Die Entwicklung von Grammatiken ist durch unterschiedliche wissenschaftliche Motivationen gepraeagt:

a) Linguistik:

Zentrales Anliegen der linguistischen Grammatikmodelle ist traditionell die Produktion/Generierung korrekter Saetze der natuerlichen Sprache (im Gegensatz zur Analyse) gewesen. Fragen der Analyse haben bis in die sechziger Jahre hinein nur eine sehr untergeordnete Rolle in den folgenden (sehr heterogenen) Randbereichen gespielt

- Aufloesung von Ambiguitaeten,
- Erklarung, warum eine bestimmte Form grammatikalisch inkorrekt ist,
- Erklarung der Schnelligkeit des Spracherwerbs,
- kontrastive Fehleranalyse beim Fremdspracherwerb.

Themenschwerpunkt

Im folgenden sollen - um die verschiedenen linguistischen Forschungsbemuehungen

richtig einschaeetzen zu koennen - zunaechst die verschiedenen wissenschaftlichen Ansprueche der Linguistik (die sicherlich historisch gepraeagt sind) kurz skizziert und in bezug auf ihre moegliche Bedeutsamkeit fuer die maschinelle Sprachgenerierung kommentiert werden.

normativer Anspruch:

Ziel der Grammatik ist es, detaillierte Normen fuer korrekte (Schrift-)sprache anzugeben. Diese Grammatiken (in erster Linie historische Grammatiken und z.T. Grammatiken fuer den Fremdspracherwerb) sind fuer das Design von Systemen zur maschinellen Sprachanalyse und -generierung bedeutungslos. Bei der konkreten Ausarbeitung von Systemen erweisen sie sich jedoch als praktisch nuetzlich, da sie zumeist die groesste Ansammlung von (haeufig) heuristischen Regeln zur Bildung korrekter Spracheinheiten (einschliesslich Ausnahmen) enthalten.

Anspruch auf Beobachtungsadaequatheit:

Ziel der Grammatik ist es, ein gegebenes Sprachkorpus adaequat zu beschreiben. Fragen der Norm spielen dabei keine Rolle. Diese Grammatiken (in erster Linie strukturalistische Grammatiken und neuere Grammatiken fuer den Fremdsprachenunterricht) dienen in erster Linie der Entwicklung von Methoden zu einer einzelsprachunabhaengigen systematischen Sprachbeschreibung. Fuer die maschinelle Sprachanalyse und -generierung sind die dort entwickelten Verfahren praktisch bedeutsame Grundlagen fuer die Entwicklung leistungsfaeiger Pattern Matcher (es entzieht sich allerdings unserer Kenntnis, wie weit ein Pattern Matcher tatsaechlich einmal auf Detailarbeiten zurueckgegriffen haette).

Anspruch auf deskriptive Adaequatheit:

Ziel der Grammatik ist es, Sprache adaequat zu beschreiben, sodass auf der Grundlage dieser Beschreibung automatisch korrekte Saetze dieser Sprache produziert und analysiert werden koennen. Grammatiken, die im Rahmen dieses Anspruchs formuliert wurden (in erster Linie kontextfreie Phrasenstrukturgrammatiken), waren Ausgangspunkt fuer maschinelle Uebersetzungsprojekte. Mit der Formulierung der Theorie der Transformationsgrammatik und ihrem Anspruch auf explanative Adaequatheit gerieten sie jedoch schnell in den Ruf eines veralteten Paradigmas. Neuere linguistische Ansaetze - insbesondere unter dem Druck der Konkurrenz von informatischer Seite - besinnen sich heute jedoch wieder auf dieses Paradigma (insbesondere GAZDAR/KLEIN/PULLUM/SAG 1985) und betreiben seine Weiterentwicklung, wobei sie heute jedoch am Anspruch der explanativen Adaequatheit gemessen werden.

Themenschwerpunkt

Anspruch auf explanative Aadaequatheit: Ziel der Grammatik ist es, ein gegebenes Sprachkorpus aadaequat zu beschreiben, gleichzeitig aber ein Modell zur Verfuellung zu stellen, das Phaenomene der Sprachproduktion, des -verstehens und des -erwerbs zu erklaren vermag. Bieten sich mehrere Modelle an, so wird das Kriterium der Einfachheit als Entscheidungskriterium angewendet. Heutige linguistische Grammatiken haben saemtlich den Anspruch auf explanative Aadaequatheit und Einfachheit des Modells und viele der Detaildiskussionen einzelner Saetze sind in erster Linie empirische Auseinandersetzung mit der Frage, ob ein gegebener Ansatz dem Anspruch auf explanative Aadaequatheit ueberhaupt genuegen kann und ob nicht vielmehr (bevor man einen Sackgassenweg geht) ein anderer (ggf. nur um Nuancen verschobener) Ansatz zu waehlen sei (vgl. z.B. die Diskussion um die Behandlung der kausativen Verben).

Die wichtigsten linguistischen Grammatikmodelle heute sind (in Klammern jeweils nur die Hauptwerke der einzelnen Richtungen):

- Transformationsgrammatik und ihre Modifikationen EST, REST und GB (CHOMSKY 1965, CHOMSKY 1975, CHOMSKY 1977, CHOMSKY 1981) oder Derivate (z.B. Generative Semantik),
- Lexical Functional Grammar (BRESNAN 1978, BRESNAN/KAPLAN 1982),
- Generalized Phrase Structure Grammar (GAZDAR 1982),
- Kasusgrammatik (FILLMORE 1969),
- Montague-Grammatik (MONTAGUE 1974),
- Dependenzgrammatik (TESNIERES 1959, HAYS 1964).

b) Informatik:

Zentrales Anliegen der von der Informatik entwickelten Grammatikmodelle ist immer die robuste Analyse von korrekten wie auch inkorrekten Saetzen der natuerlichen Sprache gewesen.

Die wichtigsten hier entwickelten Grammatiken sind (in Klammern jeweils nur die Hauptwerke der einzelnen Richtungen):

- Definite Clause Grammar (PEPEIRA/WARREN 1980)
- Functional Unification Grammar (KAY 1984)
- Augmented Transition Networks (WOODS 1970) und Cascaded Augmented Transition Networks (WOODS 1980).

Schwerpunkt wissenschaftlicher Arbeit im Bereich der maschinellen Sprachverarbeitung war stets die Analyse, die

- auf Grund praktischer Erfordernisse (Abfrage von Datenbanksystemen, maschinelle Analyse fremdspraechiger Texte) staerker gefragt war,

- als schwieriger (und damit wissenschaftlich wertvoller) als die Generierung galt (eine Einschaeztung, die vermutlich durch den Focus auf der englischen Sprache verursacht ist, bei der bereits mit Hilfe von Lueckentexten recht gute Texte generiert werden koennen).

Waehrend eine reichhaltige und auch (implementations-)technisch detaillierte Literatur zu Parsern existiert, sind die Arbeiten zur Generierung eher duenn gesaet, ueberwiegend schwer zugaenglich und zumeist noch im Stadium der Konzeption(s. dazu MANN/BATES/GROSZ/McDONALD/McKEOWN/SWARTOUT 1982, McKEOWN 1982, MORIK 1985). Erste systematische Anforderungen an eine Generierungskomponente wurden erst vor wenigen Jahren formuliert (MANN/BATES/McDONALD/McKEOWN/SWARTOUT 1981), die eigentliche Forschung auf diesem Gebiet hat erst begonnen. Allerdings zeichnet sich derzeit eine Aenderung auch fuer den deutschsprachigen Bereich ab (vgl. z.B. die im KI-Rundbrief Nr. 37 angegebene Literatur). Die wenigen bisher zugaenglichen Arbeiten sind jedoch noch so allgemeiner oder grundlagenmaessiger Natur, dass sie fuer die unten beschriebenen Generierungsprobleme (noch?) nicht brauchbar waren (DANLOS 1983, DANLOS 1984, BUSEMANN 1984a, ein paar interessante Hinweise auf einen dreistufigen Generierungsprozess komplexerer Texte finden sich in KATZ 1980) oder so an den entsprechenden Anwendungskontext und die Programmiersprache gebunden, dass sie allenfalls fuer moegliche Ideen interessant erschienen (LEHNERT 1978, McKEOWN 1983, McKEOWN 1985). Insbesondere auf dem Gebiet der Morphologie und Wortstellung des Deutschen konnten wir auf keine Veroeffentlichungen zurueckgreifen (Inzwischen liegt uns die Arbeit von BUSEMANN 1984b vor, die etliche Regeln fuer die deutsche Morphologie formuliert).

Der derzeitige Stand der Grammatiktheorie kann wie folgt zusammengefasst werden:

- Innerhalb der Linguistik verhindern der Streit um die explanative Aadaequatheit der Modelle in Verbindung mit nicht-integrierter Detailforschung die Formulierung einer Grammatik (oder zumindest der technologischen Methoden fuer ihre Formulierung), die mehr als die Generierung eines kleinen Beispielkorpus leistet.

- Innerhalb der Informatik werden quer dazu Modelle entwickelt, die fuer die maschinelle Sprachverarbeitung geeignet sind, die jedoch keinen Anspruch darauf erheben, ein Modell der natuerlichen Sprache zu sein.

Leider findet sich bisher weder von linguistischer noch von informatischer Seite eine Arbeit, die die verschiedenen Modelle

3.2 Wahl eines Grammatikmodells

Der derzeitige Stand der Grammatiktheorie erlaubt u. E. keine Ableitung klarer Empfehlungen bezüglich der Wahl eines Grammatikmodells, wir haben deshalb heuristische Argumente fuer die Entscheidungsfindung zusammengestellt:

Fuer unsere Belange sollte ein Grammatikmodell gewaehlt werden, dass fuer Analyse und Generierung gleichermaßen gut geeignet ist, so dass einmal entwickelte Grammatikteile moeglichst haeufig doppelt genutzt werden koennen.

Fuer die Verwendung einer Definite Clause Grammar sprachen mehrere Gruende:

- Wir verwenden Prolog und DCGs sind sehr einfach in Prolog formulierbar.

- Durch die Inversitaet von Prolog-Programmen (Variablen sind Output-Parameter, Konstanten Input-Parameter) ist eine Benutzung zumindest von Teilen der Grammatik sowohl fuer Generierung wie auch Analyse erleichtert.

- DCGs erlauben eine recht einfache Erweiterung der Grammatik durch Hinzufuegen weiterer Klauseln. Dies erlaubt die sehr modulare Entwicklung von Grammatiken und ist daher besonders fuer die Formulierung einer Grammatik fuer fachsprachliche Texte vorteilhaft, deren Syntax relativ unbekannt ist und bei der damit zu rechnen ist, dass im Verlaufe der Zeit eine Reihe von Ausnahmekonstrukten zusaetzlich zu behandeln ist.

- DCGs sind sehr flexibel, indem sie zwar kontextfrei geschrieben werden, es aber durch einfache Variablenuebergabe sehr einfach machen, Kontexte zu transportieren und zu beruecksichtigen (und erlauben dadurch z.B. sehr einfache Formulierung der Subjekt - Verb - Kongruenz).

- DCGs bieten sich auf Grund ihrer Flexibilitaet an, bei Bedarf auch hybride Grammatikmodelle zu implementieren. So wurden DCG-Formalismen bereits als Grundlage fuer die Implementierung von LFG-Parsern verwendet (REYLE/FREY 1983, YASUKAWA 1984) verwendet. PEREIRA/WARREN 1980 beschreiben einen informellen Algorithmus zur Ueberfuehrung von ATNs in DCGs.

so aufeinander bezieht, dass bezogen auf den Level der Formulierung von sprachlichen Eigenschaften

- a) die Grenzen der Modelle klar bezeichnet sind,
- b) Vor- und Nachteile der Modelle klar hervortreten.

(Dies trifft unseres Erachtens auch auf die Arbeit von BERWICK 1984 zu, der zwar einen guten Ueberblick ueber eine Fuelle verschiedener neuer Ansaetze gibt, aber der Terminologie der jeweiligen Autoren verhaftet bleibt und daher eher unterschiedliche theoretische Sichtweisen beschreibt als prinzipielle Gemeinsamkeiten bzw. Differenzen und deren Implikationen fuer die Leistung des Modells).

U.a. laesst die veroeffentlichte Literatur z.B. unklar,

- was eine DCG von irgendeiner in Prolog entwickelten Grammatik unterscheidet (vg. auch PEREIRA/WARRENs informellen Algorithmus zur Ueberfuehrung von ATNs in DCGs, REYLE/FREYS sowie EISELEs

Implementationen einer LFG in Prolog) und

- warum eine FUG (KAY 1984) nicht einfach als eine Realisierung einer LFG beschrieben wird.

Unsere Einschaeztung der Lage ist wie folgt:

Bei den innerhalb der Informatik entwickelten Modellen handelt es sich weniger um Modelle der natuerlichen Sprache, sondern im Vordergrund stehen Verfahren zur Verarbeitung natuerlicher Sprache. Bei den innerhalb der Linguistik entwickelten Modellen handelt es sich dagegen um Modelle der natuerlichen Sprache, bei denen das Verfahren der Verarbeitung entweder noch offen oder zumindest von zweitrangiger Bedeutung ist. Es erscheint uns unangemessen, die verschiedenen Modelle - wie es haeufig geschieht - gleichberechtigt nebeneinander stehen zu lassen. Wir behaupten, dass die Verwendung von in der Informatik entwickelten Verfahren Moeglichkeiten bietet, linguistische oder auch psycholinguistische Modelle rechnergemaess zu implementieren.

Ziel der vorliegenden Arbeit ist jedoch nicht, dem derzeitigen wissenschaftlichen Streit bis in seine Veraestelungen zu folgen oder ihn z.B. durch den Nachweise von Aequivalenzen wissenschaftlich zu klaeren zu versuchen, sondern einen Beitrag zur korrekten Generierung definierter Erklaerungstypen aus vorgegebenen semantischen Strukturen zu geben, wobei der Stand der wissenschaftlichen Forschung beachtet werden soll.

Themenschwerpunkt

- 4 Generierung aus formalen Wissensrepräsentationen
- 4.1 Stand der wissenschaftlichen Forschung

Unter formaler Wissensrepräsentation verstehen wir die eindeutige Darstellung von Wissen in einer formalen Sprache. Wissen, das von Expertensystemen verarbeitet wird, ist immer formalisiertes Wissen. Verwendete Technologien heute sind Prädikatenlogik (Hornklauseln), semantische Netze, Frames und mit einem besonderen Schwergewicht Produktionsregeln. Im Prinzip handelt es sich jedoch stets um einen Ausschnitt Prädikatenlogik. Damit ist das Problem der Generierung natürlicher Sprache zunächst das Problem der Ueberführung eines wohldefinierten Stuecks Prädikatenlogik in natürliche Sprache. Dieses Gebiet - wenn auch in umgekehrter Richtung (Ueberführung natürlicher Sprache in Prädikatenlogik) - ist ein intensives Forschungsgebiet der Semantik. U. a. liegen bereits technisch detaillierte Ansätze einer automatischen Uebersetzung vor (z.B. BOHNER/BACKER 1979, FRIEDMANN 1981).

In der Praxis zeigt sich jedoch, dass der syntaktisch und semantisch wohldefinierte Ausschnitt Prädikatenlogik nicht notwendigerweise eine syntaktisch und semantisch wohldefinierte Basis fuer die Generierung natürlichsprachlicher Sätze sein muss. So wird man etwa bei der Technologie der Expertensystem-Shell eine allgemeine Syntax und Semantik fuer die formale Wissensrepräsentation bereitstellen, innerhalb derer der Knowledge Engineer frei ist, sein Wissen zu formulieren. Die Wissensrepräsentation der TWAICE-Expertensystem-Shell erfolgt in erster Linie durch Produktionsregeln gemaess dem MYCIN-Paradigma:

```
IF <Objekt> . <Attribut> = <Wert>
THEN <Objekt> . <Attribut> = <Wert>
```

MYCIN selber verwendete zumeist (immer?) eine bestimmte Sprachform:

```
Objekt <-> Nomen
Attribut <-> Nomen
Wert <-> Adjektiv (wobei das Attribut
im amerikanischen Englisch
haeufig ein substantiviertes
Adjektiv ist)
```

Jedoch, je nach der Komplexitaet der Domaene, der Sprache der Experten und der Praeferenz des Knowledge Engineer finden sich in den einzelnen semantischen Einheiten z.T. sprachlich sehr komplexe Konstrukte wie z.B.:

```
IF Drucker . Folgerung2 = Papier fehlt
AND Drucker . Aktion Papier einspannen
durchgefuehrt = ja
```

```
AND Drucker . jetzt gruene Lampe
Bedienfeld = nein
THEN Drucker . Stoerung = Papiersensor
bzw.
```

```
IF Theorie/Praxis . Periodenwahl =
Ich moechte fabrikwoechentlich
disponieren
AND Theorie/Praxis . Wahl der
Anpassungsart bei Korrektur der
Perioden bei taeglichem Disp.
Wechsel = durch Manipulation
der Sicherheitszeit
THEN Comet . Periodenwahl (getroffen) =
woechentlich
```

Wollte man also die bestehenden Ansätze fruchtbar verwenden, so waere

- a) eine Ueberführung der natürlichsprachlichen Teile der formalen Wissensrepräsentation in Prädikatenlogik notwendig, wobei
 - das Gebiet der Analyse fachsprachlichen Telegrammstils nicht erforscht ist sowie
 - die dort vorfindlichen Konstrukte erheblich komplexer sind als die bisher bearbeiteten,

- b) eine Verknuepfung dieses Stuecks Prädikatenlogik mit der hierarchisch uebergeordneten des Wissensrepräsentationsformalismus notwendig ist,

- c) eine Ueberführung dieses Konstrukts in natürliche Sprache zu leisten ist.

Abgesehen davon, dass die bisherigen Ansätze weit entfernt von der Bewaeltigung derart komplexer Konstrukte sind, erscheint es aus pragmatischen Gruenden nicht sehr sinnvoll, alle Sprachteile, die von der Generierung direkt verwendet werden koennten, zunächst in eine praedikatenlogische Darstellung umzuformen.

Es liesse sich einwenden, dass die hier aufgezeigten Probleme selbstverursacht sind, naemlich durch Vorgabe eines bestimmten Wissensrepräsentationsformalismus und dass sie bei Verwendung einer geeigneten Mischung verschiedener Wissensrepräsentationsformalisen, wie sie beispielsweise in sogenannten hybriden Systemen angestrebt wird, gar nicht erst auftreten wuerden. Wir meinen jedoch, dass eine solche Sichtweise nicht in Rechnung stellt, dass

- auch bei hybriden Systemen nicht gewährleistet ist, dass die Semantik des natürlichsprachlichen Konstrukts der Semantik des betreffenden Formalismus entspricht,

- das Konzept der Erklärungskomponente fuer hybride Systeme ungleich komplexer und vor allem auch unangereicherter als bei nicht hybriden ist,

- hybride Systeme - zumindest derzeit noch - mit einem Verlust an Einfachheit und Komfort der Wissensrepräsentation fuer den Knowledge Engineer verbunden ist.

Am einfachsten waere es, zu komplexe bzw. mit bestimmten Eigenschaften behaftete Konstrukte zu verbieten. Aus unserer Kenntnis verschiedener Wissenbanken meinen wir jedoch jetzt schon sagen zu koennen, dass Experten bei der formalen Darstellung ihres Wissens nahezu automatisch zur Verwendung eines Telegrammstils in Kombination mit Relationen tendieren. Die Syntax der Regelsprache kommt diesem Stil durchaus entgegen. Sie ist auf Grund ihrer Offenheit fuer beliebige sprachliche Konstrukte geeignet fuer so verschiedene Bereiche wie den kaufmaennischen (der zum Ausformulieren neigt) UND den technischen (der im wesentlichen mit Abkuerzungen, Indizierungen und nur wenigen Wortkategorien auskommt). Sicherlich werden - will man natuerlich-sprachliche Saetze generieren - einige syntaktische Beschraenkungen notwendig werden (z.B. (vereinfacht): als Abkuerzung wird interpretiert, was grossgeschrieben ist), doch sollte dies nicht die Einfachheit des Gebrauchs des Formalismus und sein leichtes Verstaendnis fuer den Experten beeintraehtigen.

4.2 Wahl der formalen Wissensrepräsentation

Ausgangspunkt fuer die Generierung von Erklärungen sind fuer uns z.Zt. noch die zur Laufzeit vorhandenen Wissensrepräsentationen:

- Ziele oder Teilziele der Inferenzkomponente,
- Voraussetzungen der jeweiligen Ziele,
- Fakten oder Faktenmengen,
- Regeln oder Teilregeln,
- taxonomisches Wissen,
- Wissen ueber Fragen, Hilfen, Prozeduren, Tabellen, Standardfunktionen, Formate.

Um den oben zitierten fachsprachlichen (z.T. grammatikalisch unkorrekten) Telegrammstil korrekt verarbeiten zu koennen, sind wir von folgender Ueberlegung ausgegangen: Expertensprache setzt u.E. auf einem sprachlichen Komplexitaetslevel an, der in vielen Faellen der Generierung nicht bis auf Wortebene, sondern nur auf z.B. Nominalphrasen-, Verb- und Auxiliarebene heruntertransformiert zu werden braucht. Wir haben uns fuer daher fuer den Bau eines Analysewerkzeuges entschieden, das

1) die in der Taxonomie abgelegten (komplexen) Ausdruecke einliest, syntak-

tisch analysiert und in einer Lexikonvorstufe ablegt (aus der spaeter im Dialog das endgueltige Lexikon erstellt wird. Diese Form der Analyse ist ein Novum, da sie der LexikonERSTELLUNG dient und nicht wie bisherige Analysen auf ein Lexikon zurueckgreifen kann),

2) syntaktisch/semantisch aufeinander bezieht und zu jedem Objekt-Attribut-Paar einen Generierungsvorschlag in Form von Satzbauteilen/Syntagmen ablegt, auf den die Generierung zurueckgreift.

Ausgangspunkt fuer die Generierung von Erklärungen sind fuer uns damit langfristig neben den zur Laufzeit vorhandenen Wissensrepräsentationen zusaetzlich aus der Analyse hervorgegangene Generierungsvorschlaege.

4.3 Auswahl eines Sprachkorpus

Um zu gewaehrleisten, dass die zu entwerfende Grammatik wesentliche sprachliche Eigenschaften von Expertensprachen zu behandeln in der Lage waere, war es notwendig, diese vorab zu definieren.

Da wir auf keine Vorarbeiten in bezug auf die Deskription von Expertensprache allgemein oder Expertenerklärungen im besonderen zurueckgreifen konnten, haben wir auf vorhandene Sprachkorpora in Form eigener uns zur Verfuegung stehender Wissenbanken zurueckgegriffen. Wir haben drei Wissenbanken ausgewaehlt, die sehr verschiedene sprachliche Stile repräsentieren, von denen wir meinen, dass sie

- a) typische Expertenfachsprachen repräsentieren
- b) trotz aller Verschiedenheit doch Hinweise auf gemeinsame syntaktische Eigenschaften von Fachsprachen geben.

In bezug auf den Verallgemeinerungsgrad unserer Ergebnisse ist dieses Vorgehen jedoch nicht ganz unproblematisch: Da die Wissenbanken in der Regelsprache der TWACE Expertensystem-Shell geschrieben sind, kann nicht ausgeschlossen werden, dass ein Teil der gemeinsamen syntaktischen Eigenschaften durch den Formalismus dieser Regelsprache nahegelegt wird. Wir meinen jedoch, dass

- a) Expertensprache immer auch bereits in hohem Masse formalisierte Sprache ist,
- b) JEDE Formalisierung von Sprache Einfluss auf Syntax und Semantik hat,
- c) die Regelsprache, insbesondere mit ihrer Moeglichkeit der Verwendung komplexer sprachlicher Ausdruecke im Telegrammstil der natuerlichen Formulierung von Experten sehr entgegenkommt, also eher sprachlich reichere Formen zulässt als z.B. eine Framedarstellung.

Themenschwerpunkt

Drei Beispiele moegen zunaechst einen Einblick in den unterschiedlichen Gebrauch der Regelsprache durch verschiedene Experten unterschiedlichen Wissensgebietes geben:

- a) Konfiguration eines Computers durch einen Experten aus dem Technischen Kundendienst:

```
IF Arbeitsplatz . Peripherie = bestimmt
AND Lobby-GAA . Konfiguration = bestimmt
THEN Arbeitsplatz . Geraete = bestimmt
```

```
IF Display . Anzahl = 2
AND Arbeitsplatz . Rechner im Display KNOWN
AND Display . SKS2-Nr = kopiere erstes
AND Arbeitsplatz . Display-SKS = BA-11
THEN Display . Zwischen-SKS = BA-062
```

Typische sprachliche Eigenschaften dieser Wissensbank sind:

- Nominalstil
- Verwendung typischer Abkuerzungen (z.B. SKS-Nr).

- b) Zusammenstellung eines Versicherungspaketes durch einen Experten aus der Programmierung solcher Probleme:

```
IF K . Geschlecht = m
AND L . Endalter < 62
THEN L . BU_Alder = L . Endalter
```

```
IF K . Gefahrengruppe = A
AND U . INV_2 = nein
THEN U . Tod_F = 168
```

Typische sprachliche Eigenschaften dieser Wissensbank sind:

- extrem hohe Verwendung individueller Abkuerzungen (Schreibfaulheit des Experten/Programmierers)
- die verwendeten syntaktischen Kategorien sind beschaenkt auf Nomen und Adjektive.

- c) Parametrisierung eines kommerziellen Softwarepaketes durch einen Systemingenieur, der fuer den Entwurf der Wissensbank das Skript-Modell fruchtbar zu machen versuchte (das Schreiben in dieser Form setzt in der Praxis voraus, dass Antwortmenues und ein Spelling Korrektor vorhanden sind, damit die Werte nicht in der Langform eingegeben werden muessen. Beides ist in TWAIICE der Fall):

```
IF Deter.Methoden . Wird Teil als
    einzelne(s/r) Bau(teil/kasten)
    verwandt = ja
AND Deter.Methoden . Teil wird als
    standardisierte(s/r) Bau(teil/
    kasten) verwandt = nein
AND Deter.Methoden . Status des Teiles
    = Baukasten und kein Standardteil
```

```
IF Theorie/Praxis . Periodenwahl =
    Ich moechte fabrikwoeentlich
    disponieren
AND Theorie/Praxis . Frage nach Ab-
    gleich Periode Bedarfszeitpunkt
    = Woeentliche Dispo fuehrt bei
    der Fertigung zu Verzoegerungen
AND Theorie/Praxis . Wahl der
    Anpassungsart bei Korrektur der
    Perioden = Ich bin bereit die
    Erhoehung der DLZ zu tragen
THEN Comet . Periodenwahl(getroffen) =
    woeentlich
```

Typische sprachliche Eigenschaften dieser Wissensbank sind:

- hoher Verbalisierungsgrad bei gleichzeitigem Telegrammstil

- fuer die Generierung eines natuerlichsprachlichen Satzes sind in hohem Masse Interpretationen der Relation, Erzeugung in der Regel selbst nicht enthaltener Verben und rekursive Satzgenerierungen notwendig (Erzeugung von Nebensaetzen).

5 Prinzipien der Generierung und Architektur des Generierungsmechanismus

Im folgenden werden die Prinzipien der Generierung und architektonische Aspekte eines Generierungsmechanismus vorgestellt, die die Generierung natuerlichsprachlicher Erklarungen aus komplexen Experten-Wissensbanken leisten

Als Prinzipien der Generierung betrachten wir:

- a) bezueglich der Grammatik:

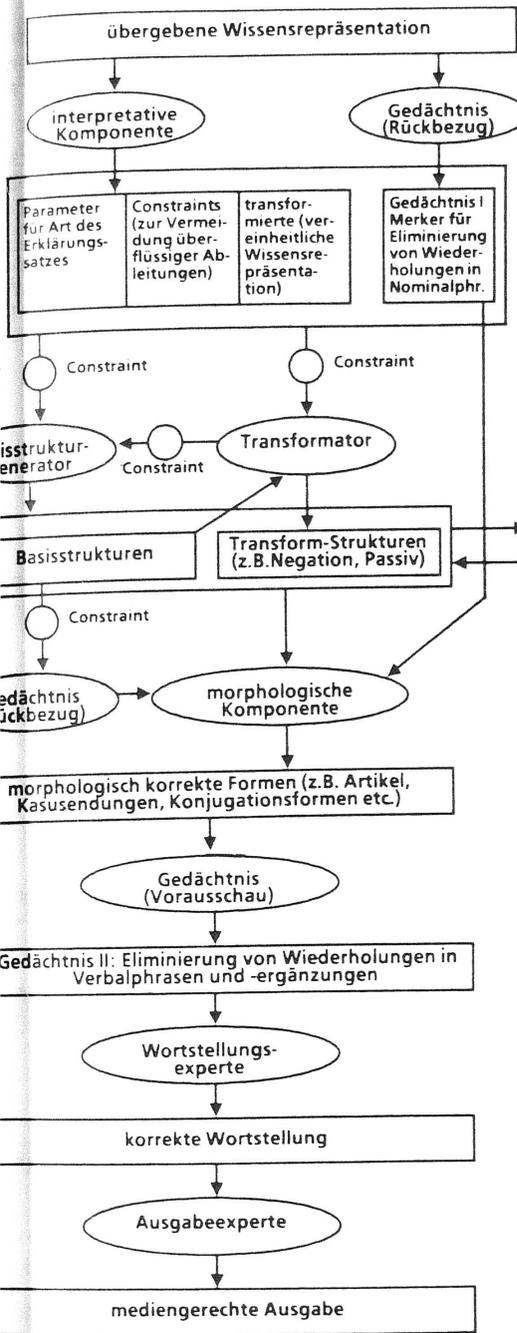
- Erzeugung einer allgemeinen (von der verwendeten Sprache unabhaengigen) syntaktischen Basisstruktur,
- Verwendung eines Minimums an Basisstrukturen,
- Verwendung von Constraints zur Vermeidung unnoetiger Generierungen,
- Generierung eines Maximums an Saetzen aus den Basisstrukturen,
- moeglichst spaete Generierung von Morphologie und Wortstellung (wegen Abhaengigkeit von der verwendeten Sprache).

- b) bezueglich des Lexikons:

- Trennung in Basis- und Fachlexikon
- Bevorzugung eines Grundformenlexikons (Generierung von Morphologie und zusammengesetzten Formen aus Regeln) gegenueber einem Vollformenlexikon,

Themenschwerpunkt

- Aufnahme einer Grundform ins Lexikon nur bei variablem Generierungskontext (sonst "canned text"),
- Modularitaet des Lexikons: (Prinzip der schrittweisen Implementation von Grundformgruppen, bedarfsorientiert, rapid prototyping),
- Modularitaet des Lexikoneintrags: der Lexikoneintrag soll in gestufter Tiefe zugaenglich sein (gestufte semantische Tiefe)
- Minimum lexikalischer Eintraege: das Fachlexikon wird aus der Taxonomie (semi-automatisch) erstellt



Lexikon:
 ● syntakt. relevante Merkmale
 ● Semantik 1. Stufe

c) bezueglich der Erstellung des Lexikons: Verwendung vorhandener Wissensbanken fuer die Lexikonerstellung:

Wir betrachten es als unzumutbar, waehrend der Erstellung eines Expertensystems gleichzeitig linguistisch relevantes Wissen zu erfassen. Die Erstellung von Expertensystemen soll weiterhin unbelastet von der Erfassung linguistisch relevanten Wissens erfolgen. KE und Experte sollen sich bei der Erstellung neuer Wissensbanken voll auf die fachliche Repraesentation ihres Wissens und seine Verarbeitung konzentrieren koennen und nicht mit sprachlichen Details belastet werden. Das eingegebene Wissen muss daher zunaechst sprachlich analysiert werden. Linguistisch notwendiges Zusatzwissen, sofern es nicht bereits aufgrund der Analyse geschlossen werden kann, wird dann im Dialog erfasst (z.B. strikte Subkategorisierung der Nomen)

Im folgenden zunaechst die Architektur, die die natuerlichsprachliche Generierung von Saetzen ermoeeglicht: Die Komponenten im einzelnen:

Abbildung 2: Architektur des Sprachgenerators

Interpretation der Wissensrepraesentation:

a) Wissensrepraesentations-Interpreter: Konzipiert und implementiert wurde ein Mechanismus zur Ueberfuehrung der TWAICE'schen Wissensrepraesentation in eine linguistisch zweckmaessige Wissensrepraesentation (z.B. einheitliche Darstellung von Praemissen und Konklusionen), die im wesentlichen ein 7-Tupel darstellt:

Satztyp, Objekt, Instanz, Attribut, Relation, Wert

Entscheidend ist, dass hierdurch nur ein genereller Interpretationsrahmen angelegt

Themenschwerpunkt

wird: Bezueglich der Komplexitaet der Parameter bestehen prinzipiell keine Grenzvorgaben (natuerlich mit Ausnahme von Speicher), es koennen z.B. einzelne Werte, Listen, Listen von Listen, Strukturen, Strukturen ueber Strukturen etc. sein. In Verbindung mit dem Constraint-Generator wird es hierdurch auf zeiteffektive und elegante Art moeglich, sehr komplexe Einbettungen zu generieren.

Grammatik:

b) Constraint-Generator (zur Vermeidung ueberfluessiger Ableitungen):

Konzipiert und implementiert wurde ein Mechanismus zur Ueberpruefung bestimmter Eigenschaften (z.B. Vorliegen einer Substantivierung oder reservierte Woerter. Hierdurch werden unnoetige Ableitungen blockiert (z.B. 'DUE.Anzahl = 1' -> 'Es gibt eine System-DUE' und nicht 'Die Anzahl der DUEs ist 1'; 'System.Konfiguration = bestimmt' -> 'Das System ist konfiguriert' und nicht 'Die Konfiguration des Systems ist bestimmt'). Die Geschwindigkeit der Generierung wird hierdurch betraechtlich erhoehrt.

c) Mechanismus zur Eliminierung von Wiederholungen:

Konzipiert und implementiert wurde ein zweistufiger Mechanismus zur Eliminierung von Wiederholungen:

Stufe 1: Gedaechnis auf der Grundlage von Rueckblick:

Eliminierung von Wiederholungen innerhalb von Nominalphrasen:

1) Eliminierung von Wiederholungen des (u.U. instanziierten) Objekts durch Pronominalisierung,

2) Eliminierung von Wiederholungen des Attributs durch Tilgung,

3) Insertion von schliessenden Partikeln (damit, auch etc.) .

Geprueft wird, ob die jeweils letzte Wissensuebergabe (im Falle einer Regel: Objekt, Instanz, Attribut) mit der aktuellen identisch sind.

Stufe 2: Gedaechnis auf der Grundlage von Vorausschau

4) Eliminierung von Wiederholungen des Auxiliars oder Verbs durch Tilgung, Insertion von Komma oder 'und' sowie spaetere Insertion der pluralisierten Form,

5) Eliminierung von Wiederholungen des/der Wert(e) oder Teilen von Werten durch Tilgung, Insertion von Komma oder 'und' sowie spaetere Insertion des getilgten Wortes,

6) Eliminierung von Wiederholungen der Konfidenz(en) durch Tilgung, Insertion von Komma oder 'und' sowie spaetere Insertion der getilgten Konfidenz,

7) Wiederholung der Satzkonjunktion durch Tilgung, ggf. Insertion von Komma und spaetere Insertion der getilgten Satzkonjunktion.

Geprueft wird, ob die nachfolgenden Basis- oder Transformstrukturen in Bezug auf Hilfsverb, Verb, Verbergaenzung oder Teil der Verbergaenzung mit der aktuellen identisch sind.

d) Transformator:

Konzipiert und implementiert wurde ein Mechanismus zur Transformation von Basisstrukturen (z.B. negierter Satz, Passivtransformation).

e) Basis-Struktur-Generator:

Konzipiert und implementiert wurde ein Mechanismus zur Generierung von Basisstrukturen (z.B.:

Praemisse,	Display,1,	Bestell-Nr.,	=,
	V	V	
	NP	NP	
V	V	V	
Satztyp	Det-Nom	-Det-Nom	...

f) Morphologische Komponente:

Konzipiert und implementiert wurde eine morphologische Komponente, die bisher folgendes enthaelt:

Regeln zur Generierung der Flexion der Nomen:

- Numerus:
 - singular, plural
- Kasus:
 - Nom., Gen., Dat., Akk.

Regeln zur Generierung der Flexion der Verben:

- Formen:
 - Infinitiv, 3. Pers. Praes., 3.Pers. Imp., 3. Pers. Perf.,

- Verbarten:
 - einfach, zusammengesetzt,

- Reflexivitaet:
 - reflexiv, nicht reflexiv

g) Wortstellungs-Experte:

Konzipiert und implementiert wurde ein Experte fuer die Wortstellung in folgender Satzformen:

- einfacher Aussagesatz
- Praemisse (wenn ...),
 - mit Konjunktion verknuepfte Praemisse
- Konklusion (dann ...),
 - mit Konjunktion verknuepfte Konklusion
- Paraphrase einer Warum-Frage(Warum ...).

Mediengerechte Ausgabe:

- h) Ausgabeexperte:
 Konzipiert und implementiert wurden Ausgabe-
 beregeln fuer die Bildschirmausgabe fuer
 folgende Situationen:
 - normale Textausgabe
 - Ausgabe einer Liste von Fakten
 - Ausgabe einer Regel (allgemeine Form)
 - Ausgabe einer Regel (Spezialformen)
 Druckregeln mit mehr als einer Druckan-
 weisung

Lexikon:

Basislexikon:
 Das Basislexikon enthaelt derzeit Artikel,
 in TWAICE erlaubte Relationen, Verbpartikel,
 Pronomen und Satzverknuepfungswor-
 ter

Fachlexikon:
 Das Fachlexikon enthaelt die in der Taxo-
 nomie enthaltenen (spaeater: die aus der
 Analyse der taxonomischen Begriffe gewon-
 nenen) Begriffe, zuzueglich ihrer lingu-
 istisch relevanten Merkmale.

Nomen:
 Z. Zt. handelt es sich um die linguisti-
 schen Standardmerkmale

- Belebtheit
- Abstraktheit
- Zaehlbarekeit
- Geschlecht
- Numerus
- Pluralbesonderheiten
 - Umlaut
 - Ausnahme

Tatsaechlich benoetigt werden bisher nur
 die letzten 3 Merkmale. Die Lexikoner-
 stellung erfolgt semi-automatisch: die
 taxonomischen Begriffe werden eingele-
 sen, im Dialog werden die linguistisch rele-
 vanten Merkmale erfasst. Dabei wird zu Beginn
 der Dialogerfassung ein Merkmalsdefault
 eingestellt, der allerdings ggf. automa-
 tisch ueberschrieben wird (z.B. Default
 fuer Substantive, die z.B. auf 'e' oder
 'heit' enden: Geschlecht = fem, zur
 Erfassung vgl. Anhang 1).

Abkuerzungen:
 ein Abkuerzungsverzeichnis wird wahlweise
 erstellt
 - waehrend der Lexikonerfassung,
 - unabhaeugig von der Lexikonerfassung.

Verben:
 Die derzeit (im Dialog) erfassten Merkmale
 sind
 - Konjugationsart
 - Reflexivitaet
 - Zusammengesetztheit
 - Transitivitaet

Selektionsbeschraenkungen sind vorgesehen,
 werden zur Zeit jedoch nicht erfasst, da
 sie in unserem Kontext bisher nicht benoe-
 tigt werden.

Ein Beispiel fuer die Erfassung findet
 sich in Anhang 2.

6 Generierung von Erklarungen

Bisher haben wir uns auf die Generierung
 folgender Erklarungen konzentriert:

im Rahmen dynamischen Wissens:
 - aktuelle Regelverknuepfung
 - abgeleitete Fakten

im Rahmen statischen Wissens:
 - Abkuerzungen
 - Beziehungen zwischen Begriffen

6.1 Generierung von Erklarungen zu dynamisch
 Wissen: Aktuelle Regelverknuepfung und ab-
 leitete Fakten

Bisher wurden die Erklarungen der Erklae-
 rungskomponente von TWAICE in Form von
 Lueckentexten und KE-Regeln ausgegeben.
 Diese Form kann zum einen u.U. sehr mono-
 ton, zum anderen - in bezug auf die Regeln
 - sehr technisch klingen und den sonst
 sprachlich fluessigen Dialog stoeren.

Erklaerungstextist derzeit:

- fuer die warum-Frage:
- die Praezisierung der (Oberflaechen-)
 Frage
 - Begruendung der Frage
 - Ausdruck der in Arbeit befindlichen
 Regel:
 - Ausgabe der bereits abgearbeiteten
 Praemissen als Fakten
 - Ausgabe der Restregel

- fuer die wie-Frage:
- die Praezisierung der Faktenfrage
 - Ausgabe des Fakts und Angabe der Her-
 leitung des Fakts

Zusammenfassend lassen sich innerhalb der
 Erklarungen zu

dynamischem Wissen 5 Erklarungssaetze
 unterscheiden:

- Praezisierungen der Fragen
 (Ich verstehe ihre Frage als: ...),
- Begruendungssaetze (Ich erfrage X, weil
 ...),
- Fakten (Fakt X: Y ist Z, Y hat den Wert
 Z, ...),
- Herleitungen von Fakten
 (Fakt X wurde hergeleitet durch ...),
- Regeln (Komplexe Wenn - dann - Saetze).

Im folgenden zeigen wir

Abbildung 3:
 einen Vergleich zwischen bisheriger und
 natuerlichsprachlicher Ausgabe der Erklae-
 rungskomponente der Expertensystem-Shell
 TWAICE

Abbildung 4:
 Beispiele fuer linguistisch komplexe

(4) Wieviele Tastaturen sollen an Arbeitsplatz-1 angeschlossen werden?
>warum
d.h. warum befassen wir uns mit dem Objekt "Tastatur" von Arbeitsplatz-1
Die Untersuchung des Objekts "Tastatur" traegt u.a. zur Bestimmung des Attributs "Peripherie" von "Arbeitsplatz-1" bei.

Folgende Praemissen sind bereits erfuellt:
(FAKT 27) Arbeitsplatz-1 . Typ = normal
daher:

```
IF Tastatur . Konfiguration = bestimmt
AND Display . Konfiguration = bestimmt
AND Drucker . Konfiguration = bestimmt
AND IDKG . Konfiguration = bestimmt
AND AKT . Konfiguration = bestimmt
THEN Arbeitsplatz-1 . Peripherie = bestimmt
( RULE 640 )
```

(4) Wieviele Tastaturen sollen an Arbeitsplatz-1 angeschlossen werden?
>warum
d.h. warum interessieren wir uns fuer das Attribut "Peripherie" von Arbeitsplatz-1?

Die Bestimmung des Attributs "Peripherie" von "Arbeitsplatz-1" laesst Rueckschluesse auf das Attribut "Geraeate" von "Arbeitsplatz-1" zu.

Folgende Praemissen sind bereits erfuellt:
(FAKT 23) Arbeitsplatz-1 . AKT-NUMBER = 1
(FAKT 23) Arbeitsplatz-1 . AKT-NUMBER KNOWN
daher:

```
IF Arbeitsplatz-1 . Peripherie = bestimmt
AND Arbeitsplatz-1 . Platzrechner = bestimmt
THEN Arbeitsplatz-1 . Geraeate = bestimmt
( RULE 460 )
```

(4) Wieviele Tastaturen sollen an Arbeitsplatz-1 angeschlossen werden?
>warum
Ich verstehe Ihre Frage als: Warum befassen wir uns mit der Tastatur des 1. Arbeitsplatzes?

Die Untersuchung der Tastatur ist notwendig, damit die Peripherie des 1. Arbeitsplatzes bestimmt werden kann.

Folgende Regel wird z. Zt. untersucht:

1) schon erfuellte Praemisse(n):
Fakt 27: der Typ des 1. Arbeitsplatzes ist normal

2) Rest der untersuchten Regel:
Wenn die Tastatur ,
der Display ,
der Drucker ,
das IDKG und der AKT konfiguriert sind ,
dann ist auch die Peripherie des 1. Arbeitsplatzes bestimmt .
(Regel 640)

(4) Wieviele Tastaturen sollen an Arbeitsplatz-1 angeschlossen werden?
>warum
Ich verstehe Ihre Frage als: Warum wird nach der Peripherie des 1. Arbeitsplatzes gefragt.

Durch die Bestimmung der Peripherie des 1. Arbeitsplatzes koennen wir den Wert der Geraeate des 1. Arbeitsplatzes erschliessen.

Folgende Regel wird z. Zt. untersucht:

1) schon erfuellte Praemisse(n):
Fakt 23: der Arbeitsplatz hat einen AKT
Fakt 23: er ist damit bekannt

2) Rest der untersuchten Regel:
wenn die Peripherie des 1. Arbeitsplatzes und sein Platzrechner bestimmt sind,
dann sind auch seine Geraeate bestimmt .
(Regel 460)

Abbildung 3: Bisherige und natuerlichsprachliche Ausgabe der TWAICE-Erklarungskomponente im Vergleich

Bisherige Ausgabe

Natuerlichsprachliche Ausgabe

```
RULE 670
IF Arbeitsplatz . Typ = normal
AND Display . Konfiguration = bestimmt
AND Drucker . Konfiguration = bestimmt
AND Arbeitsplatz . IDKG-NUMBER = 0
AND Arbeitsplatz . AKT-NUMBER = 0
THEN Arbeitsplatz . Peripherie = bestimmt
END

RULE 290
IF Arbeitsplatz . Konfiguration EVA KNOWN
THEN DUE . Anzahl 'Ihs-AP' = COUNT ( Arbeitsplatz . Platz * Ihs )
AND DUE . Anzahl 'Ihs-Kanale' = LOOKUP ( tabl . Anzahl 'Ihs-AP' , Anzahl 'Ihs-AP' )
AND DUE . Anzahl 'Ihs-Kanale' = LOOKUP ( tabl . Anzahl 'Ihs-AP' , Anzahl 'Ihs-AP' )
END

RULE 770
IF Arbeitsplatz . AKT-NUMBER = 0
AND Arbeitsplatz . Rechner im Display KNOWN
AND Display . Bestell-SKS = BA-11-11 , BA-11-12 , BA-20-11 , BA-20-12
THEN Arbeitsplatz . Platzrechner = bestimmt
END

RULE 3980 FORWARD
IF Arbeitsplatz . Ausgabe KNOWN
AND Arbeitsplatz . Druckform = II
THEN PRINT ( Arbeitsplatz . TITLE=SY )
PRINT ( Arbeitsplatz . Art , FORM=SY3 )
PRINT ( Arbeitsplatz . Installation , FORM=SY )
PRINT ( Arbeitsplatz . Anschluss , FORM=SY )
PRINT ( Lobby-GAA . Bestell-SKS , FORM=SY2 )
PRINT ( Outdoor-GAA . Bestell-SKS , FORM=SY2 )
END
```

```
Regel 670
Wenn der Typ des Arbeitsplatzes normal ,
der Display ,
der Drucker konfiguriert sind ,
der Arbeitsplatz kein IDKG und keinen AKT hat ,
dann ist seine Peripherie bestimmt .
Regelende

Regel 290
Wenn die Konfiguration des Arbeitsplatzes bereits bekannt oder ableitbar ist ,
dann
sind die Anzahl 'Ihs-AP' der DUE der Anzahl der Arbeitsplaetze gleich
unter der Bedingung , dass der Platz den Wert 'Ihs' hat ,
die Anzahl 'Ihs-Kanale' der DUE und ihre Anzahl 'Ihs-Kanale' werden
in der Tabelle 'tabl' nachgeschlagen .
Regelende

Regel 770
Wenn der Arbeitsplatz keinen AKT hat ,
sein Rechner im Display bekannt und die Bestell-SKS des Displays
BA-11-11 , BA-11-12 , BA-20-11 oder BA-20-12 ist ,
dann ist der Platzrechner des Arbeitsplatzes bestimmt .
Regelende

Regel 3980 (Vorwaertsregel)
Wenn die Ausgabe des Arbeitsplatzes bekannt und seine Druckform II ist ,
dann drucke:
den Titel SY : fuer den Arbeitsplatz ,
in der Form SY3 : seine Art ,
in der Form SY : seine Installation ,
seinen Anschluss ,
in der Form SY2 : die Bestell-SKS des Lobby-GAAs
und die des Outdoor-GAAs .
Regelende
```

Abbildung 4: Beispiele fuer linguistisch komplexe Regelgenerierungen

Regelgenerierungen

6.2 Generierung von Erklärungen zu statischem Wissen

6.2.1 Abkuerzungen

Abkuerzungen werden aus dem Abkuerzungslexikon generiert. Generiert wird - je nach Anforderungsparameter - ein Text der Form 'X ist eine Abkuerzung fuer Y' bzw. nur die Langform Y.

6.2.2 Beziehungen zwischen Begriffen

Es wird das Wissen ausgewertet, das in den unterschiedlichen Wissensquellen (Taxonomie-, Fragen-, Rat-Dateien) bereits vorliegt. Generiert werden Erklärungen nach dem Muster:

Bei dem Begriff handelt es sich um ein Objekt, ein Attribut oder einen Wert oder einen Begriff in verschiedener Bedeutung (z.B. Objekt, Attribut verschiedener Objekte, Wert verschiedener Attribute).

Falls der Begriff ein Objekt ist, wird angegeben, welche Objekte hierarchisch ueber und unter diesem Objekt stehen, welche Attribute es hat und wieviele Objekte dieser Art es geben kann.

Falls der Begriff ein Attribut ist, wird angegeben, zu welchem/welchen Objekt/en es gehoert, welche Werte es annehmen kann, ob es normalerweise erfragt oder abgeleitet wird, welche Frage ggf. gestellt wird und welcher Rat ggf. gegeben wird.

Ist der Begriff Attribut von nur einem Objekt, wird dieses Objekt angegeben und angezeigt, welche Werte das Attribut annehmen kann, welche Frage ggf. zu ihm gestellt und welcher Rat ggf. zu ihm gegeben werden.

Ist der Begriff Attribut von mehreren Objekten, so wird davon ausgegangen, dass sich die Erklarung auf das Attribut des zuletzt erklarten Objekt-Begriffs beziehen soll. Es wird eine entsprechende Erklarung ausgegeben sowie ein Hinweis darauf, wie weitere Erklarung zu den weiteren Objekt-Attributen anzufordern ist. Gibt es keinen zuletzt erklarten Objekt-Begriff, so wird nur ausgegeben, zu welchen Objekten das Attribut gehoert und ein Hinweis darauf, wie weitere Erklarung anzufordern ist.

Falls der Begriff ein Wert ist, wird angegeben zu welchem/welchen Attribut/en von welchem/welchen Objekt/en er gehoert und welche weiteren Werte es ausserdem gibt. Ist der Begriff Wert von nur einem Attribut, wird das entsprechende Objekt-Attribut-Paar angezeigt sowie die Werte, die es ausserdem annehmen kann. Taucht der Begriff mehrfach als Wert auf, so wird verfahren wie bereits bei den Attributen.

6.2.3 Partnermodellerte Ausgabe

Da wir fuer den Bereich Erklärungen zu Beziehungen zwischen Begriffen bereits unterschiedliche Benutzeranforderungen (auf der Grundlage unserer Erfahrung) formulieren koennen, haben wir verschiedene Erklarungsausgaben vorgesehen.

Die Erklärungen koennen in unterschiedlicher Form angefordert werden:

- Kurzform
- Langform
- Knowledge-Engineer-Form.

Diese Formen unterscheiden sich bezueglich

- der Wortwahl
- der Bildlichkeit
- der Informationsmenge
- der Praezision.

Sie erfuellen folgende Anforderungen:

- Die Kurzform ist eine Untermenge der Langform. Sie informiert kurz und in allgemeinverstaendlicher Form.
- Die Langform enthaelt zusaetzliche allgemeinverstaendliche Erlaeuterungen.
- Die KE-Form informiert kurz und praezise ohne Veranschaulichungen. Sie bietet die Option der vollstaendigen (aber knappen) Information.

Da Erklärungen dieser Art nach einem bestimmten Muster ablaufen, wird eine komplette Satzgenerierung hier fuer unnoetig angesehen. Sie ist machbar, benoetigt aber mehr Zeit als das von uns fuer Erklärungen dieser Art voellig hinreichend angesehene entwickelte Verfahren der Lueckentexte mit morphologischer Komponente. Dabei wird ein Lueckentext erstellt, wobei fuer die Variablen die jeweiligen gewuenschten grammatischen Eigenschaften angegeben werden, so dass ein sprachlich fluessiger, voll flektierter Text entsteht.

Abbildung 5 zeigt die Erklarung der Begriffe "Tastatur" und "Doppelnull" in Bezug auf andere Begriffe der Wissensbank fuer

- a) einen Normalbenutzer in Langform
- b) fuer einen KE

2) Tastatur ist eine Eigenschaft der Einheit Auszugdrucker.
 Auszugdrucker haben die Eigenschaften:
 - Tastatur
 - Anzeige
 - Bauart
 - Bestell-SKS
 - u.ä.m..

Eigenschaften haben Werte, die erfragt oder erschlossen werden.
 Zulässige Werte fuer Tastatur sind:
 - "Ja"
 - "nein".

Tastatur als Eigenschaft von Auszugdruckern wird normalerweise erfragt:
 Die Frage an den Benutzer lautet:
 "Soll der Auszugdrucker eine Tastatur haben?"
 Weiss der Benutzer nicht, was damit gemeint ist,
 so erhaelt er folgenden Rat:
 "Ueber die Tastatur koennen von Bankkunden zusaetzliche Eingaben, wie
 z.B. PIN, verlangt werden. Halten Sie im Zweifelsfaelle Ruecksprache mit
 dem Kunden."

=> Doppelnull

Doppelnull ist eine Eigenschaft der Einheit Tastatur.
 Tastaturen haben die Eigenschaften - Doppelnull
 - Konfiguration-ALLE
 - Bestell-SKS
 - Funktionstasten
 - u.ä.m..

Eigenschaften haben Werte, die erfragt oder erschlossen werden.
 Zulässige Werte fuer Doppelnull sind:
 - "Ja"
 - "nein".

Doppelnull als Eigenschaft von Tastaturen wird normalerweise erfragt:
 Die Frage an den Benutzer lautet:
 "Soll diese Tastatur eine Doppelnull-Taste erhalten?"
 Weiss der Benutzer nicht, was damit gemeint ist,
 so erhaelt er folgenden Rat:
 "Manche Banken wuenschen eine Doppelnull-Taste zur schnelleren Eingabe
 runder Betraege. Erfragen Sie im Zweifelsfaelle die Antwort beim Kunden."

2) Tastatur ist ein Attribut des Objekts Auszugdrucker.
 Auszugdrucker haben die Attribute Tastatur, Anzeige, Bauart, Bestell-SKS,
 Journal, u.ä.m..

Moechten Sie alle Attribute sehen? (nein)

>

Zulässige Werte fuer Tastatur sind:
 - "Ja"
 - "nein".

Tastatur als Attribut von Auszugdruckern wird normalerweise erfragt:
 Die Frage an den Benutzer lautet:
 "Soll der Auszugdrucker eine Tastatur haben?"
 Der Rat fuer den Benutzer lautet:
 "Ueber die Tastatur koennen vom Bankkunden zusaetzliche Eingaben, wie
 z.B. PIN, verlangt werden. Halten Sie im Zweifelsfaelle Ruecksprache mit
 dem Kunden."

=> Doppelnull

Doppelnull ist ein Attribut des Objekts Tastatur.
 Tastaturen haben die Attribute Doppelnull, Konfiguration-ALLE, Bestell-SKS,
 Funktionstasten, Konfiguration, u.ä.m..

Moechten Sie alle Attribute sehen? (nein)

>

Doppelnull als Attribut von Tastaturen wird normalerweise erfragt:
 Die Frage an den Benutzer lautet:
 "Soll diese Tastatur eine Doppelnull-Taste erhalten?"
 Der Rat fuer den Benutzer lautet:
 "Manche Banken wuenschen eine Doppelnull-Taste zur schnelleren Eingabe
 runder Betraege. Erfragen Sie im Zweifelsfaelle die Antwort beim Kunden."

Abbildung 5: Erklarungen zu Beziehungen zwischen Begriffen

Abkuerzungsverzeichnis

- ATN Augmented Transition Network
- Adv Adjectiv
- Aux Auxiliary
- DCG Definite Clause Grammar
- Det Determiner
- EST Extended Standard Theory (= Erweiterung derTG)
- FUG Functional Unification Grammar
- GPSS Generalized Phrase Structure Grammar
- KE Knowledge Engineer
- LFJ Lexical Functional Grammar
- Nom Nomen
- NF Nominalphrase
- TG Transformationalgrammatik
- VF Verbphrase

Literaturverzeichnis

[1] ANDERSON, John R./REISER, Brian J.: The LISP Tutor. In: BYTE April 1981, S. 159 - 176.

[2] BERTHO, Robert C./WEINBERG, Amy D.: The Grammatical Basis of Linguistic Performance. Language Use and Acquisition. Cambridge (Mass.): MIT Press 1984.

[3] SCHNEP, Herbert G./BACHER, Paul L.: Automatic English to Logic Translator. In: A Simplified Model. A Study in the Logic of Language. In: STELLWALTER, A./SALIK, G. A./SEIDLOW, James (Hrsg.): Language Research. Formal Methods. Trends in Linguistic Studies and Monographs, Nr. 8, Mouton Publ., 1979, S. 165 - 223.

[4] BRADY, Michael/BERTHO, Robert C./Hrsg.: Computational Models of Discourse. Cambridge (Mass.): MIT Press 1984.

[5] BRENNAN, Joan W.: A Realistic Transformational Grammar. In: HALL, M./BRENNAN, Joan W./MILLER, George A. (Hrsg.): Linguistic Theory and Psychologic Reality. Cambridge(Mass.): MIT Press 1978, S. 1 - 59.

[6] BUSEMANN, Steffen: Topicalization and Pronominalization. Extending a Natural Language Generation System. In: O'Shea, Tim (Hrsg.): ECAI 1984 - Proceedings of the Sixth European Conference on Artificial Intelligence. Pisa September 5-7, 1984, S. 221 - 229.

[7] BUSEMANN, Steffen: Surface Transformations During the Generation of Written German Sentences. Forschungsstelle fuer Informatik und Kuenstliche Intelligenz der Universitaet Hamburg. Report ANS-27, 1984.

[8] CHOMSKY, Noam: Aspects of the Theory of Syntax. Disser: Aspekte der Syntax. Theoret. Linguistik/M.: Suhrkamp 1965.

[9] CHOMSKY, Noam: Questions of Form and Interpretation. In: Linguistic Analysis, 1, 1975, S. 75 - 109.

[10] CHOMSKY, Noam: Conditions on Rules of Grammar. In: CHOMSKY, Noam: Essays on Form and Interpretation. New York: Amerik. Elsevier 1977, S. 163 - 210.

[11] CLANCEY, William J.: The Epistemology of a Rule-Based Expert System - a Framework for Explanation. In: Artificial Intelligence 20, 1983, H.3, S. 215 - 251.

[12] DANLOS, Laurence: Some Issues in Generation from a Semantic Representation. In: BUNDY, Alan(Hrsg.): ICAI - 83. Proceedings of the Eighth International Joint Conference on Artificial Intelligence, 8-12. August 1983, Karlsruhe, 2 Bde., Band 1, 606 - 609.

[13] DANLOS, Laurence: An Algorithm for Automatic Generation. O'Shea, Tim (Hrsg.): ECAI 1984 - Proceedings of the Sixth European Conference on Artificial Intelligence. Pisa September 5 1984, S. 213 - 216.

[14] DEPP, Marcia A./MCKEOWN, Kathleen P.: Using Focus to Generate Complex and Simple Sentences. In: Proceedings of Coling84, July 1984, Stanford University, California, S. 319 - 326.

[15] EISELE, A.: A Lexical Functional Grammar System in Prolog. LDV-Forum, 3.Jg, 1985, Nr. 2, S. 67 - 76.

[16] FILLMORE, Charles J.: Toward a Modern Theory of Case. In: BELL, D./SCHANE, S. (Hrsg.): Modern Studies in English. Eng. wood Cliffs (N.J), S. 361 - 376.

[17] FORGE, John: Physical Explanation: With Reference to the Principles of Scientific Explanation of Hempel and Salmon. McLAUGHLIN, R.: What? Where? When why?, Dordrecht(Holl.): Reidel Publ. Comp. 1982, pp. 211 - 230.

[18] FRIEDMAN, Joyce: Expressing Logical Formulas in Natural Language. In: GROENENDIJK, J. A. G./JANSSEN, T. M. V./STOR- M. B. J. (Hrsg.): Formal Methods in the Study of Language. Amsterdam: Mathematisch Centrum, 2 Baende, 1981, Band 1, S. 1 - 130.

[19] GAZDAR, Gerald: Phrase structure grammar. In: JACOBSON, P./PULLUM, G. K.: (Hrsg.): The Nature of Syntactic Representation. Dordrecht: D. Reidel 1982.

[20] GAZDAR, Gerald/KLEIN, Ewan/PULLUM, Geoffrey/SAG, Ivan: Generalized Phrase Structure Grammar, Oxford: Basil Blackwell, 1985.

[21] GILMOUR, Clark: Causal Inference and Causal explanation. McLAUGHLIN, R.: What? Where? When why?, Dordrecht(Holl.): Reidel Publ. Comp. 1982, pp. 179 - 192.

[22] FININ Timothy, W.: Providing Help and Advice in Task Oriented Systems. In: Proceedings of the Eighth International Joint Conference on Artificial Intelligence, 8-12 August 1983, Karlsruhe West Germany, Band 1, S. 176 - 178.

- [23] GOGUEN, J. A./WEINER, J. L./LINDE, C.: Reasoning and Natural Explanation. In: *International Journal of Man-Machine Studies*, 19, 1983, S. 521 - 559.
- [24] HAND, D. J.: Statistical Expert Systems: Necessary Attributes. In: *Journal of Applied Statistics*, 12, 1985, H. 1, S. 19 - 27.
- [25] HASLING, Diane Warner/CLANCEY, William J./RENNELS, Glenn: Strategic Explanations for a Diagnostic Consultation System. In: *Int. J. Man-Machine Studies*, 20, 1984, S. 3 - 19.
- [26] HAYS, D. G.: Dependency Theory. In: *Language*, 1964, 4.
- [27] HEMPEL, C. G.: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press 1965.
- [28] HINDELANG, Goetz: Zur Klassifikation der Fragehandlungen. In: HINDELANG, Goetz/ZILLING, Werner (Hrsg.): *Sprache: Verstehen und Handeln. Akten des 15. Linguistischen Kolloquiums*. Muenster 1980. Band 2, Tuebingen: Niemeyer 1981, S. 215 - 225.
- [29] JACKSON, Peter/LEFRERE, Paul: On the Application of Rule-Based Techniques to the Design of Advice-Giving Systems. In: *Int. J. Man-Machine Studies*, 20, 1984, S. 63 - 86.
- [30] JACOBS, Paul S.: Generation in a Natural Language Interface. In: BUNDY, Alan (Hrsg.): *ICAI - 83. Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. 8.-12. August 1983, Karlsruhe, 2 Bde., Band 1, S. 610 - 612.
- [31] KAPLAN, R./BRESNAN, Joan: *Lexical Functional Grammar: a Formal System for Grammatical Representation*. In: BRESNAN, Joan (Hrsg.): *The Mental Representation of Grammatical Relations*. Cambridge (Mass.): MIT Press 1982.
- [32] KATZ, Boris: A Three-Step Procedure for Language Generation. *Mass. Inst. of Technology* 1980, A.I. Memo No. 599.
- [33] KAY, Martin: Functional Unification Grammar: A Formalism for Machine Translation. In: *Proceedings of the 12th International Conference of Computational Linguistics, COLING 1984*, Stanford, S. 75 - 78.
- [34] LANG, Ewald: *Erklärungstexte*. In: DANES, Frantisek/VEHWEGER, Dieter (Hrsg.): *Probleme der Textgrammatik. Studia Grammatica IX*, Berlin: Akademie-Verlag 1976, S. 147 - 181.
- [35] LAUBSCH, J./ROESNER, D./HANAKATA, K./LESNIEWSKI, A.: Language Generation from Conceptual Structure: Synthesis of German in a Japanese/German MT Project. In: *Proceedings of Coling84*, 2-6 July 1984, Stanford University, California, S. 491 - 494.
- [36] LEHNERT, Wendy G.: *The Process of Question Answering*. Hillsdale, New Jersey: Lawrence Erlbaum Associates 1978.
- [37] MANN, William/BATES, Madeline/GROSZ, Barbara/McDONALD, David D./McKEOWN, Kathleen, R./SWARTOUT, William: Text Generation. In: *American Journal of Computational Linguistics*, Vol. 8, No 2, Apr.- June 1982, pp. 62 - 69.
- [38] McDONALD, David D.: Natural Language Generation as a Computational Problem: An Introduction. In: BRADY, M./BERWICK, R. C. (Hrsg.): *Computational Models of Discourse*. Cambridge (Mass.): MIT Press 1984.
- [39] McKEOWN, Kathleen R.: *Generating Natural Language Text in Response to Questions about Database Structure*. Ph. D. Thesis. University of Pennsylvania 1982.
- [40] McKEOWN, Kathleen, R.: Focus Constraints on Language Generation. In: BUNDY, Alan (Hrsg.): *ICAI - 83. Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. 8.-12. August 1983, Karlsruhe, 2 Bde., Band 1, S. 582 - 587.
- [41] McKEOWN, Kathleen, R.: Discourse Strategies for Generating Natural-Language Text. In: *Artificial Intelligence*, 27, 1985, S. 11 - 41.
- [42] MONTAGUE, Richard: *Universal Grammar*. In: THOMASON, Richmond H. (Hrsg.): *Formal Philosophy: Selected Papers of Richard Montague*. New Haven: Yale University Press, S. 222 - 246.
- [43] MORIK, Katharina: Eilige Thesen zur Zukunft natürlichsprachlicher Systeme. In: *LDV-Forum*, 3. Jg, 1985, Nr. 2, S. 88 f.
- [44] PEREIRA, Fernando/WARREN, David: Definite Clause Grammars for Language Analysis - a Survey of the Formalism and a Comparison With Augmented Transition Networks. In: *Artificial Intelligence*, 13, 1980, S. 231 - 278.
- [45] PEREIRA, Fernando/WARREN, David: An Efficient Easily Adaptable System for Interpreting Natural Language Queries. In: *American Journal of Computational Linguistics*, 8, 1982, S. 110 - 122.
- [46] POLYA, G.: *Vom Lösen mathematischer Aufgaben: Einsicht und Entdeckung. Lernen und Lehren*. Basel: Birkhäuser, 2 Bde. 1979.
- [47] REYLE, Uwe/FREY, Werner: A PROLOG Implementation of Lexical Functional Grammar. In: BUNDY, Alan (Hrsg.): *ICAI - 83. Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. 8.-12. August 1983, Karlsruhe, 2 Bde., Band 2, S. 693 - 695.
- [48] SAGER, Naomi/GRISHMAN, Ralph: The Restriction Language for Computer Grammars of Natural Language. In: *Communications of the ACM*, Vol. 18, No. 7, July 1975, pp. 390 - 400.
- [49] SCRIVEN, Michael D.: Definitions, Explanations and Theories. In: FEIGL, H./SCRIVEN, M./MAXWELL, G. (Hrsg.): *Minnesota Studies in the Philosophy of Science*, Vol. II, Minneapolis: Univ. of Minnesota Press 1958, pp. 99 - 195.
- [50] SWARTOUT, William P.: XPLAIN: A System for Creating And Explaining Expert Consulting Programs. In: *Artificial Intelligence* 21, 1983, H.3, S. 285 - 326.
- [51] SWARTOUT, William P.: Explaining And Justifying Expert Consulting Programs. In: CLANCEY, William J./SHORTLIFFE Edward H.: *Readings in Medical Artificial Intelligence*. Reading, Mass.: Addison-Wesley 1984, S. 382 - 398.
- [52] TESNIERES, L.: *Elements de syntaxe structurale*. Paris 1959.
- [53] WAHLSTER, Wolfgang: *Natürlichsprachliche Argumentation in Dialogsystemen*. Berlin et al.: Springer 1981.
- [54] WEINER, J. L.: BLAH, a System Which Explains Its Reasoning. In: *Artificial Intelligence*, 15, 1980, S. 19 - 48.
- [55] WILENSKY, Robert/ARENS, Yigal/CHIN, David: Talking to UNIX in English: an Overview of UC. In: *Communications of the ACM*, 27, 1984, H. 6, S. 574 - 593.
- [56] WOODS, W. A.: Transition Network Grammars for Natural Language Analysis. In: *Communications of the ACM*, 13, 1970, S. 561 - 606.
- [57] WOODS, W. A.: Cascaded ATN Grammars. In: *American Journal of Computational Linguistics*, 6, 1980, 1, S. 1 - 12.
- [58] YASUKAWA, Hideki: LFG System in Prolog. In: *Proceedings of Coling84*, 2-6 July 1984, Stanford University, California, S. 358 - 361.

Vier Seiten Anhang zur semi-automatischen Erfassung von Nomen (Anhang 1) und von Verben (Anhang 2) können bei der Redaktion angefordert werden.

UNIFICATION-BASED SPEECH PARSING WITH A CHART

Günther Görz, Clemens Beckstein

Universität Erlangen-Nürnberg, RRZE and IMMD VI

Martensstr. 1

D-8520 Erlangen

We describe GuLP, a chart parser which can be used as a syntactic module for the Erlangen Speech Understanding System EVAR. GuLP realizes an agenda-based multiprocessing scheme, which easily allows the application of various parsing strategies to fragments of the same utterance in a transparent way. The overall control mechanism is realized through a general interrupt system. In order to process speech data, a variety of new features has been incorporated: in particular the ability to perform incremental analysis, to the direction independent island parsing, to process gaps in utterances and to handle hypothesis scores. Finally, complexity estimate and a few experimental results are discussed.

1. Application Context

The General unification-based Linguistic Processor (GuLP) has been developed in cooperation with a German speech understanding project, EVAR (Niemann 1984). In principle, EVAR aims at the goals of the ARPA SUR, which are still challenging, but uses a larger dictionary. Furthermore, a lot of augmentations is required for any serious application, in particular speaker independence, size of vocabulary, linguistic coverage, and real time performance. The long-range goal of project EVAR is to realize capabilities for conducting dialogs about travel planning within the Intercity train system.

With the exception of the SRI system, the ARPA project's speech systems were not satisfactory from a linguistic point of view: HARPY used a precompiled network containing all admissible utterances explicitly HEARSAY-II (Erman 1980) employed a grammar of syntactic/semantic templates and HWIM (Woods 1976) a semantic/pragmatic ATN grammar, all of which are very difficult to extend and have to be rewritten completely for any new domain of discourse. The HARPY system, which showed the best performance, is the least flexible from a linguistic and cognitive perspective.

2. Theoretical Background

From the theoretical and methodological point of view our approach can be characterized as "pragmatics first", i.e. we see the task of natural language understanding from the viewpoint of communication as social action and interaction. If the recovery of the structure of an utterance is to be seen as part of a larger process of analyzing meaning, intentions, and goals underlying its generation, the grammar should be a *functional* one (Dik 1979, Kay 1980).

Of course, functional grammar makes use of structural analysis, but emphasizes context and situation. Our special interest is to identify the requirements for a processing model and a grammar formalism which follow from a pragmatic perspective. One insight which is common to most of the recent approaches is that the grammar formalism should be descriptive. Procedural or strategic information has to be kept separately from the grammar, but explicitly, and not as an inherent property of the processing model. In functional grammar, there is a twofold use of the concept of 'function':

- *Grammatical functions* describe role assignment to components of the utterance providing a link between situation and utterance.
- The *mathematical concept of function* is present in functional structures (f-structures) which represent grammatical functions. F-structures are functions with finite domain and range, expressed in tabular form as sets of attribute-value pairs, where values are symbols or f-structures, respectively. New f-structures are constructed by unification, which is a monotonous operation.

The theoretical foundations for our processing model are provided by Lexical Functional Grammar (LFG; Bresnan, Kaplan 1982), a salient feature of which is the principle of additive description. Besides constituent structures, LFG generates functional structures by the unification of features. It has in common with our processing model the stress on the role of the lexicon. GuLP's grammar formalism is rule-oriented, but differs from that of LFG and also from the more general PATR system (Shieber 1984) by procedural elements with an emphasis on parallelism, providing operators for process control, structure access, and unification-based generation. From the viewpoint of formal languages, the expressiveness of grammar rules is similar to that of indexed grammars.

3. The Architecture of GuLP

Because of uncertain data, in systems for processing continuous speech progress in one level of analysis can only be achieved by confirmation from other levels. Therefore a common data structure for all levels of processing and a unique control framework are essential.

The central data structure of GuLP is the *active chart* (Kay 1980; Thompson 1981, 1984). Its basic design extends the Well Formed Substring Table, a device used in many parsers to store intermediary results. The chart is represented as a directed graph, the edges of which are labelled with partial analysis results and information how to continue.

Initially, the chart is set up - in the ideal case of text analysis - as a set of vertices which mark the beginning and end of an utterance and the boundaries between words. The vertices are connected by (*inactive*) edges which are labelled by the words themselves and lexical information. Whenever a constituent is found during the parsing process, a new inactive edge is added to the chart. In contrast to that, *active edges* represent incomplete constituents; they indicate a stage in the search for a constituent. Active edges record the category of the constituent they are looking for, its structure so far known and an indication which kind of information would be necessary to complete it. In the simplest case (top-down), the analysis process is started by inserting an active edge emerging from the first vertex and pointing to it - a loop -, which is labelled by the category it is looking for: S(entence). The whole process is governed by the fundamental principle of chart parsing (quoted from Thompson 1981): "Whenever an active edge *A* and an inactive edge *I* meet for the first time, if *I* satisfies *A*'s conditions for extension, then build a new edge as follows:

- Its left end is the left end of *A*,
- Its right end is the right end of *I*,
- Its category is the category of *A*,
- Its contents are a function (dependent on the grammatical formalism employed) of the contents of *A* and the category and contents of *I*;
- It is inactive or active depending on whether this extension completes *A* or not."

Neither *A* nor *I* is modified, instead a new independent edge is introduced into the chart. Therefore it is guaranteed that in the case of ambiguity all parses will be found, independent of the order in which the individual steps are executed. So the chart is made into an active parsing agent, in particular by simulating a multiprocessing scheme which is realized by the use of agendas.

An *agenda* is a list of tasks, which are atomic, non-interruptible processing steps (i.e. transactions) to be carried out over the chart. Whenever edges are added to the chart, any new tasks which can be created as a result, are scheduled on an agenda. Scheduling of tasks on an agenda is done by its associated *scheduler* which assigns priorities to tasks. The selection of tasks from an agenda is performed by its *selector*, which can, in the extreme cases, either perform a depth-first (stack) or a breadth-first (queue) search. The rule invocation (or parsing) strategy - bottom-up or top-down - is independent of the choice of the search strategy. Different parsing strategies are reflected in different conditions for the introduction of *empty* active edges. An empty edge represents the task to search for a constituent; it points to the same vertex where it is emerging from, indicating the search direction.

A sample of word hypotheses for the sentence "Ich habe schöne Platzkarten von München nach Hamburg für Sie gefunden" and their underlying phoneme hypotheses, as they are generated by EVAR's word hypothesizing module, is shown in Fig.1. A subset of these word hypotheses, according to a certain recognition quality threshold, is incorporated into GuLP's chart.

To illustrate the working of our processing model, Fig.2 shows a simple chart for the input sentence "Der Zug fährt nach Frankfurt". Due to the excellent graphic capabilities of Kaplan's LFG system which are currently not available for GuLP, this picture was produced by the LFG implementation on a Xerox 1108 workstation.¹ Whereas the LFG system does only support text strings as input, GuLP is able to introduce multiple overlapping word hypotheses into its chart and to process them concurrently. So the chart of Fig.1 has to be understood as a subgraph of a more complex chart which is constructed by GuLP, but the underlying mechanisms are the same. Of course, the grammar and lexicon which were used for this application of the LFG system were the same as those employed by GuLP.

¹The authors are very grateful to CSLI, Stanford University, for offering its excellent facilities during a visit in summer 1985.

BOTTOM UP PARTIAL NOT FULL

Chart Window

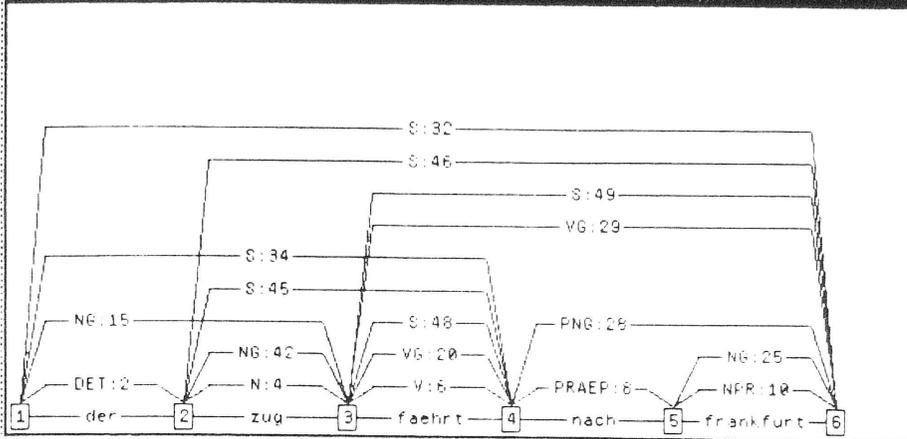


Fig.2: A chart

In this ideal case, a S-edge spanning the whole utterance is found, which leads to the construction of the constituent structure in Fig.3.

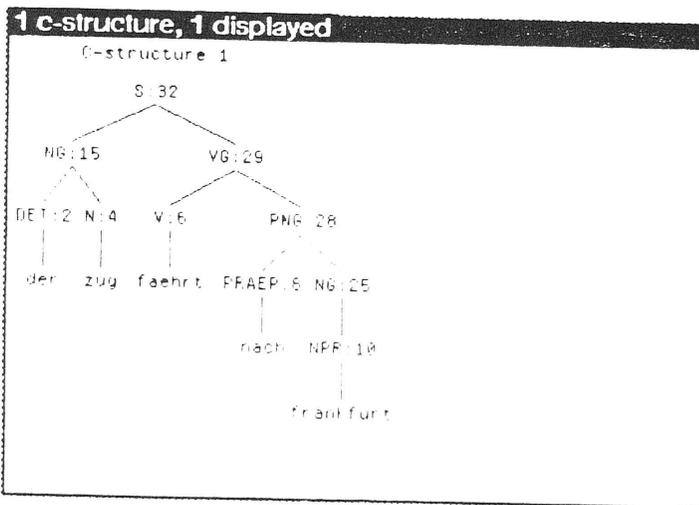


Fig.3: A constituent structure

During the construction of a constituent structure, a set of equations, the "functional description", which reflects linguistic constraints, e.g. agreement, is established and solved by *unification*. This process results in the construction of a functional structure as given in Fig.4. The main difference between GuLP and the LFG system at this point lies in the fact that GuLP solves the equations incrementally, i.e. as soon as a constituent is found, whereas the LFG system postpones the solution until a syntactic interpretation for the whole sentence is detected. Although this strategy is not essential for LFG, it is not suitable for the processing of speech data, where according to the present state of the art a lot of constituents can be found, but seldom a coherent interpretation for a whole utterance. Another difference between the two systems is that GuLP does not construct constituent structures as separate entities as LFG does, but instead builds up functional structures which contain the complete phrase structure information. Again, this property is not essential for LFG, and both systems could easily be modified to perform in the one or other way. According to Kaplan (personal communication, 1985) it is an open question whether future versions of the LFG system will still keep constituent structures as separate entities, because they are not fundamental, but merely illustrate an intermediate processing step. Due to the much more complex input data, which contain overlapping and competing word hypotheses, GuLP generates not only a few, but a lot of functional

world is deactivated and all available resources are concentrated on analysis tasks. Because there are two kinds of analysis tasks, basically one for each parsing direction, blocking one of them will enforce to follow the remaining direction.

A necessary and useful device in a speech parser is its ability to perform direction-independent *island parsing*. In order to process defective input data, our parser must be able to start its operation at any point within the chart. Expanding islands to the right is just what genuine analysis tasks ("*syntax tasks*") do anyway. Proceeding in the main parsing direction from left to right, all information required for expansion is already attached to their corresponding active edges. *Scan tasks*, a second new kind of analysis tasks, are seeking left of the island to find anchor points for syntax tasks by accessing information attached to their starting vertex. They are scheduled automatically if an island cannot be expanded to the right.

GuLP does not schedule an appropriate syntax task before an anchor point has been identified. Since constituent hypotheses are only generated by syntax tasks, fewer hypotheses than in traditional island parsers are generated. This fact results from a more adequate representation of an island's left context: expansions to the left are tried under consideration of its real left context and not just as long as only predictions exist.

In a speech understanding system, hypotheses in all levels of abstraction carry *quality scores*, which the parser must be able to handle with respect to words and constituents. GuLP reads word hypotheses from EVAR's blackboard, inserts appropriate word edges into its chart, and attaches priority scores derived from their quality scores to them as features. During parsing, scores are computed and assigned to constituents such that the discovery of the best matching interpretation of the utterance is prioritized.

If *gaps* in the utterance are recognized (i.e. there are no good word hypotheses in a certain time interval), edges are marked with the universal 'gap'-category are introduced and scored with a threshold value. Then an input task is scheduled automatically which looks whether matching word hypotheses have become available in the meantime. With each attempt to find a matching word hypothesis at that place, the gap edges' score is reduced by a certain percentage until it falls below a second threshold. In this case of failure an incomplete constituent hypothesis is generated from the available information, including the pattern which characterizes the missing word(s).

In addition, while building constituent hypotheses, GuLP could also take into consideration preference scores (or weights) for different branches in the grammar, but our present grammars do not yet employ this feature.

5. Some Reflections on Complexity

Some complexity estimates for chart parsing are well known by investigations on the context-free case (Sheil 1976). Regarding only the context-free skeleton of the grammar, where dynamic rule modification is excluded, an upper bound for the growth of the chart can be given by $O(n^{c+1})$ edges. The initial size of the chart is denoted by n , and c refers to the least number greater or equal than the number of non-terminal symbols in the right hand side of any rule, where all right hand sides containing alternatives have to be considered as being expanded into separate rules. Because the grammar rules are exhaustive, it can be shown that this bound is also valid for the maximum number of analysis steps. With the assumption that each step - including unification as one step - takes unit time, the time requirements are of the same order. This result holds independent of the parsing strategy. The size of the grammar does not influence complexity, but its branching factor, which is a measure for its degree of nondeterminism, acts as a proportionality factor.

6. Results of Experiments

GuLP has been implemented completely in InterLISP and consists of more than 160 functions in 14 modules. To carry out simulation experiments, a special user interface with a sophisticated trace system has been realized. The linguistic database consists of EVAR's lexicon which presently contains more than 1500 words, and two experimental grammars. The first grammar has been developed from a purely syntactic German ATN, the second one is a functional grammar for German based on valency theory, closely resembling to the LFG approach.

Besides particular test data, which were constructed to check various features of the parser, about a dozen experiments were conducted with real speech data. The input data, consisting of some ten best word hypotheses distributed over the utterance, were taken from the production of EVAR's current word hypothesizing module. The size of the samples is comparable to that processed by the HWIM parser (Woods 1976). Unfortunately, there are too few words (approx. 40%) carrying a quality score good enough

Themenschwerpunkt

to produce satisfying syntactic hypotheses. Nevertheless GuLP found all meaningful grammatical fragments in the quality order of the word hypotheses. The only reasonable strategy to achieve this result was table driven bottom-up because of overlaps and gaps between the available word hypotheses. Parsing times for the compiled version were of the order of 10 to 100 seconds (on a Siemens 7541).

One experience we made is that syntax does not play the role of a filtering step within a speech system. The essential bottleneck is the quality of word recognition; usually many constituent hypotheses are found which do not contain any of the really spoken words. If the grammar has a very high degree of selectivity, word recognition has to be very good. On the other hand, the grammar has to be 'robust' against ungrammaticality in speech, but must avoid that 'anything goes'.

7. References

- Dik S. (1979): Functional Grammar. Amsterdam: North-Holland
- Erman L. (1980): The Hearsay-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty. *Computing Surveys* 12, 213-253
- Goerz G. (1981): GLP: A General Linguistic Processor. In: Proc. IJCAI-81, Vancouver, B.C., 429-431
- Goerz G. (1982): Applying a Chart Parser to Speech Understanding. In: Proc. European A.I. Conference, Orsay/Paris
- Goerz G., Beckstein C. (1983): How to Parse Gaps in Spoken Utterances. In: Proc. First Europ. ACL Conf., Pisa
- Kaplan R., Bresnan J. (1982): Lexical Functional Grammar: A Formal System for Grammatical Representation. In: Bresnan (Ed.): *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press, 173-281
- Kay M. (1975): Syntactic Processing and Functional Sentence Perspective. In: Proc. TINLAP-I, Cambridge, Mass., 6-9
- Kay M. (1980): Algorithm Schemata and Data Structures in Syntactic Processing. Xerox Report CSL-80-12, Palo Alto, Calif.
- Niemann H. et al. (1984): A System for Understanding Continuous German Speech. *Information Sciences* 33, 87-112
- Sheil B. (1976): Some Observations on Context-Free Parsing. *Stat. Meth. in Linguistics* 6, 71-109
- Shieber S. (1984): The Design of a Computer Language for Linguistic Information. In: Proc. Coling-84, 362-366
- Thompson H. (1981): Chart Parsing and Rule Schemata in PSG. In: Proc. 19th Ann. Meeting ACL, Stanford, 167-172
- Thompson H., Ritchie G. (1984): Implementing Natural Language Parsers. In: O'Shea, Eisenstadt (Ed.): *Artificial Intelligence - Tools, Techniques and Applications*. New York: Harper & Row, 245-300
- Woods W. (1976): Speech Understanding Systems, Final Report - IV. Syntax and Semantics. BBN Report

SPRACHORIENTIERTE KI-Projekte in Hamburg:

zusammengestellt von Mitarbeitern der vorgestellten Projekte

Universität Hamburg
 Forschungsstelle für Informationswissenschaft und
 Künstliche Intelligenz
 Mittelweg 179
 D-2000 Hamburg 13

Universität Hamburg
 Fachbereich Informatik
 Schlüterstr. 70
 D-2000 Hamburg 13

Universität Hamburg
 Fachbereich Informatik
 Projektgruppe WISBER
 Jungiusstr. 6
 D-2000 Hamburg 36

Vier an der Universität Hamburg laufende sprachorientierte KI-Projekte werden vorgestellt: (1) LOKI, das den Entwurf und die Implementierung eines NL-Zugangssystems zu Datenbanken und Wissensbasen zum Ziel hat. (2) MARVIN, das den Benutzer eines Simulationsmodells bei seiner Arbeit unterstützt, wobei mehrere Kommunikationskanäle parallel genutzt werden können. (3) "Repräsentation und Verarbeitung räumlichen Wissens", das sich an Vorgehensweisen der Linguistik und der kognitiven Psychologie orientiert. (4) WISBER, ein umfangreiches BMFT-Verbundprojekt zur Entwicklung eines wissensbasierten Systems, das einen Beratungsdiallog mit einem "ratsuchenden" Benutzer führen soll. Die Projekte werden zunächst formal eingeordnet, Ziele und Ansätze werden kurz charakterisiert und der Stand der Arbeiten wird beschrieben.

LOKI**A LOGIC ORIENTED APPROACH TO KNOWLEDGE AND DATA BASES SUPPORTING NATURAL USER INTERACTION**

- Projektleitung: Prof. Dr. Walther von Hahn
- Mitarbeiter: Dr. Dipl.-Ing. Helmut Horacek
 Bill Imjan
 Dipl.-Inform. Claudius Pyka
 Dipl.-Inform. Martin Schröder
 Tom Wachtel, B.A., B. Phil.
- Adresse: Universität Hamburg
 Forschungsstelle für
 Informationswissenschaft und
 Künstliche Intelligenz
 Mittelweg 179
 2000 Hamburg 13
 Tel. 040/4123-3315, 4529, 2573, 2574
- Netzadresse: unido! fkinh.uuop! claude, helmut, martin,
 tom
- Förderungsprogramm: ESPRIT-Projekt 107 AIP 3.1, 2/3/4
- Projektpartner: Scicon Ltd. (London)
 Cranfield Institute of Technology
 BIM Belgian Institute of Management
 (Brüssel)
 SCS (Hamburg)
 Fraunhofer Institut IAO (Stuttgart)
 Technische Universität München
 Cretan Computer Institute (Iraklion)
- Koordination: (für Universität Hamburg)
 INCA e.V., Verein für intelligente Computer-
 Anwendungen
- Projektdauer: 1. Hauptphase: Aug. 84 - Feb. 87
 2. Hauptphase: Feb. 87 - Aug. 88 (geplant)
- Hard- und Software: Die Hardwareausstattung besteht aus einer
 SUN 2/120 unter UNIX bsd 4.2 mit vier
 Terminals sowie einem ATARI 520 ST+ für
 Textverarbeitung. Auf der SUN stehen die
 Sprachen BIMProlog und das DBMS UNIFY zur
 Verfügung.

Themenschwerpunkte

Inhaltliche Projektbeschreibung:

Inhaltliches Ziel des Hamburger Projektteils ist der Entwurf und die Implementierung eines natürlichsprachlichen Zugangs- und Dialogsystems zu Datenbanken und Wissensbasen. Als Pilotanwendung des NLI dient das Projektmanagementsystem PRADOS, des weiteren ist eine Anwendung in der Flugzeugkonstruktion geplant. Um eine einfache Portierung auf andere Problemereiche zu ermöglichen, soll der Kern des Systems weitgehend anwendungsunabhängig sein.

Das NLI wird sowohl für Deutsch als auch für Englisch realisiert. Die syntaktische Analyse orientiert sich an LFG, GPSG und ähnlichen Unifikationsgrammatiken. Für LFG steht ein Parser-Generator zur Verfügung.

Die semantische Repräsentation erfolgt durch zwei Repräsentationsformalisten: Zum einen ein KL-ONE-ähnliches semantisches Netzwerk zur Darstellung vom generischen und referenziellen Wissen, zum anderen eine Logik-orientierte Sprache zur Darstellung der Semantik natürlichsprachlicher Äußerungen.

Die Daten, die vom Projektmanagementsystem verwaltet werden sind im Datenbanksystem

UNIFY abgelegt. Der Zugriff auf das Datenbanksystem erfolgt vom BIMProlog-System aus in der Form von Prolog-Calls über ein integriertes Datenbankinterface.

Die Antworten des Dialogsystems an den Benutzer werden aus der semantischen Repräsentation generiert. Die Generierung von Äußerungen erfolgt ähnlich aufwendig wie die Analyse.

Ellipsen und Anaphern werden sowohl bei der Analyse- wie der Generierungsphase verarbeitet. Die Dialogfähigkeiten des Systems nehmen breiten Raum ein. Eine Unterstützung für globale Dialogstrategien ist eine explizite Berücksichtigung der Diskursstruktur durch Dialoggrammatik und Sprechakterkennung vorgesehen.

Das System zeigt ein weitgehend "natürliches" Verhalten durch Ausnutzen pragmatischer Informationen, z.B. pragmatische Schlüsselwörter. Auch einfache Pattern-Matching-Verfahren werden zu diesem Zweck eingesetzt. Bisher (Mai 1986) ist ein Rumpfsystem mit Schwerpunkt auf der morphologischen und syntaktischen Analyse einzelner Äußerungen und rudimentäre Pragmatik implementiert.

Veröffentlichungen:

Wachtel, Tom: Discourse Structure. LOKI-Report NLI-1.1, April 1985

Horáček, Helmut: Semantic/Pragmatic Representation Language. LOKI-Report NLI-2.1, Dezember 1985

Pyka, Claudius: Syntactic Analysis. LOKI-Report NLI-4.1, November 1985

MARVIN EIN SIMULATIONS LABORATORIUM MIT MULTIMEDIALEM ZUGANG

Mitarbeiter:

P. Scheffe, M. Hußmann und Studenten

Adresse:

Prof. Dr. Peter Scheffe
Fachbereich Informatik, Univ. Hamburg
Schlüterstr. 70
2000 Hamburg 13
Tel: 040/4123-4157

Projektrahmen:

Fachbereich Informatik, Arbeitsbereich "Anwendungen der Informatik in Geistes- und Naturwissenschaften", Universität Hamburg zur Zeit keine Drittmittelförderung

Hard- und Software:

Zur Zeit UCI LISP (erweitert um die objektorientierte Programmiersprache OBJEKT [Hußmann 85]) und Edinburgh PROLOG auf einer DECsystem-10 unter TOPS-10; ab 1986 CommonLISP und Quintus PROLOG auf einer SUN unter UNIX.

Inhaltliche Projektbeschreibung:

MARVIN ist der Nachfolger des Systems SWYSS (= "Say what You See System, 1979 - 1984). SWYSS war ein Dialogsystem, das einen natürlichsprachlichen Zugang zu Beschreibungen statischer Szenen (z.B. eines Zimmers mit verschiedenen Möbeln usw.) realisierte. Obwohl ein Teil der für SWYSS entwickelten Module und Repräsentationssprachen auch im neuen System Verwendung finden soll, ist die Konzeption MARVIN doch eine gänzlich andere:

Den Kern von MARVIN bildet ein System zur qualitativen Simulation von Ökosystemen, speziell stehenden Gewässern. In einem "Simulationslaboratorium" soll MARVIN der Benutzer (mutmaßlich ein Biologe) bei der Entwicklung von Modellen des Gegenstandsbereichs und verschiedener (alternativer oder sich ergänzender) Simulationsmodellen deren Validierung und der Auswertung ihrer Resultate unterstützen.

Der Zugang des Benutzers zum Simulationslaboratorium soll parallel über verschiedene Kommunikationskanäle und -medien erfolgen (u.a. natürliche Sprache, formale Sprachen, Grafiken, Maus- und Ikonenorientierte Kommunikation). Der Zusammenhang der Dialogäußerungen über verschiedene Kanäle soll durch die einheitliche Terminologie (einundderselbe Begriff kann mehrere, äquivalente externe Repräsentationen, z.B. als Wort einer natürlichen Sprache, als grafisches Symbol etc. besitzen) und durch die Kanalübergreifende Modellierung der Dialogstruktur (die z.B. anaphorische Bezüge auf Referenten ermöglicht, die über einen anderen Kanal eingeführt wurden) hergestellt werden. Forschungsschwerpunkte sind derzeit:

und wort-orientierten Paradigmen (hier realisiert durch eine kontextfreie Phrasenstrukturgrammatik bzw. eine durch Wortexperten implementierte Valenzgrammatik) miteinander kombiniert. Der Determinismus soll dabei allein durch die Parsingstrategie erreicht werden;

- qualitative Simulation, insbesondere Möglichkeiten der Interaktion zwischen Simulations- und Expertensystemen, wobei letztere zur Überwachung und Steuerung des Simulationsablaufs sowie zur Aufbereitung und Interpretation der Resultate eingesetzt werden sollen; die Beschreibung von Simulationsmodellen auf der Ebene der biologischen Begriffsbildung;
- die Entwicklung eines deterministischen Parsers mit begrenztem Speicher, dessen Grammatikmodell die regelorientierten

- Generierung natürlicher Sprache;
- die Erkennung kommunikativer Akte und die Modellierung von multi-medialen Dialogstrukturen;
- Wissensrepräsentation auf der Basis informationeller Konzepte, insbesondere Probleme der Definition von Begriffen und ihrer Einordnung in taxonomischen Strukturen;
- Probleme der Steuerung des Daten- und Kontrolleinflusses in komplexen modularen Systemen zur inkrementellen, pseudo-parallelen Analyse und Generierung (u.a.) natürlicher Sprache.

"MARVIN" ist im übrigen kein Akronym; vielmehr wurde das System nach einem Protagonisten aus Douglas Adams' Science Fiction Trilogie "The Hitch Hiker's Guide to the Galaxy" benannt.

Veröffentlichungen (zu SWYSS und MARVIN):

Peter Scheffé: On Foundations of Reasoning with Uncertain Facts and Vague Concepts. In: International Journal of Man-Machine Studies, 12, 1980, 35-62.

Bernd Pretschner: Die Behandlung natürlichsprachlicher Quantifizierungen in Frage-Antwort-Systemen. Diplomarbeit, Fachbereich Informatik, Universität Hamburg, 1980.

Karl-Jürgen Hanßmann: Sprachliche Bildinterpretation für ein Frage-Antwort-System. Report IfI--HH-M-74/80. Fachbereich Informatik, Universität Hamburg, 1980.

Peter Scheffé, Bernd Pretschner: SWYSS - A Natural Language Question-Answering System for Scene-Analysis. Report IfI--HH-M-88/81. Fachbereich Informatik, Universität Hamburg, 1981.

Michael Hußmann: Ellipsenrekonstruktion im Kontext: Forderungen an eine Ellipsenkomponente und ihre Realisierung in SWYSS. In: Wolfgang Wahlster (ed.): GWA1-82 - Proc. of the 6th German Workshop on Artificial Intelligence, Bad Honnef, September 1982. Berlin, Heidelberg, New York: Springer 1982, 227-236.

Michael Hußmann: Ein VEHIQUEL für den natürlichsprachlichen Zugang zu Datenbanken. In: Bernd Neumann (ed.): GWA1-83-Proc. of the 7th German Workshop on Artificial Intelligence, Dassel/Solling, September 1983. Berlin, Heidelberg, New York, Tokyo: Springer 1983, 69-79.

Peter Scheffé: Natürlichsprachlicher Zugang zu Datenbanken? In: angewandte Informatik, 10, 1983.

Michael Hußmann, Peter Scheffé: The Design of MARVIN, a Dialogue System for Scene Analysis. In: Leonard Bolc (ed.): Natural Language Communication with Pictorial Information Systems. Berlin, Heidelberg, New York, Tokyo: Springer 1984, 143-201.

Michael Hußmann, Martin Bode, Michael Borowiak, Peter Scheffé, Bernd Wronn: Extending the Natural Language System SWYSS by a Functional Database. In: Tim O'Shea (ed.): ECAI-84 - Proc. of the 6th European Conference on Artificial Intelligence, Pisa, September 1984, 217 - 220.

Michael Hußmann: IS-A Isn't Enough - Towards a Taxonomic Framework for Intensional Concepts. Vorgetragen auf der ECAI-84, Pisa, September 1984. Zweite, erweiterte Version in Vorbereitung.

Michael Hußmann: OBJECT Manual. Fachbereich Informatik, Universität Hamburg 1985.

Peter Scheffé: Zur Rekonstruktion von Wissen in neueren Repräsentationssprachen der künstlichen Intelligenz. In: Herbert Stoyan (ed.): GkAI-85-Proc. of the 9th German Workshop on Artificial Intelligence. Dassel/Solling, September 1985. Berlin, Heidelberg, New York, Tokyo: Springer 1985, 230-244.

Michael Hußmann: Die Rekonstruktion von Begriffen und Strukturen in natürlichsprachlichen Datenbankfragen. In Vorbereitung.

Michael Hußmann: Tempusbeschreibung einmal ernstgenommen - die Rekonstruktion temporaler Ausdrucksmittel des Deutschen in DSL/3. In Vorbereitung.

REPRÄSENTATION UND VERARBEITUNG RÄUMLICHEN WISSENS

Projektleiter: Prof. Dr. Christopher Habel

Adresse: Universität Hamburg, Fachbereich Informatik
Schlüterstr. 70
2000 Hamburg 13

Projektrahmen: Arbeitsbereich "Information und Dokumentation"

Inhaltliche Projektbeschreibung:

Dieses Projekt befindet sich momentan noch in der Planungsphase, der Beginn ist für den Sommer 1986 vorgesehen. Das Hauptziel soll die Entwicklung von Wissensrepräsentationsformalismen und Verarbeitungsmethoden für räumliches Wissen sein. Die Vorgehensweise wird an denen der Linguistik und der kognitiven Psychologie orientiert sein, wobei deren spezifische Ansätze KI-mäßig realisiert werden sollen.

Dabei sollen folgende Schwerpunkte bearbeitet werden:

- 1) Analyse räumlicher Ausdrücke im Deutschen
Für einen ausgewählten Diskursbereich werden die dort auftretenden lexikalischen, syntaktischen und semantischen Phänomene untersucht.
- 2) Propositionale Repräsentationen
Es soll ein formales, logik-orientiertes System zur propositionalen Repräsentation räumlichen Wissens entwickelt werden unter Berücksichtigung topologischer, metrischer und Gestalt-Eigenschaften.
- 3) Bilder und Depiktionen
Da Menschen zur Verarbeitung räumlichen Wissens, z.B. bei der Perzeption oder bei räumlichen Vorstellungen, (quasi-) analoge, "depiktionale" Repräsentationen neben den oben angesprochenen propositionalen verwenden, soll ein

weiterer Punkt die Entwicklung einer Repräsentationssysteme für Depiktionen sein.

- 4) Deskriptionen und Depiktionen
Um "bildhaftes" und propositionales Wissen erfolgreich verknüpfen zu können, müssen die Beziehungen und Übersetzungsprozesse zwischen beiden Repräsentationstypen geklärt werden.
- 5) Unvollständigkeit und Inkonsistenz
Es soll versucht werden, die speziellen Probleme bei unvollständigen und inkonsistentem räumlichen Wissen zu behandeln.

An diesem Forschungsschwerpunkt werden Mitarbeiter des Arbeitsbereiches Information und Dokumentation und Drittmittelbeschäftigte arbeiten. Das Projekt selbst wird wahrscheinlich drei Wissenschaftliche Mitarbeiter sowie studentische Hilfskräfte umfassen und von der IBM Deutschland gefördert werden.

Projektpartner sind neben der IBM Forschungsabteilung die Universitäten Stuttgart, Osnabrück und Tübingen, die im Gesamtprojekt LILOG (Linguistische und logische Methoden und Grundlagen) zusammengefaßt sind. Als Projektdauer sind zweimal 2 1/2 Jahre, also insgesamt 5 Jahre, vorgesehen.

WISBER WISSENSBASIERTER BERATUNGSDIALOG

- Projektleitung:** Prof. Dr. B. Neumann
Dr. W. Hoepfner
- Mitarbeiter:** Dipl.-Inform. H. Bergmann
Dr. R. Block
Dipl.-Inform. M. Gerlach
Dipl.-Inform. H. Marburger
- Adresse:** Universität Hamburg
Fachbereich Informatik
Projektgruppe WISBER
Jungiusstr. 6
Postfach 302762
D-Hamburg 36
Tel. 040/4123-2595
- Förderungsprogramm:** BMFT-Verbundprojekt im Themenbereich 3d:
Wissensbasierter Mensch-Maschine-Dialog
- Förderungskennzeichen:** ITW 5802 0
- Projektpartner:** Nixdorf Computer AG, Projektteam Künstliche
Intelligenz, Paderborn
SCS Organisationsberatung und Informations-
technik
GmbH, Fachgebiet Wissensbasierte Systeme,
Hamburg
Siemens AG, ZT ZTI INF 311, München
Universität Hamburg
- Koordination:** Universität des Saarlandes
- Projektdauer:** Gesamtprojekt: 1. 1. 85 - 31. 12. 88
Universität Hamburg: 1. 10. 85 - 31. 12. 88
- Hardware:** Siemens EMS 58xx, Symbolics 3600,
Nixdorf Targon/31, Modell 30
- Software:** INTERLISP-D, Zeta LISP, PROLOG
- Projektziel:** Ziel des Verbundvorhabens ist die
theoretische Fundierung, der Entwurf und die
Implementierung des deutschsprachigen
Beratungssystems WISBER (ausgesprochen wie
englisch "whisper"), mit dem ein ziel- und
problemorientierter kohärenter Beratungsdialog
in dem exemplarischen Anwendungsbereich
der Geldanlageberatung geführt werden kann.

Inhaltliche Projektbeschreibung:

Kennzeichnend für Beratungsdialoge - im Gegensatz zu anderen Dialogformen - ist die folgende Kommunikationssituation: einer der Partner gesteht dem anderen auf einem bestimmten Gebiet Beratungskompetenz zu, die ihm für bestimmte eigene Zwecke nützlich erscheint, d.h. die Rollenverteilung in einem Beratungsgespräch ist festgelegt. Der Berater verfügt im Gegensatz zum Ratsuchenden über umfangreiches Sach- und Methodenwissen auf dem Problemgebiet. Dieses Wissen wird vom Ratsuchenden anerkannt und die Aussagen des Beraters werden als seriöse Äußerungen angesehen. Der ratsuchende Partner hat die Erwartung, daß der Experte sich auf seine Probleme einstellen kann und ihm aus der Menge aller möglichen Lösungen den für ihn optimalen Lösungsvorschlag macht. Dabei muß der Berater sich auf jeden einzelnen Ratsuchenden einstellen können: nicht das aktuelle Problem, lediglich das Problemgebiet liegt vor Gesprächsbeginn fest. Der Ratsuchende hat die Option, die Empfehlungen des Beraters zu akzeptieren oder zu erwerben.

In der Mensch-Maschine-Interaktion wird die Rolle des Ratgebers vom System übernommen, das über umfangreiches, beraterungsgebietsspezifisches Wissen verfügt. Wesentliche, für die Beratungsleistung notwendige Systemfähigkeiten sind dabei das Erkennen des Wissensstandes und der Ziele des Benutzers sowie das adäquate Auswählen und Repräsentieren von Detailwissen. Natürlich setzen diese beratungsimmanenten Fähigkeiten die für natürlichsprachliche Dialogsysteme erforderlichen Eigenschaften voraus.

Da sich das Projekt in Hamburg noch in einem sehr frühen Stadium befindet (Stand 1.4.86), können noch keine detaillierten Aussagen über entwickelte Systemkomponenten gemacht werden. Die Ausführungen beschränkten sich deshalb auf eine Beschreibung der geplanten Entwicklungen, die z.T. in Zusammenarbeit mit den Projektpartnern durchgeführt werden sollen.

Themenschwerpunkt

Evaluationsstudie

Erstellung einer wissenschaftlichen Beurteilung von Systemen in Hinblick auf Beratungssituationen. Untersucht werden u.a. bestehende vollständige natürlich-sprachliche Dialogsysteme, Systeme, die beraterrelevante Aspekte beinhalten (Tutorsysteme, Hilfesysteme), und auch Teilsysteme, die spezielle Fähigkeiten von Beratungssystemen modellieren (Erkennen und Erzeugung von Sprechakten, kooperatives Verhalten bei Fehlannahmen usw.). Diese Studie bildet die Grundlage für den Entwurf der Systemarchitektur von WISBER.

Inhaltsorientierte Repräsentationssprache

Entwurf und Realisierung einer zentralen Repräsentationssprache, die neben der semantischen Repräsentation der Benutzer- und Systemäußerungen auch die Darstellung der pragmatischen Funktion von Äußerungen erlaubt. Diese Repräsentationssprache dient als Zielsprache der Analysekomponenten, als Quellsprache der Generierungskomponenten, sowie als vermittelndes Element zwischen den verschiedenen Wissensquellen des Systems.

Benutzermodellierung

Die Benutzermodellierung darf sich wegen der besonderen Bedeutung in Beratungsdialogen nicht auf einzelne Aspekte beschränken, sondern muß für alle Systemkomponenten gleichmäßig ausgebaut sein. Eine adäquate Benutzermodellierung verlangt, neben der Modellierung des Wissens über den Benutzer (z.B. Vorannahmen, dialogbasiertes Wissen, Begrifflichkeiten, usw.), der Modellierung seiner kommunikativen und aufgabenorientierten Ziele und Pläne, auch die Modellierung des Wissens über den Sachbereich (Planwissen, Weltwissen) sowie Wissen über Beratungen an sich (Problemspezifikationswissen, Ableitungen von Rahmenbedingungen aus Benutzeraußerungen usw.).

Semantisch-pragmatische Analyse

Hauptaufgabe der semantisch-pragmatischen Analyse ist das Erkennen von Sprechakten, von kommunikativen und aufgabenorientierten Zielen und Plänen des Benutzers, sowie die Überführung der Benutzeraußerung in die zentrale Repräsentationssprache. Von den zahlreichen Untersuchungsschwerpunkten in diesem Bereich seien hier einige genannt: Unterscheidung zwischen hypothetischen und

realen Sachverhaltsschilderungen, Einsatz von Wissen über Dialogstand und Dialogfortsetzungsmöglichkeiten zur Disambiguierung von Benutzeraußerungen, Auflösung von anaphorischer Referenzen, die sich auf Sachverhalte, Handlungsbeschreibungen und Dialogobjekte beziehen.

Generierung von Systemäußerungen

In einem Beratungsgespräch mit gemischter Initiative muß das System neben der direkten Beantwortung von Benutzerfragen eine Anzahl heterogener Sprechakte erzeugen, z.B. Fragen, Rückversicherungen, Vorschläge, Handlungsanweisungen, Bestätigungen, Empfehlungen, Korrekturen, die darauf abzielen, in größeren Texteinheiten zusammengefaßt werden. Entwicklungsschwerpunkte liegen sowohl auf der Bestimmung des Inhalts der Äußerungen (Auswahl der relevanten Teile der Wissensquellen, Bestimmung des Detaillierungsgrades, Antizipation des Verstehensprozesses des Benutzers) als auch auf der Bestimmung der Form der Äußerung (Linearisierung, Ellipsengenerierung, Pronominalisierung, Topikalisierung, Wortwahl, Wahl der syntaktischen Gestalt).

Dialogsteuerung

Innerhalb eines Beratungsdialogs sind unterschiedliche Dialogphasen identifizierbar, z.B. Problemschilderung, Lösungsvorschlag, Aushandeln des Lösungsvorschlages. Diese groben Phasen sind in feinere Dialogschritte zerlegbar. Um einen ziel- und problemorientierten, kohärenten Dialog führen zu können, muß das System über eine Vielzahl von Dialogstrategien verfügen, sowie über Kriterien, die entscheiden, welche Strategie in welchem Stand des Dialogs und des Problemlösungsprozesses die geeignetste ist.

Wissensbasisverwaltungssystem

Die Dialog- und Beratungsfähigkeit eines natürlich-sprachlichen Systems stützt sich auf eine große Anzahl heterogener, z.T. dynamischer, diskursbereichsabhängiger, u.U. unabhängiger Wissensquellen, die auf vielfältige Art und Weise untereinander verknüpft sind. Zur Unterstützung der Wissensakquisition, der Wartung dieser Wissensquellen und zur Optimierung der Zugriffe werden notwendige Teile eines Wissensbasisverwaltungssystems entwickelt.

TOPIC II / TOPOGRAPHIC II

ZUR ENTWICKLUNG EINES WISSENSBASIERTEN VOLLTEXT-INFORMATIONSSYSTEMS AM LEHRSTUHL FÜR INFORMATIONSWISSENSCHAFT DER UNIVERSITÄT KONSTANZ

Udo Hahn, Rainer Hammwöhner, Ulrich Reimer, Ulrich Thiel

Universität Konstanz
Informationswissenschaft
Projekt TOPIC II / TOPOGRAPHIC II
Postfach 5560
D-7750 Konstanz 1

Am Lehrstuhl für Informationswissenschaft der Universität Konstanz wird seit 1982 an der Entwicklung eines Prototypen eines wissensbasierten Volltext-Informationssystems gearbeitet. Das TOPIC-System soll deutschsprachige Fachtexte zusammenfassen, das TOPOGRAPHIC-System soll entsprechend kondensierte Textrepräsentationen graphisch-interaktiv zugänglich machen. Die wesentlichen Funktionsmerkmale und die methodischen Konzepte beider Systeme werden beschrieben.

Von der zunehmenden Verbreitung rechnergestützter Verfahren zur Erstellung, Speicherung und Verteilung von Texten, etwa in Textsystemen, Volltext-Informationenbanken, Textkommunikations- (text filing) und elektronischen Publikationssystemen geht unmittelbar ein qualitativ neuartiges Nachweis- und Präsentationsbedürfnis für Volltexte aus. Denn es stellt sich vor dem Hintergrund solcher Texttechnologien das Informationsproblem, relevante Texte, die prinzipiell nicht in gedruckter, sondern allein in maschinenlesbarer Form vorliegen, nachzuweisen, informationell zu verdichten bzw. signifikante Fakten aus den Originaltexten gezielt bereitzustellen. Dieser Anspruch hinsichtlich referenz-, kondensat- und faktenorientierter Retrievaloperationen auf Volltexten verlangt Textanalyseverfahren, die in der Lage sein müssen, sowohl die Ausprägung und Organisation thematischer Schwerpunkte in Texten zu erkennen als auch die Darstellung aus Texten extrahierter Fakten zu ermöglichen und dieses Wissen in entsprechende 'Textwissensbasen' zu überführen. Das Retrieval auf diesen Textwissensbasen muß ebenso an die dort abgelegten komplexen Wissensstrukturen angepaßt werden wie die Interaktions- und Präsentationsstile beim Browsing und bei der Frageformulierung. Unter Retrievalaspekten müssen neben ausschließlich referenzorientierten Modi des klassischen Information Retrieval auch dem Niveau der Textanalyse entsprechende kondensat- und faktenorientierte Retrievalmechanismen sowie Zugriffsmechanismen auf die Volltexte selbst unterstützt werden. Unter Schnittstellenaspekten sollte die Benutzerinteraktion insbesondere auch auf die informationellen Sondierungs- und Zugriffsmuster von Informationssuchenden zugeschnitten sein. Der eingeschränkte Funktionsvorrat bislang gebräuchlicher Auswahl- und Orientierungstechniken (Anwahl von Textstellen durch Vorgabe von Stringmustern, Seiten- oder Zeilenangaben, Blättern usw.) erweist sich bei einer so

umfassenden Funktionalität eines Volltext-Informationssystems als strukturell zu einfach. Vielmehr bietet sich wegen der hohen Informationsdichte, Abstraktionskraft und Anschaulichkeit graphischer Darstellungsmittel insbesondere bei der Verfügbarkeit strukturierter Wissensrepräsentationsformen ein graphischer statt natürlichsprachlicher Interaktions- und Retrievalmodus mit Textwissensbasen an. Schließlich folgt aus dem massiven Mengenwachstum elektronisch gespeicherter Textbestände in der Informationspraxis auch ein beachtlicher quantitativer Impuls, der neben der oben skizzierten Bedürfnis- auch die Bedarfskomponente für die Überwindung von einfachen Volltextspeicherungssystemen in Richtung auf entwickeltere Volltextinformationssysteme unter Einbeziehung der oben skizzierten Textkondensierungs- und Textretrievalkonzepte unterstreicht.

Am Lehrstuhl für Informationswissenschaft der Universität Konstanz wird daher seit 1982 an der Entwicklung eines Prototypen eines wissensbasierten Volltext-Informationssystems in der Zuständigkeit von zwei Projektgruppen gearbeitet:

- * die Zusammenfassung deutschsprachiger Fachtexte aus dem Gebiet der Informationstechnik ist Gegenstand des TOPIC-Systems (HAHN/REIMER 1985a, b, 1986)
- * die interaktiv-graphische Präsentation und das Retrieval auf entsprechend kondensierten Textrepräsentationsstrukturen beschreibt den Aufgabenbereich der TOPOGRAPHIC-Komponente (HAMMWÖHNER/THIEL 1984, THIEL/HAMMWÖHNER 1986)

Beide Systeme werden im folgenden unter den Gesichtspunkten ihrer wesentlichen Funktionsmerkmale und methodischen Konzepte kurz vorgestellt:

TOPICs Funktionalität kann durch folgende Merkmale charakterisiert werden:

Textzusammenfassung (Abstracting)

TOPIC erzeugt eine Graph-Repräsentation der thematischen Schwerpunkte eines Texts. Diese Form der inhaltlichen Zusammenfassung beruht auf Textrepräsentationsstrukturen, deren Explizitheitsgrad unterschiedlich verschärft werden kann: Sie variiert von relativ generischen Beschreibungen (durchaus einem System von Index-Terms vergleichbar) über kohärente Darstellungen der thematischen Organisation des Texts bis zu spezialisierten Informationen zu relevanten Themen in Form von Fakten, neu gelernten Konzepten und ihnen zugeordneten Eigenschaften. Dieses Konzept unterschiedlicher informationeller Verdichtungsstufen der Textzusammenfassung, das wir als kaskadierte Textkondensierung (KUHLEN 1984) bezeichnen, unterscheidet sich grundlegend von anderen Ansätzen, die auf der Voreinstellung eines Interessenprofils oder einer zu erwartenden Textstruktur basieren.

Faktenextraktion / Akquisition neuer Konzepte

Für diese qualitativ anspruchsvolle Form der Textzusammenfassung werden nicht nur einfache Formen der Wissensextraktion aus Texten bereitgestellt - etwa die Zuweisung bestimmter Eigenschaften zu bereits bekannten Konzepten -, sondern zusätzlich auch Heuristiken für das Lernen neuer Konzepte und ihnen zuzuordnender Eigenschaften (eine dafür besonders charakteristische Erweiterung des Wissensrepräsentationsmodells beschreibt REIMER 1986b).

Verknüpfung von Themenbeschreibungen mit Textpassagen

TOPICs Analyseverfahren sind aus Gründen der angestrebten Systemfunktionalität nicht daraufhin ausgelegt, das gesamte im Text kodierte Wissen zu erfassen. Die Textrepräsentationsstrukturen können daher unvollständig sein. Deswegen werden die Beschreibungen thematischer Schwerpunkte eines Texts mit den ihnen entsprechenden Passagen des Originaltexts verbunden, so daß Anfragen an die Textwissensbasis unter bestimmten Umständen mit signifikanten Ausschnitten des Originaltexts beantwortet werden können (passage retrieval).

Diese Funktionalität des TOPIC-Systems leitet sich aus folgenden methodischen Grundlagen ab:

Partielles Text-Parsing

Die aktuelle Version des TOPIC-Systems ist auf ein eher "flaches" Textverstehen ausgerichtet (HAHN/REIMER 1984a). Sie konzentriert sich dabei auf die thematischen (nicht-propositionalen) Textinhalte (aboutness) und ihnen entsprechenden Fakten.

Damit ist ein Niveau des Textverstehens abgesteckt, das für die Bereitstellbarkeit indikativer Textkondensate völlig hinreichend ist. Dieses partielle Parsing wird durch eine Beschränkung der Textanalyse auf vorwiegend taxonomische Wissensrepräsentationsstrukturen und linguistische Spezifikationen realisiert, die diesem eingeschränkten Repräsentationsniveau angepaßt sind. Konzepte des Weltwissens sind daher primär Nomen bzw. nominale Gruppen und ihre Attribute (Adjektive, numerische Werte usf.).

Durch Integritätsbedingungen angereichertes Frame-Repräsentationsmodell

Das der Textanalyse zugrunde liegende Weltwissen wird durch ein Frame-Repräsentationsmodell dargestellt (REIMER/HAHN 1983; REIMER 1986a). Empirisch unterstützt die starke Schematisierung von Wissen, wie sie Frame-Modellen inhärent ist, die Textanalyse durch die unmittelbare Verfügbarkeit der semantischen Umgebungen der im Text aktualisierten Konzepte - das ist besonders wichtig für grundlegende Mechanismen der Textkohäsion. Formal bewirkt die Spezifikation von Integritätsbedingungen eine scharfe Kontrolle der auf Welt- und Textwissensbase durchführbaren Operationen und sichert damit in hohem Maß die Validität dieser Wissensbasen - ein besonders bei realistisch dimensionierten Textmengen, die charakteristisch für Anwendungen im Bereich des Information Retrieval sind, wesentliches Modellierungskriterium.

Text-Parsing mit Schwerpunkt auf Textkohäsions- und Textkohärenzmustern

Die Textualität von Texten beruht (etwa in der Abgrenzung zu Einzelsätzen) auf zwei Klassen von Phänomenen: lokaler Kohäsivität (Textkohäsion durch Anaphora, lexikalischer Kohäsion, Koordination, Ellipsen usw.) und globaler thematischer Wohlgeformtheit (Textkohärenz durch thematische Progressionsmuster, funktionale Kohärenzrelationen etc.). Ihre Mißachtung durch satzorientierte Parsing-Verfahren führte unweigerlich zu invaliden (Textkohäsion) und unterstrukturierten (Textkohärenz) Textwissensbasen. Textverstehenssysteme müssen diesen Bedingungen daher durch entsprechend modellierte Textgrammatiken gerecht werden.

Lexikalisch verteilte Organisation einer semantischen Textgrammatik

Die Mehrzahl dieser Textkohäsions- und -kohärenzmuster können durch primär semantische Regularitäten erklärt werden. Daher bietet sich unmittelbar ein semantischer Ansatz für die Formulierung einer Textgrammatik an, der sowohl konzeptuelle Restriktionen der Diskursdomäne (sie wird durch das Frame-Modell repräsentiert) als auch strukturelle Eigenschaften der behandelten Text(klass)e (sie sind wesentlicher Bestandteil textgrammatischer Wohlgeformtheitsbedingungen) einbezieht. Da sich die semantische Basierung des Textverstehens

Themenschwerpunkt

direkt - insbesondere bei einem auf thematische Schwerpunkte ausgerichteten Analysemodell - in der Interaktion der Bedeutungen der im Text auftretenden lexikalischen Elemente widerspiegelt, bilden sie den unmittelbaren Ansatzpunkt für eine lexikalische Verteilung grammatikalischen Wissens im TOPIC-System (HAHN/REIMER 1983, HAHN 1984, 1986b) und seine Organisation im Format eines generalisierten Wortexperten-Modells (HAHN 1986a).

Textgraphen als Repräsentationsstruktur für Textkondensate

Die beim Text-Parsing aufgebauten Textrepräsentationsstrukturen werden einem Kondensierungsprozeß unterworfen, der sie in eine Kondensatrepräsentation auf unterschiedlichen Ebenen inhaltlicher Verdichtung überführt und in Form eines Textgraphen organisiert (HAHN/REIMER 1984b). Aus dieser Repräsentationsstruktur sind dann auch Zugriffe auf das Faktenwissen, das während der Textanalyse gewonnen wird, möglich. Die Form des Zugriffs und der Interaktion mit diesen Wissensstrukturen ist mit dem Retrieval-Interface TOPOGRAPHIC ausschließlich graphisch organisiert - TOPIC verfügt also über keine eigenen natürlichsprachlichen Textgenerierungsmechanismen.

Das bis 12/86 laufende Projekt TOPIC II dient der experimentellen Überprüfung der bislang vorliegenden Wissensrepräsentations-, Text-Parsing- und Textkondensierungsmodelle sowie der kontinuierlichen Erweiterung der empirischen (Wortexperten- und Frame-)Basis des Gesamtsystems, um im Rahmen von Test/Änderungszyklen diese Modelle schrittweise experimentell zu modifizieren und zu erweitern. Darüber hinaus wird in der zweiten Projektphase der typischen Information-Retrieval-Situation Rechnung getragen, relevante Informationen aus einer Dokumentenkollektion auszuwählen und die Beschränkung auf Einzeltexte (wie sie für TOPIC I noch galt) aufzugeben. TOPIC II wird dafür ein geeignetes frame-basiertes Retrievalmodell fuer Textkondensate sowie ein daran angepaßtes Textwissensverwaltungssystem (auf der Basis des fuer das Weltwissen realisierten Verwaltungssystems) bereitstellen. Ein wesentliches Forschungsproblem ist dabei die Entwicklung partieller Matching-Verfahren für komplexe Kondensat- und Anfrage-Graphen. Das Ergebnis des zweiten Projektteils soll ein experimentell verifiziertes, formal beschriebenes Modell für das zugrunde gelegte indikative Niveau der Textkondensierung und ein dazugehöriger lauffähiger Prototyp sein. Zusammen mit TOPOGRAPHIC soll ein Textkondensierung und Text-Retrieval integrierender Prototyp eines wissensbasierten Volltextinformationssystems implementiert werden, wobei im Sinne einer analytischen Trennung von TOPIC als einer Kondensatmaschine und TOPOGRAPHIC als

einer Interface- und Retrievalmaschine gesprochen werden kann.

T O P O G R A P H I C

Im Projekt TOPOGRAPHIC werden mit einer experimentellen Benutzerschnittstelle wissensbasierte Informationsvermittlungsformen entwickelt. Als Basis dient das oben beschriebene Textkondensierungssystem TOPIC, das die Inhalte der analysierten Texte über Wissens- und Textrepräsentationsstrukturen in gestaffelten Kondensierungsgraden bereithält. Damit bietet sich die Möglichkeit, mit graphisch-interaktiven Mitteln ein Analogon zum Abstracting zu modellieren, das gleichzeitig Indexstrukturen für das Retrieval von Textinformationen zur Verfügung stellt.

Im Gegensatz zu der bisher häufig vertretenen Auffassung, die Ausgabe eines Informationssystems sollte möglichst nahe an einer natürlichsprachigen Form realisiert werden, soll alternativ dazu eine Präsentation mit graphischen Mitteln entwickelt werden. Dies hat folgenden Hintergrund:

1. Zur Darstellung komplexer, stark vernetzter Strukturen, wie sie für TOPICs Analyseresultate (Textgraphen) charakteristisch sind, eignen sich Graphiken besonders gut und sind eventuell sogar textuellen Darstellungsformen überlegen (HAMMÖHNER/THIEL 1985, 1986a), weil sie Textstrukturen transparent machen und relevante Information hervorheben.
2. Spezielle Techniken erlauben eine Anpassung des Informationsniveaus an die Bedürfnisse des Benutzers und erleichtern ihm die Orientierung:
 - Das zielgerichtete Navigieren im Weltwissen erlaubt ein schnelles Aufsuchen der relevanten Information.
 - Falls ein präzise formulierter Informationsbedarf vorliegt, kann er durch direkte Anwahl von Suchbegriffen gedeckt werden.
 - Zooming-Techniken gestatten die Darstellung von Wissen in unterschiedlichen Abstraktionsgraden. Optionen dieser Art sind ohne Graphik nicht realisierbar und stellen eine eigene Informationsvermittlungsvariante dar.
 - Sukzessive Umgestaltung des Bildschirminhaltes unterstützt die Perception der dargestellten Strukturen. Dabei können zwischen einzelnen Begriffen bestehende konzeptuelle Bezüge ausgenutzt und gegebenenfalls auch verstärkt werden.

In der ersten Projektphase (bis 12/84) ist ein prototypisches Benutzerinterface für Systementwickler implementiert worden, das vor allem auf infrastrukturelle Dienste bezogen war. Mit ihm wurde die experimentelle Visualisierung von stark vernetzten

Wissensstrukturen in parallel dazu entwickelte infrastrukturelle Software-Werkzeuge integriert. Die auf diese Weise graphisch unterstützten Wortexperten-, Welt- und Textwissens-Editoren gestatten den Zugriff auf diese Systemkomponenten und den Aufbau bzw. die Veränderung von Wortexperten, Frames und Textrepräsentationsstrukturen (THIEL/HAMMÖHNER 1985, HAMMÖHNER/THIEL 1985). Graphisch-interaktive Manipulationsmöglichkeiten und Varianten der Benutzerführung wurden empirisch auf ihre Angemessenheit getestet.

In der aktuellen Projektphase (bis 12/86) werden folgende weiter gesteckte Ziele verfolgt:

Mithilfe eines prototypischen Endbenutzer-Interface sollen die in der Konzeptionsphase postulierten kognitiven Hypothesen, insbesondere über das geeignete Verhältnis von textueller und graphischer Information, einer experimentellen Überprüfung unterworfen werden. Die gewonnenen Erkenntnisse bilden das Fundament für die Erweiterung des Systems, die primär der Evaluierung graphischer Dialogformen und den Dialog steuernder Retrievalheuristiken dient.

Das von TOPIC und TOPOGRAPHIC gebotene Funktionsspektrum (Welt- bzw. Textwissensverwaltung, Graphiksoftware zur Präsentation informationeller Strukturen) wurde in PROLOG eingebettet (HAMMÖHNER/THIEL 1986b), so daß in einem "rapid prototyping" eine graphisch-interaktive "Retrievalmaschine" (THIEL/HAMMÖHNER 1986) entwickelt werden kann, die durch folgende Konzepte charakterisierbar ist:

- Das Weltwissen wird als Suchbegriffsraum verwendet, in dem mit den Operationen von TOPOGRAPHIC navigiert werden kann.
- Die Definition von Suchbegriffsprofilen erfolgt durch Auswahl von Teilnetzen und deren Manipulation mit einer graphischen Retrievalsprache.
- Das ausgewählte Profil wird mit Textwissensstrukturen verglichen, wobei durch partielles Matching von Textgraphen auch Texte berücksichtigt werden können, die dem Suchprofil nur teilweise entsprechen.
- Die Annäherung an das endgültige Retrievalergebnis erfolgt iterativ, indem mit den jeweiligen Zwischenergebnissen in einem Feed-Back-Verfahren die optimale Abfrage erarbeitet wird.
- Durch "informationelles Zooming" kann der Benutzer sukzessive detailliertere Informationen über den Textinhalt erhalten, beginnend mit einer graphischen Darstellung der thematischen Struktur des Textes (Textgraph). Fakten werden zunächst in tabellarischer Form angeboten, können aber auch in ihrem textuellen Kontext betrachtet werden, indem der Benutzer sich die zugehörige Textpassage (Textfragment) zeigen läßt. Bei

Bedarf ist schließlich der Gesamttext zugänglich.

Das TOPIC-System ist in C, TOPOGRAPHIC in C und Prolog unter UNIX implementiert. TOPIC wird von U. Hahn und U. Reimer, TOPOGRAPHIC von R. Hammöhner und U. Thiel unter der Projektleitung von Prof. Dr. R. Kuhlmann entwickelt und auf einem CADMUS 9200 implementiert. Kontakte zu beiden Forschungsgruppen können über die oben angegebene Adresse aufgenommen werden, unter der auch die den Veröffentlichungen entsprechenden technischen Berichte bei Bedarf angefordert werden können.

Literaturhinweise zu TOPIC / TOPOGRAPHIC

Hahn, U. (1984)

Textual Expertise in Word Experts: An Approach to Text Parsing Based on Topic/Comment Monitoring. In: COLING-84; Proc. of the 10th Int. Conf. on Computational Linguistics & the 22nd Annual Meeting of the Association for Computational Linguistics. Stanford/CA, July 2-6, 1984. ACL 1984, pp.402-408. (= Bericht TOPIC-10/84)

Hahn, U. (1986a)

A Generalized Word Expert Model of Lexically Distributed Text Parsing. In: ECAI'86. Proc. of the 7th European Conf. on Artificial Intelligence. Brighton, U.K., July 21-25, 1986. (= Bericht TOPIC-21/86)

Hahn, U. (1986b)

On Lexically Distributed Text Parsing: A Computational Model for the Analysis of Textuality on the Level of Text Coherence and Text Coherence. In: F. Kiefer (ed): Linking in Text. Dordrecht: Reidel, 1986. (= Bericht TOPIC-20/85)

Hahn, U. / U. Reimer (1983)

Wortexperten-Parsing: Text-Parsing mit einer verteilten lexikalischen Grammatik im Rahmen des automatischen Textkondensierungssystems 'TOPIC'. In: Linguistische Berichte 1983, No.88, pp. 56-78. (= Bericht TOPIC-6/83)

Hahn, U. / U. Reimer (1984a)

Heuristic Text Parsing in 'TOPIC': Methodological Issues in a Knowledge-Based Text Condensation System. In: H.J. Dietschmann (ed): Representation and Exchange of Knowledge as a Basis of Information Processes. Amsterdam etc.: North-Holland, 1984, pp.143-163. (= Bericht TOPIC-5/83)

Hahn, U. / U. Reimer (1984b)

Computing Text Constituency: An Algorithmic Approach to the Generation of Text Graphs. In: C.J. van Rijsbergen (ed): Research and Development in Information Retrieval. Cambridge etc.:

Themenschwerpunkt

Cambridge UP., 1984, pp.343-368 (= Bericht TOPIC-8/84)

Hahn, U. / U. Reimer (1985a)

The TOPIC Project: Text-Oriented Procedures for Information Management and Condensation of Expository Texts. Final Report. Konstanz: Univ Konstanz, Informationswissenschaft, Mai 1985, 155pp. (= Bericht TOPIC-17/85)

Hahn, U. / U. Reimer (1985b)

Entwurfsprinzipien und Architektur des Textkondensierungssystems TOPIC. In: B. Endres-Niggemeyer / J. Krause (eds): Sprachverarbeitung in Information und Dokumentation. Berlin etc.: Springer, 1985, pp.165-179. (= Bericht TOPIC-14/85)

Hahn, U. / U. Reimer (1986)

TOPIC Essentials. In: COLING 86: Proc. of the 11th Int. Conf. on Computational Linguistics. August, 25-29, 1986, Bonn, F.R.G. ACL 1986. (= Bericht TOPIC-19/86)

Hammwöhner, R. / U. Thiel (1984)

TOPOGRAPHIC: eine graphisch-interaktive Retrievalschnittstelle. In: G. Enderle et al. (eds): MICROGRAPHICS: Graphik auf Arbeitsplatzrechnern. Bonn: GI, 1984, pp.155-169. (= Bericht TOPOGRAPHIC-2/84)

Hammwöhner, R. / U. Thiel (1985)

Graphisch-interaktive Manipulation komplexer Wissenstrukturen mit dem Framnetz-Editor TOPOGRAPHIC I. In: H.-J. Bullinger (ed): Software-Ergonomie '85. Mensch-Computer-Interaktion. Stuttgart: Teubner, 1985, pp.119-128.

Hammwöhner, R. / U. Thiel (1986a)

Zur Konzeption eines graphischen Interface für ein wissensbasiertes Textkondensierungssystem. In: Kognitive Aspekte der Mensch-Computer-Interaktion. Berlin etc.: Springer, 1986. (= Bericht TOPOGRAPHIC-1/84)

Hammwöhner, R. / U. Thiel (1986b)

Die Behandlung graphischer Seiteneffekte beim Backtracking in Prolog. In: Tagungsband der 16. Jahrestagung der Gesellschaft für Informatik in Berlin, 6.-10. Oktober 1986. Berlin etc.: Springer, 1986. (= Bericht TOPOGRAPHIC-6/86)

Kühlen, R. (1984)

Some Similarities and Differences between Intellectual and Machine Text Understanding for the Purpose of Abstracting. In: H.J. Dietschmann (ed): Representation and Exchange of Knowledge as a Basis of Information Processes. Amsterdam etc.: North-Holland, 1984, pp.87-109. (= Bericht TOPIC-7/83)

Reimer, U. (1986a)

A System-Controlled Multi-Type Specialization Hierarchy. In: L. Kerschberg (ed): Expert Database Systems. Menlo Park/CA: Benjamin/Cummings, 1986, pp.173-187. (= Bericht TOPIC-11/84)

Reimer, U. (1986b)

A Representation Construct for Roles. In: Data and Knowledge Engineering 2. 1986. (= Bericht TOPIC-16/86)

Reimer, U. / U. Hahn (1983)

A Formal Approach to the Semantics of a Frame Data Model. In: IJCAI-83: Proc. of the 8th Int. Joint Conf. on Artificial Intelligence. 8-12 August 1983, Karlsruhe, West Germany. Los Altos/CA: W. Kaufmann, 1983, pp.337-339. (= Bericht TOPIC-3/83)

Reimer, U. / U. Hahn (1985)

On Formal Semantic Properties of a Frame Data Model. In: Computers and Artificial Intelligence 4. 1985, No.4, pp.335-351. (= Bericht TOPIC-13/85)

Thiel, U. / R. Hammwöhner (1985)

TOPOGRAPHIC: Die graphisch interaktive Schnittstelle des Textkondensierungssystems TOPIC. Bericht zur ersten Projektphase. Eggenstein-Leopoldshafen: FIZ Energie Physik Mathematik GmbH Karlsruhe, Dez. 1985, 82pp. (= BMFT-FB-ID 85-006 & Bericht TOPOGRAPHIC-4/85)

Thiel, U. / R. Hammwöhner (1986)

Graphical Interaction with a Full-Text Oriented Information System: The Retrieval Component of the End User Interface TOPOGRAPHIC. In: Proc. of the 2nd Int. Conf. on the Application of Micro-Computers in Information, Documentation and Libraries. Amsterdam etc.: North-Holland, 1986. (= Bericht TOPOGRAPHIC-5/86)

PROJEKTE "XTRA", "VITRA" UND "SC" AN DER UNIVERSITÄT SAARBRÜCKEN

Universität des Saarlandes
Fachbereich Informatik
Im Stadtwald 15
D-6600 Saarbrücken

SFB 314 - TEILPROJEKT NS1:

"XTRA": EIN NATÜRLICHSPRACHLICHES ZUGANGSSYSTEM FÜR EXPERTENSYSTEME

Ziel des Projekts ist die theoretische Fundierung, der Entwurf und die Entwicklung eines natürlichsprachlichen Zugangssystems für Expertensysteme. Dieses System soll dem Benutzer eine einfachere Interaktion mit Expertensystemen gestatten. Unter anderem soll XTRA die folgenden Aufgaben erfüllen:

- Extraktion von Daten, die für das Expertensystem relevant sind, aus der natürlichsprachlichen Eingabe des Benutzers.
- Beantwortung begrifflicher Fragen des Benutzers.
- Benutzerangepasste Aufbereitung und natürlichsprachliche Verbalisierung von Erklärungen, die das Expertensystem liefert.
- Kooperative und kommunikativ adäquate Antwortgenerierung (u.a. Überbeantwortung, Elliptifizierung und Pronominalisierung)
- Natürlichsprachlicher Wissenserwerb für konzeptuelles Wissen.

Der gegenwärtige Anwendungsbereich von XTRA ist der Zugang zu einem sich in Entwicklung befindlichen Expertensystem für die Unterstützung beim Ausfüllen eines Lohnsteuerjahresausgleichsformulars. Innerhalb dieser Domäne werden zur Zeit folgende Probleme näher untersucht:

- a) *Integration von Zeigen und Sprechen*: Dem Benutzer soll die Möglichkeit geboten werden, sich auf Felder oder größere Bereiche des auf dem Bildschirm wiedergegebenen Antragsformulars sowohl natürlichsprachlich als auch mit Hilfe von Zeigeoperationen zu beziehen. Auch Kombinationen sollen gestattet sein.
- b) *Referenzauflösung unter Zuhilfenahme mehrerer Wissensquellen*: Identifikation der vom Benutzer angesprochenen Referenten unter Berücksichtigung der verwendeten Nominalphrasen und Zeigeoperationen, von Kasusrestriktionen sowie von Informationen aus dem Dialoggedächtnis.
- c) *Komplexität von Parsingalgorithmen*: Kontrastive Untersuchungen an bekannten Grammatiktypen bzw. Versuch der Erweiterung und Neudefinition.
- d) *Interaktion von Wissensbasen*: Theoretische Untersuchung der Interaktion multipler Wissensbasen und Realisierung einer zweckmäßigen und effizienten Interaktion der XTRA-Wissensbasis mit der Wissensbasis des Expertensystems.
- e) *Dialogkontext*: Untersuchung der Auswirkungen des Dialogkontexts auf die Antwortplanung.

Themenschwerpunkt

Projektleitung: Prof. Dr. W. Wahlster, Dr. Alfred Kobsa
(0681) 302-2363 (Sekretariat)
Telefon: EUNET: sbsvax! < nachname mitarbeiter >
-Mail: CSNET: < nachname > %sbsvax.uucp@germany.csnet
1.4.85-31.12.87
Projektdauer: Deutsche Forschungsgemeinschaft, Universität des Saarlandes
Ordner Institutionen: Lisp-Maschine Symbolics 3600 (gemeinsame Nutzung mit dem
Technerausstattung: SFB-Projekt VITRA), VAX 11/780 unter UNIX, ETHERNET-
Kopplung zwischen Symbolics und VAX mit TCP/IP
Franzisp, Flavors, Commonlisp, BABYLON
Software-Ausstattung:

MITARBEITER

Jürgen Allgayer (App. 2474)

Studium der Informatik an der Univ. Kaiserslautern (1981 - 86). 1984 Mitarbeit bei der Entwicklung einer Benutzerschnittstelle zu einem Expertensystem im Rahmen des Industrie-Kooperationsprojekts "IXMO: Motorendiagnostik auf Prüfständen in der Serienproduktion" (Univ. Kaiserslautern / INPRO Berlin).

Aufgabenschwerpunkt: Analysekomponente für Zeigeoperationen (Diplomarbeit), Wissensrepräsentation im Expertensystem.

Karin Harbusch, Dipl.-Inform. (App. 2865)

Studium der Informatik und der Mathematik (Schwerpunkt Formale Sprachen) an der Univ. des Saarlandes (1978-1983). Von 1980 bis 1983 neben- bzw. von 1983 bis 1985 hauptberufliche wissenschaftliche Mitarbeiterin im SFB 100 - Elektronische Sprachforschung.

Aufgabenschwerpunkt: Analyse der natürlichsprachlichen Benutzereingabe.

Alfred Kobsa, Dr.techn., Dipl.-Ing. Mag. (App. 2474)

1975-80 Studium der Informatik und Betriebsinformatik an der Univ. Linz. 1980-81 Assistent am dortigen Institut für Informatik. Studien in Kognitiver Psychologie, Linguistik, Logik und Philosophie an der Univ. Salzburg (1980-82). Mitarbeit am natürlichsprachlichen Dialogsystem VIE-LANG des Instituts für Med. Kybernetik u. Artificial Intelligence der Univ. Wien (1982-85). Dissertation über Benutzermodellierung (erschieden bei Springer 1985). Theodor-Körner- und Heinz-Zemanek-Preis 1985.

Aufgabenschwerpunkt: Projektleitung, Wissensrepräsentation, Schnittstelle XTRA - XPS.

Carola Reddig-Siekmann, Dipl.-Inform. (App. 2474)

Studium der Informatik an der TU Berlin (1975-1981). Mitarbeit in den Projekten "Automatische Erstellung semantischer Netze" (1979-1981) und KIT (1981-1982). 1982-84 an der Univ. Osnabrück Mitarbeit im Linguistik-Projekt "Simulation grammatischer Kreativität". 1984 Konstruktion der Benutzerschnittstelle zum XPS im Industrie-Kooperationsprojekt "IXMO: Motorendiagnostik auf Prüfständen in der Serienproduktion" (Univ. Kaiserslautern / INPRO Berlin).

Aufgabenschwerpunkt: Referentielle Analyse komplexer Nominalphrasen; Repräsentation und kommunikative Funktion unterschiedlicher Objektbeschreibungen im Dialog.

Norbert Reithinger, Dipl.-Inform. (App. 2865)

1978 bis 1983 Studium der Informatik an der Univ. Erlangen-Nürnberg mit Hauptfach Mustererkennung (Prof. Dr. H. Niemann). 1980 bis 1983 Mitarbeit am Erlanger Spracherkennungsprojekt auf den Gebieten Syntaxanalyse und Generierung von Sprache. Von 1984 bis Anfang 1985 Angestellter der TA Triumph-Adler AG in Nürnberg in der Hauptabteilung Neue Technologien/Basisentwicklung.

Aufgabenschwerpunkt: Generierung natürlichsprachlicher Ausgabe.

Dagmar Schmauks, Mag.phil. (App. 3496)

Programmierausbildung 1969/70 bei der DFVLR in Porz-Wahn. Lehramtsstudium in München (1973-76), anschließend 8 Jahre Lehrtätigkeit. Von 1981-85 Diplomstudium Philosophie (Schwerpunkt: Sprachphilosophie, Erkenntnistheorie, Geschichte der Philosophie) und Linguistik in Salzburg. Begleitende Studien in Germanistik und Kognitionspsychologie.

Aufgabenschwerpunkt: Linguistisch-philosophische Untersuchung von Deixis.

Wolfgang Wahlster, Prof. Dr.rer.nat., Dipl.-Inform. (App. 2363)

Studium der Informatik und Linguistik an der Univ. Hamburg (Promotion 1981). Mitarbeiter und Projektleiter in den Projekten HAM-RPM und HAM-ANS. 1982 Ruf nach Saarbrücken, ordentlicher Professor für Künstliche Intelligenz und Datenbanksysteme. Koordinator des BMFT-Verbundvorhabens WISBER, Mitherausgeber u.a. der Zeitschriften 'Artificial Intelligence', 'Computational Linguistics', 'Informatik - Forschung und Entwicklung' und der Buchserie 'Symbolic Computation'. 1982-85 Sprecher des Fachausschusses 'Künstliche Intelligenz und Mustererkennung' der GI.

Aufgabenschwerpunkt: Gesamtleitung, Zeigehandlungen, Erklärungskomponente.

Studentische Mitarbeiter

Hans-Peter Beiche

Studium der Informatik an der Univ. Saarbrücken seit 1981 (Schwerpunkt Künstliche Intelligenz und Datenbanksysteme).

Aufgabenschwerpunkt: Entwicklung eines Expertensystems zur Bearbeitung von Lohnsteuerformularen (Diplomarbeit).

Wolfgang Finkler

Studium der Informatik an der Univ. Saarbrücken seit 1982 (Schwerpunkt Künstliche Intelligenz und effiziente Programmierung).

Aufgabenschwerpunkt: Tätigkeiten im Bereich des Expertensystems für die Bearbeitung von Lohnsteuerformularen.

Johannes Reinert

Studium der Informatik an der Univ. Saarbrücken seit 1983 (Schwerpunkt Künstliche Intelligenz, Nebenfach Wirtschaftsinformatik).

Aufgabenschwerpunkt: Tätigkeiten im Bereich des Expertensystems für die Bearbeitung von Lohnsteuerformularen.

Bärbel Ripplinger

Studium der Informatik an der Univ. Saarbrücken seit 1979 (Schwerpunkt Künstliche Intelligenz).

Aufgabenschwerpunkt: Plankontrollierte Benutzereingabe (Diplomarbeit).

VERÖFFENTLICHUNGEN

Allgaver, J. (1986): Eine Graphikkomponente zur Integration von Zeigehandlungen in natürlichsprachliche KI-Systeme (im Erscheinen).

Kobsa, A. (1985): Using Situation Descriptions and Russellian Attitudes for Representing Beliefs and Wants. Proceedings of the 9th International Joint Conference on Artificial Intelligence, Los Angeles, CA, 513-515.

Kobsa, A., J. Allgaver, C. Reddig, N. Reithinger, D. Schmauks, K. Harbusch and W. Wahlster (1986): Combining Deictic Gestures and Natural Language for Referent Identification. Proceedings of the 11th International Conference on Computational Linguistics, Bonn, West Germany (forthcoming).

Schmauks, D. (1986): Formulardeixis und ihre Simulation auf dem Bildschirm. Ein Überblick aus linguistischer Sicht. Memo Nr. 4, SFB 314, Fachbereich Informatik, Universität Saarbrücken.

Wahlster, W. (1986): The Role of Natural Language in Advanced Knowledge-Based Systems. In: H. Winter, ed. Artificial Intelligence and Man-Machine Systems. Berlin etc., Springer (in press).

Wahlster, W. and A. Kobsa (1986): Dialog-Based User Models. Proceedings of the IEEE, Special Issue on Natural Language Processing (in press).

SFB 314 - TEILPROJEKT NS2:

"VITRA": KOPPLUNG BILDVERSTEHENDER UND SPRACHVERSTEHENDER SYSTEME

Das Projekt VITRA (VIsual TRANslator) beschäftigt sich mit den Grundfragen der Beziehung zwischen Sprache und Sehen. Experimentelle Studien werden im Bereich der Kopplung von bildverstehenden und sprachverstehenden Systemen mit dem Ziel der Entwicklung eines Systems zur natürlichsprachlichen Beschreibung von Bildfolgen durchgeführt. Dies erscheint auch sinnvoll im Hinblick auf die Erleichterung des Verstehens der Ausgabe von bildverstehenden Systemen für menschliche Benutzer.

Zwei Domänen wurden für die Arbeit in VITRA ausgewählt. Bei der ersten handelt es sich um eine fiktive Stadtrundfahrt durch einen interessanten Teil einer bestimmten Stadt. Ein Programm namens CITYTOUR, das den Führer simuliert, beantwortet Fragen über die räumlichen Relationen zwischen (statischen oder bewegten) Objekten und anderen (statischen) Objekten in der Szene. Der Forschungsschwerpunkt in die

themenschwerpunkt

ser Domäne liegt auf der Semantik von Pfadpräpositionen (wie "entlang" oder "vorbei") und ihrem Gebrauch für die Beschreibung von Trajektorien bewegter Objekte sowie auf der Abhängigkeit der räumlichen Relationen von den wechselnden Beobachterstandorten (außerhalb der Szene oder an verschiedenen Standorten innerhalb der Szene). In diesem Zusammenhang ist auch die Verwendung von deiktischen Ausdrücken von Interesse.

Bei der zweiten Domäne handelt es sich um Fußballspiele. Die Eingabe für das in der Entwicklung befindliche System wird aus symbolischen Beschreibungen der Orte aller beweglichen Objekte in der Szene zu aufeinanderfolgenden Zeitpunkten bestehen. Diese symbolischen Beschreibungen werden von dem System MORIO, das in Zusammenarbeit mit VITRA von der FhG in Karlsruhe entwickelt wird, aus den Bilddaten extrahiert werden. Zur Zeit wird in VITRA auch an einem Trajektorien-Editor gearbeitet, mit dem die Eingabedaten per Hand erzeugt werden können.

Das Ziel ist es, natürlichsprachliche Beschreibungen zu Bildfolgen von einigen Minuten Länge zu erzeugen. Diese Beschreibungen sollen den Kommentaren von Fußballreportern im Radio in gewissen Aspekten ähneln (Simultanreport; Beschreibung für Hörer, die die Szene selbst nicht sehen), sich von diesen aber in anderen Aspekten unterscheiden (z.B. objektive Beschreibungen ohne Emotionen; keine Metaphern und Übertreibungen; grammatische Sätze). Um die Tiefe des Verstehens testen zu können, sollte auch ein Dialog mit dem System über das Spiel möglich sein können.

Forschungsschwerpunkte in dieser Domäne sind:

- die algorithmische Formulierung und Verfeinerung psycholinguistischer Modelle der Koordination von Sehen und Sprechen
- die Verwendung von Erwartungen für die Beschreibung (im Präsens) von Ereignissen, die noch nicht abgeschlossen sind
- die Erkennung höherer visueller Bewegungskonzepte, gestaffelt nach der Anzahl der beteiligten Akteure und nach der Länge der betrachteten Zeiträume
- die Interpretation von Handlungen, d.h. die Erkennung von nicht-visuellen Konzepten wie Intentionen und Plänen
- Fokussierung im Hinblick auf Gedächtnis und Verbalisierung
- die Generierung von Nominalphrasen, insbesondere die Beschreibung von bewegten Objekten in Beziehung zu anderen bewegten Objekten oder mithilfe von früheren Ereignissen
- die Verwendung von deiktischen und anaphorischen Ausdrücken
- Partner-Modellierung durch Re-Repräsentation (Visualisierung) der generierten Beschreibung.

Projektleitung :

Prof. Dr. Wolfgang Wahlster
Dipl.-Inform. Gudula Retz-Schmidt

Schrift :

Fachbereich Informatik
Universität des Saarlandes
D - 6600 Saarbrücken 11

Telefon :

0681-302-2363
0681-302-2016

E-Mail :

wahlster@sbsvax.uucp@germany.csnet
retz-schmidt@sbsvax.uucp@germany.csnet

Internet :

sbsvax!wahlster
sbsvax!retz-schmidt

Projektdauer :

1. Januar 1985 - 31. Dezember 1987

- Fördernde Institutionen* : DFG (Sonderforschungsbereich 314 : Künstliche Intelligenz - Wissensbasierte Systeme)
Universität des Saarlandes
- Projektpartner* : Fraunhofer-Institut für Informations- und Datenverarbeitung
Fakultät für Informatik der Universität Karlsruhe
- Hardware-Ausstattung* : LISP-Maschine Symbolics 3600 (gemeinsame Nutzung mit Projekt NTRA im SFB 314)
COLOR-OPTION OP36-C108 für LISP-Maschine
VAX 11/780 unter UNIX
ETHERNET-Kopplung zwischen Symbolics und VAX mit TCP/IP
- Software-Ausstattung* : CommonLISP
FUZZY
Flavors

MITARBEITER

Gudula Retz-Schmidt, Dipl.-Inform. (Tel. 302-2016)

Studium der Informatik und Linguistik an der Universität Hamburg von 1978 bis 1985. Zwischendurch Auslandsstudium an der School of Epistemics der University of Edinburgh (1982/83). Diplomarbeit im Projekt NAOS an der Universität Hamburg. Mitarbeiterin im Projekt VITRA seit Oktober 1985.

Aufgabenschwerpunkte : Projektleitung; Erkennung und Verbalisierung von Intentionen und Plänen.

Wolfgang Wahlster, Prof. Dr. (Tel. 302-2363)

Studium der Informatik und Linguistik an der Universität Hamburg. 1977 Diplom. 1981 Promotion. Mitarbeiter und später Projektleiter in den Projekten HAM-RPM und HAM-ANS. Seit 1982 ordentlicher Professor für Künstliche Intelligenz und Datenbanksysteme an der Universität des Saarlandes.

Aufgabenschwerpunkte : Gesamtleitung; Partnermodellierung.

Studentische Mitarbeiter

Elisabeth Andre, cand. inform.

Studium der Informatik an der Universität des Saarlandes seit 1981 (Schwerpunkt: Künstliche Intelligenz).

Aufgabenschwerpunkte : Generierung natürlicher Sprache; komplexe Nominalphrasen für die Beschreibung bewegter Objekte.

Guido Bosch, cand. inform.

Studium der Informatik an der Universität des Saarlandes seit 1981 (Schwerpunkt: Künstliche Intelligenz).

Aufgabenschwerpunkte : CITYTOUR; Parsing von räumlichen Beschreibungen; Semantik von Pädpräpositionen.

Gerd Herzog, cand. inform.

Studium der Informatik an der Universität des Saarlandes seit 1981 (Schwerpunkt: Künstliche Intelligenz).

Aufgabenschwerpunkte : Wissensrepräsentation; Ereigniserkennung; Fokussierung.

Thomas Rist, cand. inform.

Studium der Informatik an der Universität des Saarlandes seit 1981 (Schwerpunkt: Künstliche Intelligenz).

Aufgabenschwerpunkte : Wissensrepräsentation; Ereigniserkennung; Fokussierung.

VERÖFFENTLICHUNGEN

Andre, E.; Bosch, G.; Herzog, G.; Rist, T. (1985): CITYTOUR. Ein natürlichsprachliches Anfragesystem zur Evaluierung räumlicher Präpositionen. Fortgeschrittenenpraktikum Prof. Dr. W. Wahlster, Wintersemester 1984/85. Fachbereich Informatik. Universität des Saarlandes.

Andre, E.; Bosch, G.; Herzog, G.; Rist, T. (1986): Characterizing Trajectories of Moving Objects Using Natural Language Path Descriptions. Memo Nr. 5. Fachbereich Informatik. Universität des Saarlandes.

Retz-Schmidt, G. (1985): Script-Based Generation and Evaluation of Expecta-

Themenswerpunkt

Communications in Traffic Scenes. Mitteilung 136. Fachbereich Informatik. Universität Hamburg. Erscheint auch in: H. Stoyan (Hrsg.) (1986): GWAI-35, 9. Fachtagung über Künstliche Intelligenz, Informatik-Fachberichte, Berlin/Heidelberg/New York/Tokyo: Springer.

"SC": EIN AKTIVES, WISSENSBASIERTES HILFESYSTEM FÜR SINIX

Ziel des Projektes SC (SINIX Consultant) ist die Entwicklung eines intelligenten Hilfesystems für das Betriebssystem SINIX, das sowohl in einem passiven Modus als auch natürlichsprachliches Hilfesystem arbeitet, als auch aktiv beratend und unterstützend tätig wird. SC soll damit zum einen Anfänger und Laien bei der Einarbeitung in die Benutzung des SINIX-Systems unterstützen, zum anderen aber auch fortgeschrittenen Benutzern gezielte Informationen bieten können.

Die Anforderungen an SC können durch folgende Charakteristika umrissen werden:

Beantwortung natürlichsprachlicher Anfragen

- begriffliche Fragen

Hierunter fallen Fragen zur Definition von Konzepten, Objekten und Kommandos des SINIX-Systems, z.B. 'Was ist eine Pipe?' oder 'Wie funktioniert das mail-Kommando?'

- anwendungsbezogene Fragen

Diese Fragen beinhalten Umschreibungen von Aktionen oder Zielzuständen, z.B. 'Wie kann ich eine Datei in ein anderes Dateiverzeichnis kopieren?' oder 'Wie verschicke ich geheime Post?'. Zur Beantwortung derartiger Fragen ist ein komplexer Problemlösungsprozeß auf der Basis einer semantischen Repräsentation der Kommandos erforderlich.

Dialogfähigkeit

SC soll nicht nur in der Lage sein, einzelne Fragen zu beantworten, sondern es soll einen zusammenhängenden kooperativen Dialog mit dem Benutzer führen.

Aktive Hilfe

Über die Beantwortung natürlichsprachlicher Anfragen hinaus soll das System aktiv, ohne explizite Aufforderung durch den Benutzer Ratschläge erteilen. Als Hilfestellung kann der Benutzer auf Kommandos zur Vervollständigung seines Plans und/oder auf einen effizienteren Plan zur Erreichung seines Ziels aufmerksam gemacht werden.

Benutzerangepasstheit

Das Verhalten des Systems bzgl. Dialogfähigkeit und aktiver Hilfe soll sich am Wissensstand und den Zielen des Benutzers orientieren.

Grundlegend für die Durchführung der o.g. Aufgaben ist die Integration verschiedener Wissensquellen in das System :

- o Wesentlich für die Funktion des System ist die Repräsentation des *Domänen-Wissens*, die strukturelle Beschreibungen von SINIX-Konzepten, -Kommandos und -Objekten sowie Planschemata für die Lösung komplexer Aufgaben beinhaltet. Darüberhinaus sollen Semantik-Definitionen für Kommandos und Pläne auf der Basis eines 'deep modelling' als Zustandsänderungen von Objekten spezifiziert werden. Um einen Anschluß an die sprachlich-kommunikative Ebene zu erreichen, sollen konzeptuelle Vorstellungen des Benutzers vom SINIX-System in der Modellierung berücksichtigt werden.

- o *Wissen über den Benutzer* (Ziele, Pläne, Kenntnisse) wird unter verschiedenen Aspekten in SC verwendet, z.B. für die Planerkennung und die Antwortformulierung. Das

Benutzermodell besteht zum einen aus a-priori-Informationen über prototypische Benutzerklassen, zum anderen aus aktuellen Annahmen über den individuellen Benutzer, die aus der Kommunikation des Benutzers mit SC und SINIX extrahiert werden.

- Weiterhin notwendig ist *Wissen über Planungsprozesse*, das zur Planerkennung und Plangenerierung eingesetzt wird.

Die vorläufige Konzeption der Systemarchitektur sieht folgende Komponenten vor :

- ein *robuster Parser*, der Anfragen analysiert im Hinblick auf den Fragetyp und verwendete Schlüsselbegriffe.
- eine *Antwortformulierungskomponente*, die ein dem Fragetyp und dem Benutzer angemessenes Antwortmuster auswählt und mit entsprechenden Inhalten der Wissensbasis füllt.
- eine *Dialogsteuerung*, die die Analyse- und die Generierungskomponente kontrolliert und Wissen über den jeweiligen Dialogzustand bereithält.
- eine *Problemlösungskomponente*, die das zur Beantwortung von Fragen notwendige SINIX-Wissen bereitstellt.
- eine *Planerkennungskomponente*, die den momentanen Plan des Benutzers erkennt und das damit verbundene Ziel ggf. inferiert.

Zu einem späteren Zeitpunkt soll das System um eine *Plangenerierungskomponente* erweitert werden, die selbständig komplexe Pläne auf der Basis einer Semantik-Definition der primitiven Kommandos erzeugt.

Projektleitung :	Prof. Dr. W. Wahlster
Anschrift :	Projekt SINIX-Consultant FR 10.2 Informatik IV Universität des Saarlandes Im Stadtwald 15 D-6600 Saarbrücken
Telefon :	(0681) 302-3393 oder 2363 (Sekretariat)
CSnet :	hecking% sbsvax.uucp@Germany.csnet kemke% sbsvax.uucp@Germany.csnet
Projektdauer :	01.05.1985 - 31.12.1987
Fördernde Institution :	Siemens AG Universität des Saarlandes
Rechnerausstattung :	1 Siemens APS 5815 (XEROX 1108-105) 1 Siemens PC-MX
Software-Werkzeuge :	Interlisp-D, LOOPS

MITARBEITER

Matthias Hecking, Dipl.-Inform. (DW 3393)

Von 1979 bis 1984 Studium der Informatik und Linguistik an der Universität Hamburg mit den Schwerpunkten: Software Engineering und Künstliche Intelligenz. Von 1984 bis 1985 Mitarbeiter bei InterFace Computer GmbH in München.

Aufgabenschwerpunkte : Planerkennung, Plangenerierung, natürlichsprachliche Schnittstelle, Wissensrepräsentation

Christel Kemke, Dipl.-Inform. (DW 3393)

Von 1977 bis 1985 Studium der Informatik und der Theoretischen Medizin an der Universität Dortmund und der Ruhruniversität Bochum mit den Schwerpunkten : Modellierung und Simulation, Künstliche Intelligenz und Neurophysiologie.

Aufgabenschwerpunkte : Wissensrepräsentation, Plangenerierung, Gesamtkonzeption

Themenschwerpunkt

Studentische Mitarbeiter

Bernd Jung, cand.inform.

Seit 1980 Studium der Informatik mit Nebenfach Wirtschaftswissenschaften an der Universität des Saarlandes (Schwerpunkte: KI und Datenbanksysteme). 1983 - 1985 freier Mitarbeiter bei IBM Stuttgart im Projekt 'Entwicklung interaktiver Lernprogramme'.

Aufgabenschwerpunkte : Aufbau der SINIX-Wissensbasis, SINIX-Experte

Peter Neisser, cand.inform.

Seit 1979 Studium der Informatik mit Nebenfach Informationswissenschaften an der Universität des Saarlandes (Schwerpunkte: KI, insbesondere natürlichsprachliche Systeme).

Aufgabenschwerpunkte : Entwicklung eines robusten Parsers

Erich Nessen, cand.inform.

Seit 1981 Studium der Informatik mit Nebenfach Wirtschaftswissenschaften an der Universität des Saarlandes (Schwerpunkte: KI, Datenbanksysteme, Parallele Algorithmen).

Seit 1984 freier Mitarbeiter bei IBM Stuttgart im Projekt INTEGRAS (Integriertes Text- und Graphik-Anwendungssystem).

Aufgabenschwerpunkte : Benutzermodellierung

VERÖFFENTLICHUNGEN

C.Kemke (1985) : *Entwurf eines aktiven, wissensbasierten Hilfesystems für SINIX*, Bericht Nr.5, FR.10.2 Informatik IV, Universität des Saarlandes, Saarbrücken

C.Kemke (1985) : *SC - Ein intelligentes Hilfesystem für SINIX*, LDV-Forum Nr.2 Dezember 1985, S.43-60



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

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

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

Fachbeitrag

ZUM PROPOSITIONALEN UND TEXTUELLEN WISSEN FÜR DAS VERSTEHEN FRANZÖSISCHER WETTERBERICHTE

Dieter Seelbach

Institut für Allgemeine und Vergleichende Sprachwissenschaft
Universität Mainz
Welderweg 18
D-6500 Mainz

Es werden **Voruntersuchungen** zum (fach-)sprachlichen Wissen vorgestellt, das für das Verständnis von französischen Wetterberichten aus "Le Monde" gebraucht wird. Genauer untersucht wird die Zuordnung von **semantischen Funktionen** innerhalb der Propositionen sowie die Zuordnung von **Sinnrelationen** assoziativ-anaphorischen **Relationen**. Die Zuweisung der semantischen Funktionen wird im Sinne der "Lexical Functional Grammar" **über die grammatischen Funktionen** wie Subjekt, Objekt usw. und die Zuordnung der **Prädikatsklassen** geleistet. Die grammatischen Funktionen erhält man über eine "Finite Clause Grammar". Mit den vorgelegten linguistischen Ergebnissen werden zugleich Vorarbeiten für ein Übersetzungssystem geliefert.

Ziel des Beitrages ist die linguistische Beschreibung einer Fachsprache, die als Grundlage für eine maschinelle Analyse dienen könnte. Die theoretische Basis für diese Beschreibung liefern strukturalistische, generativ-linguistische und textlinguistische Ansätze. Im wesentlichen sind es **drei Pfeiler**, auf denen die Beschreibung des propositionalen Wissens aufbaut: **die syntaxorientierte Lexikalische Semantik** im Sinne etwa von **Apresjan, Gross und Grelmas**, Elemente der "Lexical Functional Grammar" (**Bresnan**) und die "Metamorphosis Grammar" (**Colmerauer**). Die textlinguistischen Aspekte der Beschreibung gehen auf **van Dijk** und **Fradin** zurück.

1 Propositionales Wissen

1.1 Prädikatsklassen

Arbeiten zur Syntax und Semantik französischer Verben im distributionellen und transformationellen Rahmen (z. B. **Gross 1975, Ruwet 1972, Seelbach 1981**) haben zusammen mit der Aus-

wertung der in etwa 100 Wetterberichten aus "Le Monde" (Météorologie Nationale) enthaltenen **ben und prädikativen Adjektive** einige für die Texttyp charakteristische (syntaktisch und) semantisch homogene Klassen ergeben. Verben bilden auch unabhängig von Fachgebieten - dann syntaktisch und semantisch homogene Klassen, wenn eine größere Menge von distributionellen und transformationellen Eigenschaften gemeinsam haben. Darüber hinaus noch semantische Gemeinsamkeiten in Form von (komplexen) semantischen Merkmalen aufweisen. Wenn sie z. B., zusätzlich zu einer ähnlichen syntaktischen Verhalten, eine **Bewegung**, eine **psychologische Reaktion**, **Kommunikation**, eine **sinnliche Wahrnehmung** etc. ausdrücken, muß betont werden, daß derartige Klassen und Prädikatsklassen zunächst rein formal definiert sind, d. h. auf der Basis von rein formalen Eigenschaften vorab definiert sind. Dies sei an einem Beispiel erläutere.

Nur eine kleine Menge von Verben passen in die Lücke von (1) sowie in die Position **V** in jeder angegebenen distributionellen Satzkonstruktion:

(1) Les vents d'ouest une zone de pluie des côtes atlantiques vers l'intérieur du continent.

N_0 V N_1 de N_2 Prep N_3

N_1 être V-é de N_2 Prep N_3 par N_0

N_0 V N_1 Prep N_3

N_0 y V N_1

De N_2 N_1 être V-é Prep N_3

De N_2 N_0 + pro V N_1 Prep N_3

N_0 V Prep N_3 N_1

(N_0 , N_1 , N_2 und N_3 sind mehr oder weniger komplexe Nominalgruppen)

Es sind Verben wie **pousser, envoyer, diriger** etc. Schon eine oberflächliche Analyse der Verbbe-
 deutung ergibt, daß hier eine semantische Gemein-
 samkeit in Form eines "Klassesms" (**Grelmas**), ei-
 nes allen Elementen der Prädikatsklasse gemein-
 samen (komplexen) semantischen Merkmals, vorliegt.
 Es handelt sich um Verben, die die **Verursachung**
inner Bewegung ausdrücken, die **kausativen Be-**
wegungsverben (verbes causatifs de mouvement).
 N₁ wird von N₀ zu einer Ortsveränderung von N₂
 nach N₃ veranlaßt.)

Wir haben 17 derartige semantisch (und weit-
 gehend auch syntaktisch) homogene Prädikatsklas-
 sen in den Wetterberichten ermittelt (vgl. Anhang
 I). Ausgedrückt werden durch:

1. **ZIEL-BEWEGUNGSVERBEN**
 die Richtung oder das Ziel einer Orts-
 veränderung: wo sich etwas hinbewegt (*wo-*
hin?):
 a) *entrer, se diriger, b) atteindre, gagner*
2. **URSPRUNG-BEWEGUNGSVERBEN**
 der Ausgangspunkt oder die Richtung, aus
 der etwas kommt (*woher?*):
 a) *s'absenter, s'éloigner, b) quitter, aban-*
donner, c) tourner, s'orienter (bei Windbe-
 wegungen)
3. **URSPRUNG-ZIEL-BEWEGUNGSVERBEN**
 der Ausgangspunkt und der Zielpunkt einer
 Ortsveränderung (*woher? wohin?*):
 a) *se déplacer, arriver, b) progresser, se*
décaler
4. **SZENE-URSPRUNG-ZIEL-BEWEGUNGS-**
VERBEN
 der Ausgangspunkt, das Ziel und die Szene,
 wo sich die Ortsveränderung abspielt (*wo-*
her? wohin? wo?):
traverser, parcourir
5. **KAUSATIVE BEWEGUNGSVERBEN**
 (verbes causatifs de mouvement)
 die Verursachung einer Ortsveränderung:
 a) *diriger, apporter, b) être poussé, être en-*
voyé
6. **SZENE-BEWEGUNGSVERBEN**
 (mouvement sur place)
 die Bewegung innerhalb einer Szene oder in-
 nerhalb eines Gebietes:
circuler, onduler
7. **POSITIONSVERBEN**
 die Positionierung im geografischen Raum:
se trouver, être situé
8. **SZENE-PROZESSVERBEN**
 durative, inchoative und terminative Pro-
 zesse, die an eine Szene gebunden sind:
Maintien: *se maintenir, persister*
Appartition: *se former, s'installer*
Dispartition: *se disloquer, se dissiper*
9. **SZENE-VERÄNDERUNGSVERBEN** (Chan-
 gement)
 Prozesse, die eine qualitative oder quantita-
 tive Veränderung ausdrücken:
se renforcer, se faire rare, devenir beau

10. **KAUSATIVE VERÄNDERUNGSVERBEN**
 (verbes causatifs de changement)
 die Verursachung einer Veränderung:
 a) *réduire, ralentir, affaiblir, b) être renforcé,*
être augmenté

11. **ZUSTANDSVERBEN**
 Zustände:
être stationnaire, prédominer, rester/être
fréquent

12. **PSYCHOLOGISCHE VERBEN**
 die Auslösung einer psychologischen Reak-
 tion bei den Einwohnern einer Region, die Be-
 einflussung durch eine meteorologische Er-
 scheinung:
 a) *affecter, intéresser, b) subir, c) profiter,*
 d) *échapper, e) être affecté*

13. **FESTSTELLUNGS- UND WAHRNEH-**
MUNGSVERBEN
 Feststellung und Wahrnehmung:
 a) *voir, observer, remarquer, b) être observé,*
être remarqué, c) assister

14. **ASPEKT- UND MODALITÄTSVERBEN**
 Aspektualität und Modalität:
commencer, risquer, pouvoir

15. **METEOROLOGISCHE VERBEN**
 meteorologische Erscheinungen:
 a) *pleuvoir, neiger, geler, b) souffler (vent),*
briller (soleil), se couvrir (ciel)

16. **URSACHE-WIRKUNGS-VERBEN**
 kausale Relationen:
 a) *provoquer, entraîner, favoriser, b) être*
provoqué, être favorisé

17. **TEMPORALE RELATIONSVERBEN**
 temporale Relationen:
 a) *précéder, suivre, b) alterner, se succéder,*
 c) *faire place, se substituer*

Dazu zwei Textpassagen:
La zone dépressionnaire centrée sur le sud
de la Scandinavie dirige sur la France un
flux faible de nord-ouest, très frais et hu-
mide. Une nouvelle perturbation abordera
samedi soir les régions du nord-ouest du
pays. (03.05.85)

centrer: Positionierung; **diriger:** Kausative
 Bewegung; **aborder:** Ziel-Bewegung

Lente dégradation du temps sur
la France. Les hautes pressions de l'Europe
s'affaiblissent et se décalent vers l'est. Les
perturbations qui circulaient sur l'océan et
les îles Britanniques vont pénétrer sur notre
pays. (19.04.85)

dégradation: Szene-Veränderung; **s'affaiblir:**
 Szene-Veränderung; **se décaler:** Ursprung-Ziel-
 Bewegung; **circuler:** Szene-Bewegung; **pénétrer:**
 Ziel-Bewegung

1.2 Zuordnung von semantischen Funktionen über Prädikatsklassen und grammatische Funktionen

In der von **Bresnan, Kaplan, Rappaport** u. a. entwickelten "Lexical Functional Grammar" (LFG) ist von "semantischen Prädikat-Argument-Relationen" die Rede, die auch unter den Namen Tiefenkasus(relationen), "participant roles", Argumentrollen oder Kasusrollen in der linguistischen Literatur bekannt wurden und in erster Linie in der Kasusgrammatik **Fillmores** sowie in der "Systemic Functional Grammar" von **Halliday, Fawcett** u. a. untersucht wurden. Ziel der LFG ist - vereinfacht gesagt - die Zuordnung dieses Relationstyps, den wir im folgenden **semantische Funktionen** nennen, zu den Konstituenten und Kategorien der Oberflächenstruktur. In den Worten von **Kaplan/Bresnan 1982, p. 174**, geht es darum "to characterize the mapping between semantic predicate-argument relations and surface word and phrase configurations by which they are expressed". Diese Zuordnung von semantischen Funktionen und Oberflächenstruktur wird über die **grammatischen Funktionen** wie Subjekt, Objekt etc. und über ein Lexikon geleistet, dessen Einträge u. a. auch die Zuordnung von semantischen Funktionen wie Agent, Patient zu den grammatischen Funktionen spezifizieren. Das grammatische Subjekt kann z. B. in Abhängigkeit vom Verb(typ) unterschiedliche semantische Funktionen haben, wie die folgenden Beispiele zeigen:

- (2) Cet article fourmille de fautes
LOCATION
- (3) La fumée agace tout le monde
PHENOMENON oder THEME
- (4) Cette décision a provoqué son échec
AGENT
- (5) Ce livre a été brûlé
PATIENT
- (6) Tout le monde déteste la fumée
EXPERIENCER

Syntaxregeln (angereicherte kontextfreie PS-Regeln) schließlich ordnen den Konstituenten der Oberflächenstruktur die grammatischen Funktionen zu. Diese ergeben sich - grob - aus Konstituentenstrukturkonfigurationen wie "NP vor dem Verb" (für das Subjekt) in Sprachen wie Englisch und Französisch oder aus morphologischen Indikatoren, z. B. Verbkongruenz und Kasusmarkierungen, wie Nominativ (für das Subjekt) im Deutschen, sowie ebenfalls aus den Lexikoninformationen für jedes einzelne Prädikat.

Das Ergebnis einer Analyse im LFG-Rahmen sei an (1a) angedeutet. Über die Syntaxregeln ergeben sich folgende Konstituenten und grammatische Funktionen:

(1a) Les vents d'ouest poussent une zone de pluie des côtes atlantiques vers l'intérieur du pays

NP. SUBJECT V NP. DIR-OBJ PP. DE-OBJ PP.VERS-OBJ

Über die im Lexikoneintrag spezifizierte Zuordnung von grammatischen und semantischen Funktionen bei *pousser*

pousser: [Agent Patient Source Goal
 SUBJ DIR/OBJ DE-OBJ PREP-OBJ]

(PREP = à, vers, sur, dans, jusqu'à, à travers)

können die semantischen Funktionen den Konstituenten zugewiesen werden.

Nun zu den semantischen Repräsentanten der Propositionen der Wetterberichte. Auf Grundlage der über die rein formalen Eigenschaften der Verben gefundenen Prädikatsklassen im Sinne der LFG die Location-, Source-, Goal-, Agent-, Patient-, Experienter/Location- und Phenomenon-Funktion, die zwischen Subjekt und Nomen bzw. Nominalgruppen bestehen, und zugleich die Verben selbst semantisch charakterisiert werden. Eine wesentliche Rolle spielen hier die (syntaktischen) Subjekt- und unterschiedlichen Objektfunktionen, insbesondere die unterschiedlichen präpositionalen Objekte. Die Präpositionen *de* bzw. *vers* etc. sind dabei spielsweise bei **Bewegungsverben** als Indikatoren für die Source- bzw. Goal-Funktion zu gebrauchen oder anders ausgedrückt, die semantischen Funktionen **Source** bzw. **Goal** werden mit Hilfe der grammatischen Funktionen des präpositionalen Objektes mit *de* bzw. *vers* etc. bei **Bewegungsverben** realisiert.

Einige der Realisierungen der sieben erwähnten semantischen durch unterschiedliche grammatische Funktionen in Abhängigkeit von den Prädikatsklassen seien im folgenden angedeutet und mit selbsterläuternden Beispielen erläutert.

I. Die **LOCATION-Funktion** wird z. B. folgendermaßen realisiert (vgl. Anhang 2):

- 1. **Direktes Objekt** von **Bewegungsverben** vom Typ **Szene-Ursprung-Ziel**
- 2. **à, sur, dans-Objekt** von **Szene-Bewegungsverben** (mouvement sur place), **Positionalverben**, **Veränderungsverben**, **Zustandsverben**

II. Die **SOURCE-Funktion** wird realisiert durch

- 1. **De-Objekt** von **Bewegungsverben** der Subklassen **Ursprung, Ursprung-Ziel, Szene-Ursprung-Ziel** sowie von **kausativen Bewegungsverben**
- 2. **Direktes Objekt** von **Ursprung-Bewegungsverben** vom Typ b)
- 3. **A-Objekt** von **Ursprung-Bewegungsverben** vom Typ c)

III. Die **GOAL-Funktion** wird realisiert durch

- 1. **à-, sur-, dans-, vers-, Jusqu'à-, à travers-Objekt** von **Bewegungsverben** der Subklassen **Ziel, Ursprung-Ziel, Szene-Ursprung-Ziel**, sowie von **kausativen Bewegungsverben**
- 2. **Direktes Objekt** von **Ziel-Bewegungsverben** vom Typ b)

IV. Die **AGENT-Funktion** wird realisiert durch

- 1. **Subjekt** von **Bewegungsverben** (sämtliche Subklassen)
- 2. **Subjekt** von **Ursache-Wirkung-Verben**
- 3. **Subjekt** von **kausativen Veränderungsverben**
- 4. **Par-Objekt** von **kausativen Bewegungsverben** und **kausativen Veränderungsverben**

VI. Die **PATIENT-Funktion** wird realisiert durch
Direktes Objekt von **kausativen Bewegungs-Verben**

VI. Die **EXPERIENCER/LOCATION-Funktion** wird realisiert durch

1. **Direktes Objekt** von **psychologischen Verben** vom Typ a)
2. **Subjekt** von **psychologischen Verben** vom Typ b), c), d) und e)

VII. Die **PHENOMENON-Funktion** wird realisiert durch

1. **Subjekt** von **Position-, Veränderungs- und Zustandsverben**
2. **Direktes Objekt** von **Ursache-Wirkungs-Verben**
3. **Subjekt** von **psychologischen Verben** vom Typ a)
4. **Unterschiedliche Objektfunktionen** von **psychologischen Verben** vom Typ b), c), d) und e).

Beispiele:

I. LOCATION

1. *Un front orageux traversera les régions méditerranéennes*

2. *L'anticyclone* { circule
se trouve
s'affaiblit
est stationnaire } *sur les îles Britanniques*

II. SOURCE

1. *L'anticyclone* { s'éloigne du midi
monte du midi aux régions du centre }

*Une zone de pluie va parcourir les régions méridionales des Pyrénées jusqu'à la frontière italienne
 La dépression des îles Britanniques dirige une perturbation de l'Atlantique sur les côtes vendéennes*

2. *La zone pluvieuse quittera la Méditerranée*
3. *Le vent tourne à l'ouest*

III. GOAL

1. *Une dépression* { se dirige sur la France
se décale des îles Britanniques vers le Danemark
apportera des nuages et des averses dans nos régions }

Une zone pluvieuse traverse le pays du nord-ouest au sud-est

2. *La zone pluvieuse* { atteindra
gagnera } *le Midi dans la soirée*

IV. AGENT

1. *L'anticyclone se dirige vers la France*
2. *L'anticyclone de l'Irlande du Nord dirige des perturbations vers le pays*
3. *La hausse du champ de pression atténue l'activité du mauvais temps*

V. PATIENT

Le flux du nord pousse les averses vers l'intérieur

VI. EXPERIENCER/LOCATION

1. *La pluie* { concernera
affectera
intéressera
menacera } *les régions du nord*

2. *Les régions du nord* { subissent
restent sous l'influence de
échappent à
sont concernées par } *la pluie*



VII. PHENOMENON

- 1. Les nuages $\left\{ \begin{array}{l} \text{se situent} \\ \text{diminuent} \\ \text{sont faibles} \end{array} \right\}$ sur la côte atlantique
- 2. Une zone dépressionnaire $\left\{ \begin{array}{l} \text{entraînera} \\ \text{provoquera} \\ \text{donnera} \end{array} \right\}$ un temps plus variable
- 3. La pluie $\left\{ \begin{array}{l} \text{concernera} \\ \text{affectera} \\ \text{intéressera} \end{array} \right\}$ les régions du nord
- 4. Les régions du nord $\left\{ \begin{array}{l} \text{subissent} \\ \text{restent sous l'influence de} \\ \text{échappent à} \\ \text{sont concernées par} \end{array} \right\}$ la pluie

Dieselben Zuordnungen sind ausgehend von den grammatischen Funktionen folgendermaßen zu formulieren:

Das **DIREKTE OBJEKT**

- von **Szene-Ursprung-Ziel-Bewegungsverben** drückt die **Location-Funktion** aus: (*traverser*)
- von **Ursprung-Bewegungsverben vom Typ b)** drückt die **Source-Funktion** aus: (*quitter*)
- von **Ziel-Bewegungsverben vom Typ b)** drückt die **Goal-Funktion** aus: (*gagner*)
- von **kausativen Bewegungsverben** drückt die **Patient-Funktion** aus: (*diriger*)
- von **psychologischen Verben vom Typ a)** drückt die **Experlencer/Location-Funktion** aus: (*affecter*) etc.

Das **SUBJEKT**

- von **sämtlichen Subklassen der Bewegungsverben** sowie von **Ursache-Wirkungsverben** und **kausativen Veränderungsverben** drückt die **Agent-Funktion** aus: (*monter, provoquer, renforcer*)
- von **psychologischen Verben vom Typ b), c), d) und e)** drückt die **Experlencer/Location-Funktion** aus: (*subir, souffrir, être soumis, être concerné*)
- von **psychologischen Verben vom Typ a)** sowie von **Positlions-, Veränderungs-, und Zustandsverben** drückt die **Phenomenon-Funktion** aus: (*affecter, se situer, s'affaiblir, prédominer*) etc.

Das **A-OBJEKT**... das **DE-OBJEKT**... das **PAR-OBJEKT**... etc.

1.3 Zuordnung der grammatischen Funktionen über eine "Metamorphosis Grammar" oder "Definite Clause Grammar"

Nun zur Frage der Syntaxregeln, mit deren Hilfe die grammatischen Funktionen zugewiesen werden. Im LFG-Rahmen wird versucht, über (angereicherte) kontextfreie PS-Regeln die grammatischen Funktionen aus den Konstituentenstrukturkonfigurationen "herauszufiltern". Dabei wird ausgehend von einer K-Struktur (mit Kategorien als Knoten), die mit aus Lexikoneinträgen stammenden grammatischen Informationen angereichert ist, über ein komplexes Gleichungssystem eine F-Struktur aufgebaut.

Über die (angereicherten) kontextfreien PS-Regeln entstehen bei der Analyse mehrere Bäume (Übergenerierung), erst beim Aufbau und der

Lösung des Gleichungssystems entsteht der zuträufende, die F-Struktur, in der die für die semantische Interpretation wesentlichen grammatischen Funktionen erfaßt und den Konstituenten und Kategorien zugeordnet sind. Das linguistisch Befriedigende und Interessante an dieser Analyse ist die saubere Trennung von Kategorien und Funktionen.

Da der "Filter" (Aufbau der F-Struktur über den die grammatischen Funktionen zugeordnet werden, höchst komplex und computerlinguistisch (noch) umstritten ist, schlagen wir als Alternative eine "Metamorphosis Grammar" (MG) (Colmeaer 1975, Sabatler 1985), genauer eine "Definite Clause Grammar" (DCG), als Grundlage für das Parsing der Wetterberichte vor. In diesem Rahmen besteht die Möglichkeit, die grammatischen Funktionen **auf direktem Wege** zuzuweisen.

Der entscheidende Unterschied des N-Formalismus zu den klassischen formalen Grammatiken besteht in der Tatsache, daß terminale und nicht-terminale Symbole **komplex**, d. h. "TERMS" sein können und keine atomaren Symbole sein müssen.

TERMS sind entweder

- Variablen oder
- funktionale Symbole ohne Argumente ("atoms") oder
- funktionale Symbole mit Argumenten, die bereits TERMS sind.

Beispiele für Symbole in der "Metamorphosis Grammar":

np(X)
 subject(full)
 sentence(X, Y)
 verb(comp(direct))
 sentence(no-subject, with-subject)

(Die Variablen sind unterstrichen)

Der Expandierung derartiger Symbole ist eine Unifikationsoperation vorausgehen (Robins 1965).

Die Möglichkeit, die Symbole mit Variablen versehen, erlaubt eine kontextfreie Formulierung von kontextabhängigen Phänomenen.

Eine MG, deren Regeln links von einem nicht-terminales Symbol aufweisen (das möglicherweise komplex ist) heißt "Definite Clause Grammar". Die folgende DCG kann als eine Etappe auf dem Weg der Entwicklung einer Analysegrammatik französischer Wetterberichte angesehen werden.

ntence(S,Q,V) → SUBJ(S) vp(Q,V) $\left. \begin{array}{l} \text{subordinate} \\ \text{participle} \\ \text{infinitive} \\ \text{coordinate} \end{array} \right\}$

ntence(S,Q,V) → SUBJ(S) vp(Q,V) adjunct

ntence(S,Q,V) → $\left. \begin{array}{l} \text{subordinate} \\ \text{participle} \\ \text{infinitive} \\ \text{coordinate} \end{array} \right\}$ SUBJ(S) vp(Q,V) adjunct

ntence(S,Q,V) → adjunct SUBJ(S) vp(Q,V)

BJ(empty) → \emptyset

BJ(full) → np $\left\{ \begin{array}{l} \text{relative} \\ \text{participle} \end{array} \right\}$

BJ(full) → de np $\left\{ \begin{array}{l} \text{relative} \\ \text{participle} \end{array} \right\}$

BJ(full) → { il, elle, ils, elles, (l')on, celui-ci, celle-ci, ceux-ci, celles-ci, c' }

→ noun

→ det noun

→ det adj(1) noun

→ det noun adj(2)

→ det noun adv adj(X)

→ det noun N-Comp

→ det noun adj(2) N-Comp

→ det adj(1) noun adj(2) N-Comp

→ det adj(1) noun N-Comp

→ det noun $\left\{ \begin{array}{l} \text{relative} \\ \text{participle} \end{array} \right\}$

Comp → prep np

(Q,V) → aux(V) verb(C,V) V-Comp(Q,C)

x(V-fin) → \emptyset

x(V-inf) → { peut, peuvent, pourra, pourront, pourrait, pourraient, doit, doivent, devra, devront, devrait, devraient, risque de, commence à, cesse de, arrête de, s'arrête de, continue à, achève de, recommence à, a tendance à, tend à, finit de, vient, va, termine de, menace de, se met à, débute à ... }

Comp(Q,comp(direct)) → DIR-OBJ(Q)

Comp(Q,comp(de)) → DE-OBJ(Q)

Comp(Q,comp(à)) → A-OBJ(Q)

V-Comp(\underline{Q} , comp(loc(\underline{X}))) \rightarrow PREP-OBJ(\underline{Q} , loc(\underline{X}))

V-Comp(\underline{Q} , comp(direct, loc(\underline{X}))) \rightarrow DIR-OBJ(full) PREP-OBJ(\underline{Q} , loc(\underline{X}))

V-Comp(\underline{Q} , comp(de, loc(\underline{X}))) \rightarrow DE-OBJ(\underline{Q}) PREP-OBJ(\underline{Q} , loc(\underline{X}))

V-Comp(\underline{Q} , comp(direct, de, loc(\underline{X}))) \rightarrow DIR-OBJ(full) DE-OBJ(\underline{Q}) PREP-OBJ(\underline{Q} , loc(\underline{X}))

V-Comp(\underline{Q} , comp(par, de, loc(\underline{X}))) \rightarrow PAR-OBJ(\underline{Q}) DE-OBJ(\underline{Q}) PREP-OBJ(\underline{Q} , loc(\underline{X}))

V-Comp(\underline{Q} , comp(par, loc(\underline{X}))) \rightarrow PAR-OBJ(\underline{Q}) PREP-OBJ(\underline{Q} , loc(\underline{X}))

V-Comp(\underline{Q} , comp(par)) \rightarrow PAR-OBJ(\underline{Q})

PREP-OBJ(empty, \underline{X}) \rightarrow \emptyset

PREP-OBJ(full, \underline{X}) \rightarrow prep(\underline{X}) np $\left\{ \begin{array}{l} \text{relative} \\ \text{participle} \end{array} \right\}$

prep(loc(1)) \rightarrow { à, sur, dans }

prep(loc(2)) \rightarrow { à, vers, sur, dans, jusqu'à, à travers, en direction de }

DIR-OBJ(empty) \rightarrow \emptyset

DIR-OBJ(full) \rightarrow np $\left\{ \begin{array}{l} \text{relative} \\ \text{participle} \end{array} \right\}$

DIR-OBJ(full) \rightarrow de np $\left\{ \begin{array}{l} \text{relative} \\ \text{participle} \end{array} \right\}$

adjunct \rightarrow \emptyset

adjunct \rightarrow np

adjunct \rightarrow prep np

adjunct \rightarrow adv

relative \rightarrow \emptyset

relative \rightarrow qui sentence(empty, full, U-fin)

relative \rightarrow que sentence(full, empty, U-fin)

relative \rightarrow où sentence(full, empty, U-fin)

relative \rightarrow $\left\{ \begin{array}{l} \text{dans} \\ \text{sur} \end{array} \right\} \left\{ \begin{array}{l} \text{lequel} \\ \text{laquelle} \end{array} \right\}$ sentence(full, empty, U-fin)

subordinate \rightarrow \emptyset

subordinate \rightarrow $\left\{ \begin{array}{l} \text{parce que} \\ \text{puisque} \\ \text{alors que} \\ \text{tandis que} \\ \text{quand} \\ \text{pendant que} \\ \text{bien que} \\ \text{si} \end{array} \right\}$ sentence(full, full, U-fin)

coordinate \rightarrow \emptyset

Fachbeitrag

coordinate → $\left\{ \begin{array}{l} \text{et} \\ \text{mais} \\ \text{puis} \\ \text{,} \end{array} \right\}$ sentence(empty,full,V-fin)

coordinate → $\left\{ \begin{array}{l} \text{et} \\ \text{mais} \\ \text{puis} \\ \text{,} \end{array} \right\}$ sentence(full,full,V-fin)

participle → \emptyset

participle → $\left\{ \begin{array}{l} \text{et} \\ \text{mais} \\ \text{en} \\ \text{,} \end{array} \right\}$ sentence(empty,full, $\left\{ \begin{array}{l} \text{V-ant} \\ \text{V-é} \end{array} \right\}$)

participle → sentence(empty,full, $\left\{ \begin{array}{l} \text{V-ant} \\ \text{V-é} \end{array} \right\}$)

participle → sentence(full,full, $\left\{ \begin{array}{l} \text{V-ant} \\ \text{V-é} \end{array} \right\}$)

infinitive → \emptyset

infinitive → pour sentence(empty,full,V-inf)

verb(comp(direct),V-inf) → { gagner, quitter, affecter, subir ... }

verb(comp(direct),V-ant) → { gagnant ... }

verb(comp(direct),V-é) → { gagné ... }

verb(comp(direct),V-fin) → { gagne, gagnent, gagnera, gagneront ... }

verb(comp(loc(1)),V-inf) → { prédominer, circuler, se trouver, pleuvoir, être observé, être augmenté ... }

verb(comp(loc(1)),V-ant) → { prédominant ... }

verb(comp(loc(1)),V-é) → { prédominé ... }

verb(comp(loc(1)),V-fin) → { prédomine, prédominent, prédominera, prédomineront ... }

verb(comp(loc(2)),V-inf) → { entrer, se diriger ... }

verb(comp(loc(2)),V-ant) → { entrant ... }

verb(comp(loc(2)),V-é) → { entré ... }

verb(comp(loc(2)),V-fin) → { entre, entrent, entrera, entreront ... }

verb(comp(de),V-inf) → { profiter, s'absenter, être précédé, être suivi ... }

verb(comp(de),V-ant) → { profitant ... }

verb(comp(de),V-é) → { profité ... }

verb(comp(de),V-fin) → { profite, profitent, profitera, profiteront ... }

- verb(comp(à),V-inf) → { échapper, tourner, succéder, assister ... }
- verb(comp(à),V-ant) → { échappant ... }
- verb(comp(à),V-é) → { échappé ... }
- verb(comp(à),V-fin) → { échappe, échappent, échappera, échapperont ... }
- verb(comp(direct,loc(1)),V-inf) → { affaiblir, noter, provoquer ... }
- verb(comp(direct,loc(1)),V-ant) → { affaiblissant ... }
- verb(comp(direct,loc(1)),V-é) → { affaibli ... }
- verb(comp(direct,loc(1)),V-fin) → { affaiblit, affaiblissent, affaiblira, affaibliront ... }
- verb(comp(de,loc(2)),V-inf) → { se déplacer, se décaler ... }
- verb(comp(de,loc(2)),V-ant) → { se déplaçant ... }
- verb(comp(de,loc(2)),V-é) → { déplacé ... }
- verb(comp(de,loc(2)),V-fin) → { se déplace, se déplacent, se déplacera, se déplaceront ... }
- verb(comp(direct,de,loc(2)),V-inf) → { traverser, diriger ... }
- verb(comp(direct,de,loc(2)),V-ant) → { traversant ... }
- verb(comp(direct,de,loc(2)),V-é) → { traversé ... }
- verb(comp(direct,de,loc(2)),V-fin) → { traverse, traversent, traversera, traverseront ... }
- verb(comp(par,de,loc(2)),V-inf) → { être poussé ... }
- verb(comp(par,loc(1)),V-inf) → { être renforcé ... }
- verb(comp(par),V-inf) → { être concerné, être provoqué ... }
- adj(1) → { grand, petit, beau, haut, bas, mauvais, vaste, nouveau, premier, deuxième, troisième ... }
- adj(2) → { blanc, bleu, gris, sec, humide, chaud, frais, froid, méridional, atmosphérique, nuageux, actif, local, négatif, central ... }
- adv → { plus, peu, assez, de plus en plus, à nouveau, en général, parfois, souvent, plutôt, adj+ment ... }

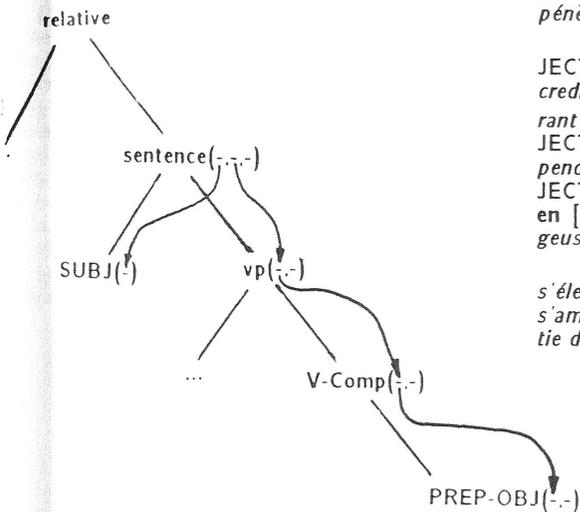
Die lexikalischen Kategorien **det**, **noun** und **prep** sind nicht weiter expandiert. Die Ersetzung von **aux**, **verb**, **adj(1)**, **adj(2)** und **adv** durch lexikalische Elemente ist nur angedeutet. **Adjuncts** sind nicht-satzförmige Satzadjunkte oder Angaben.

S, **O**, **V**, **C** und **X** sind Variablen. Beim Lesen der Grammatik muß man sich immer wieder vor Augen halten, daß **In Jeder Regel** Variablen mit demselben Namen dasselbe bezeichnen. Das Sym-

bol sentence ist mit drei Variablen oder Argumenten versehen: **sentence(-,-,-)**. Das erste Argument (die Variable **S**) spezifiziert die Eigenschaft **vs. full** des Subjekts, das zweite (**O**) dieselbe Eigenschaft des (jeweils unterschiedlichen Typs) Objekts, das dritte (**V**) die Verbform (finit, Infinitiv, Partizip). Die Symbole **vp**, **V-Comp**, **PREP** und **verb** sind mit jeweils zwei Variablen oder Argumenten versehen. Die Symbole **SUBJ**, **DIR-DE-OBJ**, **A-OBJ**, **PAR-OBJ**, **aux**, **adj** und

nd mit jeweils einer Variablen oder einem Argument versehen. Das erste Argument des Symbols $p(-,-)$, die Variable O , spezifiziert die Eigenschaft **ill vs. empty** des Objekts, das zweite, die Variable V , die Form des Verbs. Das erste Argument des Symbols $verb(-,-)$ sowie das zweite Argument des Symbols $V-Comp(-,-)$, jeweils die Variable C , spezifizieren den Typ der verbalen Ergänzung. $V-Comp$ hängt ja ab vom Verbtyp. Die Ersetzung von $p(O,V)$ erfolgt unter der Bedingung, daß $verb$ und $V-Comp$ einander "entsprechen". Dies ist immer dann der Fall, wenn die Variable C von $verb$ und $V-Comp$ denselben Wert annimmt. Das Argument von $loc(-,-)$, die Variable X , spezifiziert den Typ der lokalen Präposition, die ebenfalls vom Verbtyp abhängt.

Man benutzt die Argumente der Symbole, um Restriktionen von einem Symbol zum anderen weiterzuleiten. Durch die komplexen Symbole $sentence(-,-,-)$, $vp(-,-)$ und $V-Comp(-,-)$. Iso über die Variablen S und O , wird beispielsweise die Abhängigkeit zwischen Relativpronomen *qui, que, où* etc.) und (leerem) Subjekt oder Objekt ($SUBJ(-)$, $DIR-OBJ(-)$ oder $PREP-OBJ(-)$) des Relativsatzes weitergereicht. Vgl.



Damit sind auch die unterschiedlichen (elliptischen) Objekttypen in mit *que* oder *où* eingeleiteten Relativsätzen mit dieser Grammatik rekonstruierbar, oder genauer, mit dem "antécédent" des Relativpronomens identifizierbar: vgl.

La vallée du Rhône
qui (SUBJ(empty)) échappera aux averses,
 ...
qui (SUBJ(empty)) subira des averses, ...
où le vent se renforcera (PREP-OBJ(empty)), ...
où le vent augmentera son activité (PREP-OBJ(empty)), ...
que la tramontane traversera (DIR-OBJ(empty)), ...

Über diese Grammatik sind nicht nur für die vollständigen Propositionen die jeweiligen Subjekte, direkten Objekte und unterschiedlichen Typen von präpositionalen Objekten identifizierbar, sondern es sind vor allem auch die getilgten oder elliptischen Subjekte von Infinitivsätzen, Partizipialsätzen, koordinierten Sätzen und Relativsätzen zu erkennen (als "empty-subjects"), so daß jeweils vollständige Propositionen rekonstruierbar sind. Das dem Relativsatz und Partizipialsatz unmittelbar vorausgehende Nomen oder die Nominalgruppe sowie das Subjekt

des vorausgehenden Satzes im Falle der Infinitivkonstruktion mit *pour*, der Partizipialsätze und koordinierten Sätze können auf diese Weise als Subjekt einem Satz mit "empty-subject" zugeordnet werden. Dabei sind, wie wir gesehen haben, "long distance dependencies" beschreibbar. Aus der Grammatik ergeben sich beispielsweise die folgenden Analysen hinsichtlich der Subjekte der Propositionen, die jeweils, sofern leer, mit den hervorgehobenen vorausgehenden nps zu unifizieren sind.

LA ZONE DÉPRESSIONNAIRE [(SUBJECT(empty)) *centrée sur le sud de la Scandinavie*] *dirige sur la France un flux faible de nord-ouest.* (03.05.85)

LES HAUTES PRESSIONS *qui* [(SUBJECT(empty)) *recouvrent l'Atlantique débordent sur la France,*] [(SUBJECT(empty)) *rejetant les perturbations sur les îles Britanniques.*] (22.08.85)

LE SYSTÈME DÉPRESSIONNAIRE [(SUBJECT(empty)) *induisant un mauvais temps pluvieux sur une grande partie de la France,*] *se décale lentement vers le nord-ouest, tandis que* [(SUBJECT(full)) *une nouvelle perturbation atlantique pénètre sur les régions de l'ouest.*] (07.05.85)

LES HAUTES PRESSIONS *qui* [(SUBJECT(empty)) *s'installent sur la France mercredi et jeudi,*] *rejetent temporairement le courant perturbé vers les îles Britanniques,* [(SUBJECT(full)) **UN FRONT FROID** *abordera cependant l'ouest du pays, jeudi soir,*] *puls* [(SUBJECT(empty)) *pénétrera lentement vers le Sud-Est*] *en* [(SUBJECT(empty)) *prenant une activité orageuse modérée.*] (06.05.85)

[(SUBJECT(full)) *Le champ de pression s'élevait progressivement par l'ouest,*] *le temps s'améliorera dès dimanche sur la plus grande partie du pays.* (23.03.85)

LES CONDITIONS ANTICYCLONIQUES *se maintiennent sur la majeure partie de la France,* [(SUBJECT(empty)) *rejetant*] **LE FLUX PERTURBÉ** *qui* [(SUBJECT(empty)) *circule dans un courant de nord-ouest*] *sur les régions du nord et du nord-est.* (21.02.85)

Les autres régions seront encore épargnées avant l'approche de **CE SYSTÈME PLUVIEUX** *qui* [(SUBJECT(empty)) *se décalera vers l'est*] *pour* [(SUBJECT(empty)) *toucher en soirée les régions du Sud-Ouest au Centre.*] (23.08.85)

Für die beiden folgenden Passagen wird das Ergebnis der Analyse auf der propositionalen Ebene, nämlich die semantischen Repräsentationen für die einzelnen Propositionen, angedeutet.

La dépression des îles Britanniques se décale vers le Danemark. Le flux perturbé s'oriente à l'ouest, et un temps plus doux va s'établir sur la France. Lundi matin, des résidus d'une première perturbation vont encore se maintenir le matin sur le Jura, les Alpes, la Provence et la Corse. Une zone pluvieuse venant d'Espagne va aborder les régions pyrénéennes, et une nouvelle perturbation modérément active va arriver sur les côtes de la Manche. (26.03.85)

- P1 La dépression des îles Britanniques se décale vers le Danemark
 A Ursprung-Ziel-Bewegungsverb G
- P2 Le flux perturbé s'oriente à l'ouest
 A Ursprung-Bewegungsverb S
- P3 Un temps plus doux va s'établir sur la France
 Phen Szene-Prozeß-Verb (apparition) L
- P4 Des résidus d'une première perturbation vont se maintenir sur la Jura, les Alpes ...
 Phen Szene-Prozeß-Verb (maintien) L
- P5 (Une zone pluvieuse) vient d'Espagne
 A Ursprung-Ziel-Bewegungsverb S
- P6 Cette zone pluvieuse va aborder les régions pyrénéennes
 A Ziel-Bewegungsverb G
- P7 Une nouvelle perturbation va arriver sur les côtes de la Manche
 A Ursprung-Ziel-Bewegungsverb G

Une vaste dépression centrée au nord de l'Irlande dirige un flux cyclonique de sud-ouest à travers la France. Une perturbation située sur les régions de l'Ouest vendredi matin s'enfonce lentement à travers le pays et favorisera une évolution orageuse sur le relief. (09.08.85)

- P1 (Une vaste dépression) (est) centrée au nord de l'Irlande
 Phen Positionsverb L
- P2 Cette dépression dirige un flux cyclonique à travers la France
 A Kausatives Bewegungsverb P G
- P3 (Une perturbation) (est) située sur les régions de l'ouest
 Phen Positionsverb L
- P4 Cette perturbation s'enfonce (lentement) à travers le pays
 A Ziel-Bewegungsverb G
- P5 (Cette perturbation) favorisera une évolution orageuse
 A Ursache-Wirkungs-Verb Phen

Semantische Funktionen:
 A = Agent
 P = Patient
 G = Goal (Ziellokativ)
 S = Source (Ursprungslokal)
 L = Location (Szene)
 Phen = Phänomenon

Textuelles Wissen

Die für eine maschinelle Analyse wesentlichen textlinguistischen Phänomene dieser Textklasse sind die Sinnrelationen, die thematischen Propositionen, die Globalstruktur und die Anaphorik. Hier soll allgemein die Verweisproblematik. Hier soll kurz auf zwei dieser Komponenten des textuellen Wissens eingegangen werden:

1.1 Sinnrelationen oder logische Verknüpfungen zwischen Propositionen

Logische Verknüpfungen zwischen Propositionen sind zwar auch in Fachtexten bisweilen implizit, alltags-sprachlich

- (1) *Il pleut. Je reste à la maison.*

Parallel dazu:

- (2) *La pression s'élève. Le temps s'améliore.*
 (3) *La dépression des îles Britanniques se décale vers le Danemark. Le flux perturbé s'oriente à l'ouest.*

In (2) und (3) hilft nur Fach- und Sachwissen, um die Ursache-Wirkung-Relation herzustellen. Dies ist im Fall der Wetterberichte geographisches und meteorologisches Wissen. - Voraussetzungen für die Entwicklung einer Fachwissenschaftskomponente liegen übrigens bei Habel und Rollinger (1984) und hinsichtlich des geografischen Wissens in Frankreich bei Sabatier/Duchler (1982) vor. - Die logischen Verknüpfungen, auch Sinnrelationen genannt, werden jedoch erfreulicherweise häufiger in Fachtexten explizit gemacht. Indikatoren für diesen Relationstyp sind hier charakteristischerweise Verben und, wie in der Gemeinsprache, Konnektoren. In folgenden Beispiel (4) signalisieren *entraîner* und *laisser* eine Ursache-Wirkung-Relation:

- (4) *Un bel anticyclone se développe sur la France, entraînant une période de beaux temps qui persistera pratiquement toute la semaine.* (16.04.85)

Les hautes pressions de l'Europe centrale vont régresser vers l'est laissant les perturbations pénétrer plus profondément sur la France. (13.02.85)

Die zeitliche Abfolge wird mit Hilfe der Verben *être associé à* und *laisser la place à* in den folgenden Beispielen ausgedrückt.

- (5) *Le flux de sud-ouest qui va progressivement se généraliser à tout le pays est associé à une remontée sensible des températures.* (29.03.85)

Enfin, sur les régions du quart Sud-Ouest, le phénomène le plus marquant sera la présence des brumes et de brouillards parfois épais, qui laisseront la place à un ciel souvent dégagé. (02.02.85)

Im ersten Fall besteht die Relation der Gleichzeitigkeit, im zweiten die Relation der Vor-/Nachzeitigkeit.

Dieselben Relationen, doch zusätzlich die der Gegensätzlichkeit werden durch Konnektoren explizit gemacht. Hinter dieser Klasse von sprachlichen Ausdrücken, die sich in der Textlinguistik und Dialogforschung als nützlich erwiesen hat, verbergen sich die traditionellen Wortarten Konjunktionen, Präpositionen und Adverbien. Verursachung, Gleichzeitigkeit und Vor-/Nachzeitigkeit werden beispielsweise mit Hilfe von *ainsi*, *avec* und *après* explizit gemacht.

Besonders häufig sind Konnektoren der Gegensätzlichkeit, wie *toutefois*, *bien que*, *mais* sowie *sauf* und *excepté* in der typischen Konstruktion

{ *sauf* / *excepté* } + Szene + Relativsatz

Charakteristisch für die Textklasse ist schließlich ein weiterer ganz spezifischer Typ von Sinnrelation, die Kombination aus oder Überlagerung von Gleichzeitigkeit und Gegensätzlichkeit, wie wir sie von der traditionellen Grammatik her in der Konjunktion *tandis que* vereint finden:

- (6) *Un front de faible activité intéressera le nord et l'est du pays tandis qu'une dorsale persistera dans l'ouest.* (27.04.85)

Diese Sinnrelation wird außerdem durch *alors que* und durch Kombinationen von Konnektoren und/oder Indikatoren an der Oberfläche repräsentiert wie *mais en même temps*, *mais associé à*, *en étant toutefois*, *mais avec* und *avec*, *cependant* sowie *avec*, *toutefois*. Vgl. (7)

- (7) *Les perturbations vont continuer à se succéder avec, toutefois, une activité pluvieuse moins importante.* (10.04.85)
 Vgl. Anhang 3

2.2 Assoziative anaphorische Beziehungen

Ein anderer Typ der Verbindung zwischen Propositionen kommt durch Verweisung oder Anaphorik zustande. Hier verweisen Elemente in einer Proposition, die Anaphern, auf Elemente in vorangehenden Propositionen, die Antezedenten. Komplizierter noch als pronominale Anaphern für die Analyse sind nominale Anaphern vom Typ *le N* oder *son N*, auch *ce N*, die im Material ebenso häufig sind wie die pronominalen. Auch hier sind triviale Fälle der Wiederaufnahme durch Wiederholungen oder synonyme Ausdrücke von den interessanteren Fällen zu unterscheiden, wo Anapher und Antezedent unterschiedliche Bedeutung, d. h. nicht denselben Referenten haben. Dieses Phänomen der assoziativen Anapher findet sich in (1a) und (1b).

- (1a) *Un camlon est passé. LE CONDUCTEUR m'a salué*
 (1b) *Je me suis acheté un vélo. LES FREINS fonctionnent pas*

Hier liegt keine logische Verknüpfung vor. In der Praxis findet man jedoch häufig logische und anaphorische Verknüpfung parallel oder überlagert. Beispiele für assoziative Anaphern im Material:

- (2a) *Un front pluvio-orageux concernera environ une moitié est du pays. LES VENTS souffleront du secteur sud et seront assez forts à forts.*
 (2b) *Une première perturbation située le matin de la Normandie aux Pyrénées atteindra le soir les régions de l'Est; elle sera précédée de faibles précipitations neigeuses; L'ACTIVITÉ sera toutefois plus marquée de l'Aquitaine au golfe de Lion.*
 (2c) *La France restera dans un courant cyclonique du nord-ouest. LES AVERSES seront fréquents, et LE VENT soufflera en rafales.*

(2d) **La dépression centrée au large du Portugal, qui apporte de l'air chaud, remontera vers le nord-ouest de l'Espagne, puis vers le centre-est de la France. LA PERTURBATION ASSOCIÉE réactivera la masse nuageuse située de la Bretagne au Centre et aux Alpes.**

(2e) **La dépression des îles Britanniques se décale vers le Danemark. LE FLUX PERTURBÉ s'oriente à l'ouest.**

Welche Regularitäten liegen hier zugrunde oder konkreter: Wie können im Fall der assoziativen Anaphorik die Antezedenten erkannt werden? Es

lassen sich zwei unterschiedliche Relationen zwischen Antezedent und Anapher unterscheiden. Die **Prädikat-Argument-Relation** (oder **aktantelle Relation**) und die **nicht-aktantelle Relation**. z. B. **est-un-** oder **est-avec-**Relation. Im Falle sind es die untersuchten **semantischen Funktionen**, also die aus dem propositionalen Wissen bekannten Relationen zwischen Prädikaten und Argumenten oder zwischen Argumenten ein und dem Prädikats, die die anaphorische Verweisung stellen. Das sei zunächst an einem gemeinsamen Beispiel verdeutlicht, vgl. (3):

(3a) $\left\{ \begin{array}{l} \text{Pierre se promenait dans la gare.} \\ \text{Pierre devait se rendre à Paris.} \\ \text{Pierre attendait le train.} \end{array} \right\}$ Une demi-heure avant le départ, il a acheté un

(3b) Le départ est fixé pour demain matin. $\left\{ \begin{array}{l} \text{La voiture est en panne} \\ \text{La destination est inconnu} \\ \text{La petite ville de Province est sans aucun int}$

(3c) Pierre est allé à la gare. $\left\{ \begin{array}{l} \text{Il n'y avait pas de train le Dimanche.} \\ \text{Paris était loin d'ici.} \end{array} \right\}$

Vgl. Pierre est parti en voiture de la maison à son lieu de travail partir [A. S. G. I]

Die semantisch spezifizierten Argumente **partir** (Ursprung, Ziel, Instrument, Agens) können als Anapher oder Antezedent von **départ** oder **partir** (vgl. (3a) und (3b)) funktionieren, sie können aber auch für sich allein jeweils die Anapher- oder Antezedentfunktion übernehmen (vgl. (3c)). Sie eröffnen sich willkommene Zusammenhänge zwischen textuellem und propositionalem Wissen. Soziale anaphorische Beziehungen sind über die Prädikat-Argument-Struktur (oder auch semantische Verbvalenz) zu erklären und aufzudecken.

Diese für die Verweisproblematik wichtige Kenntnis läßt sich auf die Wetterberichterstattung übertragen. Wie in (1a) (**conducteur**), so ist es in (2b) (**activité**) jeweils die Nominalisierung des Prädikats, mit der auf ein zugehöriges Argument verwiesen wird. Vgl. auch

(4) Le conducteur conduit le camion (1a)
La perturbation est active (2b)

$\left\{ \begin{array}{l} \text{Prädikat-Argument-} \\ \text{Aktantielle} \end{array} \right\}$ Relation

Der zweite Relationstyp, die **nicht-aktantelle Relation** zwischen Antezedent und Anapher, wird in unseren Beispielen ausschließlich durch die **est-avec-**Relation repräsentiert. Diese Relation ist

... X (der Antezedent) est avec Y (die Anapher). In (1a), (2a), (2c), (2d) und (2e) liegt die Relation vor. Vgl. auch

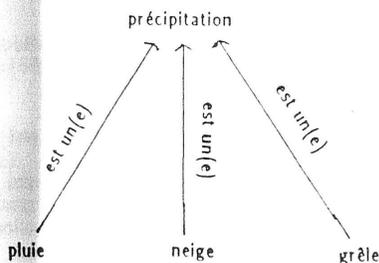
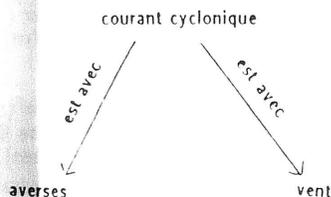
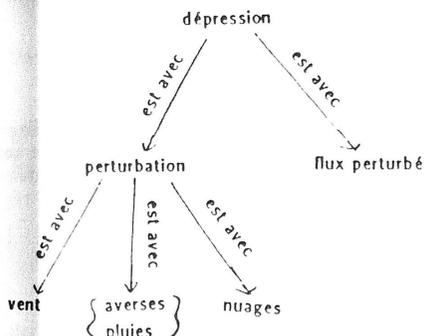
(5) $\left\{ \begin{array}{l} \text{Le vélo est avec des freins (1b)} \\ \text{Un front pluvio-orageux est avec des vents (2a)} \\ \text{Un courant cyclonique est avec des averse, du vent (2c)} \\ \text{Une dépression est avec une perturbation (2d)} \\ \text{Une dépression est avec un flux perturbé (2e)} \end{array} \right\}$

est-avec-Relation

Dieser Relationstyp ist offenbar eng verwandt mit der Tell-von-Relation. Nicht-aktantielle Relationen sind in der Fachwissenskategorie in Form von semantischen Netzen repräsentiert, die sich zumindest teilweise beim Lesen der Texte aufbauen lassen. Vgl.

Ausschnitte aus semantischen Netzen

(Knoten = Begriffe, Kanten = Relationen):



Textausschnitte:

1. Une zone de temps couvert et accompagnée de quelques pluies fines ... Cette perturbation

2. Une perturbation... Les nuages de la perturbation

Die der Anapher vorangehenden Nomina oder Nominalgruppen können also aufgrund ihrer semantischen Beschreibung als Antezedent ermittelt werden. In diese Beschreibung fließt sowohl propositionales Wissen als auch (aus den Texten (re)konstruierbares) Fachwissen mit ein.

3 Ausblick

Der vorgestellte Entwurf einer Analysegrammatik bedarf selbstverständlich der Ergänzung und Überarbeitung. Es fehlt z. B. noch eine vollständige Nominalgruppenanalyse, in der analog zur vp-Analyse die Zuordnung von semantischen Funktionen zu den Attribut- oder Modifiertypen in Abhängigkeit vom noun(-Typ) beschrieben werden kann.

La dépression des îles Britanniques

DE-MOD
Location

L'orientation du vent à l'ouest

DE-MOD A-MOD
Agent Source

Les régions de la Vendée à la Normandie

DE-MOD A-MOD
Source Goal

Le renforcement des vents

DE-MOD
Phenomenon

Les vents du sud

DE-MOD
Source

Die Nominalgruppenanalyse baut auf der vorliegenden Beschreibung auf. Auch hier sind in Abhängigkeit von der Prädikatsklasse über unterschiedliche grammatische Funktionen (hier Modifiertypen) die semantischen Funktionen zuzuordnen.

Dieser Beitrag versteht sich als Ergänzung zu den KI-orientierten Arbeiten von Habel, Rollinger und anderen an den deutschen Wetterberichten aus dem "Berliner Tagesspiegel". Im weitesten Sinne wurden hier und in Berlin Vorarbeiten geleistet für die Entwicklung eines Qualitätsanspruchs genügenden Übersetzungssystems Französisch-Deutsch bzw. Deutsch-Französisch für Wetterberichte, wie es beispielsweise für im Elsaß erscheinende Tageszeitungen gebraucht werden könnte. Gleichzeitig hoffen wir, Impulse gegeben zu haben in Richtung auf ein Fachtextverstehen und eine Fachtextgenerierung, bei denen neben dem Fach- und Deduktionswissen in gleicher Weise sprachliches Wissen berücksichtigt werden sollte.

Anhang 1: Liste der (syntaktisch und) semantisch homogenen Verbklassen

1. Ziel-Bewegungsverben

- a) entrer, se diriger
buter (stoßen auf), déborder (übergreifen), se diriger (sich wenden) s'enfoncer (tief eindringen), entrer (eintreten), s'étendre (sich ausweiten), s'évacuer (wegziehen), se heurter (stoßen auf), s'engouffrer (hineinströmen), pénétrer (eindringen)
- b) atteindre, gagner
atteindre (erreichen), envahir (eindringen), contourner (umgehen), franchir (überqueren), gagner (erreichen), toucher (berühren), rallier (dazustoßen), escalader (hochklettern), gravir (erklimmen), grimper (klettern), pénétrer (eindringen), rencontrer (treffen), heurter (stoßen), aborder (sich nähern)

2. Ursprung-Bewegungsverben

- a) s'absenter, s'éloigner
s'absenter (sich absentieren), sortir (rausgehen), s'échapper (entwischen), s'éloigner (sich entfernen), s'évader (entrinnen), provenir (herkommen)
- b) quitter, abandonner
- c) tourner, s'orienter (bei Bewegung von Luftmassen)

3. Ursprung-Ziel-Bewegungsverben

- a) se déplacer, arriver
s'en aller (weggehen), arriver (ankommen), s'avancer (voranschreiten), déboucher (münden, herauskommen), se déplacer (sich bewegen), descendre (hinuntergehen), s'envoler (wegfliegen), filer (flitzen), monter (steigen), partir (abziehen), se précipiter (sich stürzen), remonter (wieder hochsteigen), rentrer (zurückkehren), repartir (wieder wegziehen), se replier (sich zurückziehen), se retirer (sich zurückziehen), retourner (zurückkehren), revenir (zurückkommen), se ruier (sich stürzen), se sauver (abhauen), dévaler (herunterkommen), venir (kommen)
- b) progresser, se décaler
aller (gehen), courir (laufen), déferler (sich wälzen), se décaler (sich verschieben), foncer (flitzen), se glisser (sich einschleichen), s'installer (einziehen/sich breit machen), s'orienter (sich orientieren), passer (vorbeiziehen), progresser (fortschreiten), se propager (sich verbreiten), reculer (zurückweichen), repasser (wieder vorbeiziehen), tourner (einbiegen), voler (fliegen), viser (zielen), régresser (zurückschreiten)

4. Szene-Ursprung-Ziel-Bewegungsverben

- traverser, parcourir
traverser (überqueren), longer (entlangziehen), parcourir (durchziehen), sillonner (durchziehen), survoler (überfliegen), recouvrir (überdecken), franchir (überqueren), monter (hochgehen), descendre (hinuntergehen), remonter (wieder hochgehen)

5. Kausative Bewegungsverben

- a) diriger, apporter
amener (mitbringen), attirer (anlocken), conduire (geleiten), emmener (mitnehmen), entraîner (mit sich reißen), envoyer (schicken), escorter (eskortieren), expédier (senden), mener (führen), ramener (zurückbringen), renvoyer

(zurückschicken), transporter (befördern), véhiculer (befördern), apporter (mitbringen), diriger (lenken), emporter (mitnehmen), lancer (lanzieren), pousser (schieben), propager (ausbreiten), propulser (schießen), ramener (zurückbringen), refouler (zurückdrücken), repousser (zurückschieben), renvoyer (zurückwerfen)

- b) être poussé, être envoyé etc.

6. Szene-Bewegungsverben (mouvement place)

circular, onduler

7. Positionsverben

se trouver, être situé
se trouver (sich befinden), être situé (sich befinden), se situer (liegen), être centré (zentrieren), être installé (liegen), rester (bleiben), résider (bleiben), être établi (festsitzen), régner (herrschen), se localiser (liegen)

8. Szene-Prozeßverben

s'établir, se former

Maïntien

se maintenir (sich aufrecht erhalten), s'étendre (sich ausweiten), se dérouler (abspielen), se produire (stattfinden), persister (bestehen bleiben), se développer (entwickeln), s'accumuler (sich anhäufen), généraliser (sich ausbreiten)

Appartien

s'organiser (sich aufbauen), s'établir (festsetzen), se former (sich bilden), se fixer (sich fixieren), s'installer (sich aufbauen), manifester (sich manifestieren), apparaître (erscheinen), réapparaître (wieder erscheinen)

Dispartien

disparaître (verschwinden), se dissiper (sich auflösen), se dissiper (sich auflösen), se désagréger (sich auflösen)

9. Szene-Veränderungsverben (Changement)

se renforcer, devenir beau
s'aggraver, devenir + Adj., (re)prendre l'activité, perdre de l'activité, se dégrader, s'affaiblir, se gêner, s'atténuer, s'améliorer, se renforcer, diminuer, faiblir, évoluer, faire rare, prendre un caractère, s'intensifier, s'accélérer, se voiler

10. Kausative Veränderungsverben (verbes satellites de changement)

affaiblir, renforcer
réduire, ralentir, affaiblir, renforcer, diminuer, augmenter, accélérer, intensifier, être forcé, être augmenté etc.

11. Zustandsverben

être stationnaire, prédominer
prédominer, exister, subsister, être stationnaire/fréquent, rare/abondant, important ...

12. Ursache-Wirkungs-Verben (kausale Relationsverben)

provoquer, donner
provoquer, entraîner, induire, donner, maintenir, donner lieu à, permettre, favoriser, commander, laisser + Inf, être associé, faire + Inf, rendre, être lié à, être provoqué, être favorisé

Psychologische Verben

- a) affecter
affecter, influencer, intéresser, toucher, épargner, concerner, menacer
- b) subir
- c) profiter
- d) échapper
- e) être concerné
subir, être/rester sous l'influence de, souffrir de, profiter de, échapper à, bénéficier de, être à l'abri de, être épargné de/par, craindre, rester/être soumis à, être affecté

Feststellungs- und Wahrnehmungsverben

- voir, observer
voir, observer, noter, remarquer, constater, être observé, être remarqué, assister à

Temporale Relationsverben

précéder, suivre

- a) N₁ V N₂ oder N₂ est V-é de/par N₁
précéder, suivre, accompagner, remplacer.
- b) N₁ et N₂ V
alterner, se succéder, être entrecoupé, se suivre, être en alternance, se poursuivre
- c) N₁ V à N₂
succéder, faire place, laisser la place, être associé, se substituer, céder (peu à peu), céder la place

Aspekt- und Modalitätsverben

- commencer, risquer
recommencer, continuer, commencer, risquer, achever, avoir tendance, tendre, pouvoir, devoir, finir, venir + Inf, aller + Inf, terminer, menacer, arrêter, s'arrêter, se mettre à, débiter, cesser

Meteorologische Verben

- a) pleuvoir, neiger
il pleut, il neige, il vente, il gèle, il fait beau, il fait doux, il fait frais/froid, il fait agréable, il fait mauvais ...
- b) souffler, briller
souffler (vent), briller (soleil), se couvrir (ciel), tomber (vent), se lever (vent), se creuser (dépression), éclater (orage)

Anhang 2: Zuordnungen von grammatischen Funktionen zu semantischen Funktionen**1. Ziel-Bewegungsverben**

- a) entrer, se diriger
A G
SUBJ PREP-OBJ
- b) atteindre, gagner
A G
SUBJ DIR-OBJ

2. Ursprung-Bewegungsverben

- a) s'absenter, s'éloigner
A S
SUBJ DE-OBJ
- b) quitter, abandonner
A S
SUBJ DIR-OBJ

c) tourner, s'orienter (+ Wind)

A S
SUBJ A-OBJ

3. Ursprung-Ziel-Bewegungsverben

- a) se déplacer, arriver
A S G
SUBJ DE-OBJ PREP-OBJ
- b) progresser, se décaler
A S G
SUBJ DE-OBJ PREP-OBJ

* A S
SUBJ DE-OBJ

4. Szene-Ursprung-Ziel-Bewegungsverben

traverser, parcourir
A L S G
SUBJ DIR-OBJ DE-OBJ PREP-OBJ

5. Kausative Bewegungsverben

- a) diriger, apporter
A P S G
SUBJ DIR-OBJ DE-OBJ PREP-OBJ
- b) être poussé, être envoyé
P A S G
SUBJ PAR-OBJ DE-OBJ PREP-OBJ

6. Szene-Bewegungsverben (mouvement sur place)

circuler, onduler
A L
SUBJ A/SUR/DANS-OBJ

7. Positionsverben

se trouver, être situé
Phen L
SUBJ A/SUR/DANS-OBJ

8. Szene-Prozeß-Verben

s'établir, se former
Phen L
SUBJ A/SUR/DANS-OBJ

9. Szene-Veränderungsverben

se renforcer, se faire rare, devenir beau
Phen L
SUBJ A/SUR/DANS-OBJ

10. Kausative Veränderungsverben (verbes causatifs de changement)

- a) affaiblir, renforcer
A Phen L
SUBJ DIR-OBJ A/SUR/DANS-OBJ
- b) être renforcé, être augmenté
Phen A L
SUBJ PAR OBJ A/SUR/DANS-OBJ

11. Zustandsverben

être stationnaire, prédominer
Phen L
SUBJ A/SUR/DANS-OBJ

12. Psychologische Verben

- a) affecter
Phen E/L
SUBJ DIR-OBJ
- b) subir
Phen E/L
DIR-OBJ SUBJ
- c) profiter
Phen E/L
DE-OBJ SUBJ
- d) échapper
Phen E/L
A-OBJ SUBJ
- e) être concerné
Phen E/L
PAR-OBJ SUBJ

13. Feststellungs- und Wahrnehmungsverben

- a) voir, noter
Phen E L
DIR-OBJ SUBJ A/SUR/DANS-OBJ
- b) être observé, être remarqué
Phen L
SUBJ A/SUR/DANS-OBJ

14. Ursache-Wirkungs-Verben

- a) provoquer, entraîner
Phen A
DIR-OBJ SUBJ

Anhang 3: Liste der Indikatoren für Sinnrelationen

1. Ursache-Wirkung-Relation

Konnektoren: à cause de, grâce à, ainsi, puisque, donc, en raison de, à la faveur de
Verben: provoquer, permettre, entraîner, favoriser, induire, laisser + Inf, apporter, faire + Inf, donner, être associé à, donner lieu à, être lié à, rendre + np + adj, amener, mener à

2. Temporale Relation

- a) **Gleichzeitigkeit**
Konnektoren: alors, en même temps, simultanément, avec, d'autre part, (à/pendant) + Nominalisiertes Verb oder Adjektiv, pendant que, en V-ant, tandis que
Verben: accompagner, être accompagné de, être associé à, être lié à
- b) **Vor-/Nachzeitigkeit**
Konnektoren: (d'abord) ... ensuite, (d'abord) ... puis, après, avant, pour + Inf., V-é (Participe passé), ultérieurement, plus tard
Verben: précéder, suivre, remplacer, faire place à, laisser la place à, succéder à, céder à, céder la place à, succéder à, être suivi de, être précédé de

- c) **Wechsel**
Konnektoren: tantôt ... tantôt
Verben: alterner, alterner avec, se succéder, se poursuivre, être en alternance, se suivre

3. Gegensätzlichkeit

Konnektoren: mais, par contre, en revanche, toutefois, cependant, malgré, en

dépôt de, tandis que, alors que, saucepté, seul(e)(s), tout de même, ai sauf/excepté + Szene + Relativsatz

4. Gleichzeitigkeit und Gegensätzlichkeit
Konnektoren: tandis que; mais en temps; avec, toutefois; mais associé étant toutefois; mais avec; avec, cepen

Literaturangaben

Apresjan, J. D.: Ideen und Methoden der demen strukturellen Linguistik, München 1971
 Bresnan, J. (Hrsg.): The Mental Representation of Grammatical Relations, Cambridge, MA 1978
 Colmerauer, A.: Les Grammaires Métamorphoses, Rapport GIA, Luminy, Université Aix-Marseille 2, 1975; "Metamorphosis Grammatical Relations in Natural Language Communication with Computers", in: L. Bolc (Hrsg.), Springer 1978
 van Dijk, T. A.: Textwissenschaft. Eine interdisziplinäre Einführung, München 1980
 Fradin, B.: Anaphorisation et Stéréotypes syntaxiques, in: Lingua 64, S. 325 - 369, North-Holland 1984
 Greimas, A. J.: Sémantique structurale, Paris 1966
 Gross, M.: Méthodes en syntaxe, Paris 1971
 Fillmore, Ch. J.: "Some Problems for Case Grammar" Ohio State University and Center for Advanced Studies in the Behavioral Sciences; und O'Brien (Hrsg.): 22nd Annual Round Table, S. 56, Georgetown 1971
 Kaplan, R. / Bresnan, J.: "Lexical-Functional Grammar. A Formal System of Grammatical Representation" In: Bresnan (Hrsg.): The Mental Representation of Grammatical Relations, Cambridge, MA 1982
 Rappaport, M.: On the Nature of Derived Morphemes, in: L. Levin, M. Rappaport, A. Zaenen (Hrsg.): Papers in LFG, Bloomington (Ind) 1983
 Robinson, J.: "A machine-oriented logic based on the resolution principle", Journal of the ACM Vol. 12, 1962
 Rollinger, C.-R. (Hrsg.): Probleme der Textverstehens. - Ansätze der künstlichen Intelligenz, Tübingen 1984
 Ruwet, N.: Théorie syntaxique et syntaxe française, Paris 1972
 Sabatier, P.: Variables in Metamorphosis, Grammars and Related Formalisms, Paris, vielfältig 1985
 Sabatier, P / Duchier, D.: Interfrance, un prototype d'interface en langue française, Paris 1983
 Seelbach, D.: Syntax und Semantik französischer Verben, IRAL XIX/4, 1981, S. 293-315
 Seelbach, D.: Oberflächensyntax französischer Verben und lokale Tiefenkasusrelationen. Auf dem Wege zu einer lexikalisch-funktionalen Grammatik mit syntaktisch und semantisch motivierten Verbklassen. (Manuskript) erscheint in: Dietrich/Gauger/Geckeler (Hrsg.): Grammatik und Wortbildung romanischer Sprachen. Beiträge zum Deutschen Romanistentag in Siegen 1985, Tübingen 1986

TEXTBEARBEITUNG MIT DEM PERSONAL COMPUTER EIN ERFAHRUNGSBERICHT ÜBER DAS BYU CONCORDANCE SYSTEM

Jan Brustkern, Gert Frackenpohl

Universität Bonn
Institut für Kommunikationsforschung und Phonetik
Poppelsdorfer Allee 47
D-5300 Bonn 1

Universität zu Köln
D-5000 Köln

BYU Concordance ist ein vielseitiges und interaktives Text Retrieval System, das an der Brigham Young University (Provo, USA) implementiert wurde. Es zeichnet sich durch seine Funktionsvielfalt und die übersichtliche Fenstertechnik aus. Der Funktionsumfang und die Nutzungsmöglichkeiten des Systems werden dargestellt. Es zeigt sich, daß BYU Concordance ein sehr benutzerfreundliches Werkzeug für Lexikographen und Sprachwissenschaftler darstellt.

Nutzungsbereiche und Grundfunktionen

BYU Concordance ist ein Text Retrieval System, das es erlaubt, Textstellen interaktiv in Textdokumenten aufzufinden. Das System wurde an der Brigham Young University (Provo, USA) erstellt und zeichnet sich durch seine Funktionsvielfalt und die übersichtliche Fenstertechnik aus. Die benutzergesteuerte Benutzerführung wird durch ausführliche Erläuterungsbildschirme unterstützt. Der Benutzer wird trotz der Vielzahl der Funktionen schon nach kurzer Einarbeitung in die Lage versetzt, seinen Text in der gewünschten Form zu bearbeiten. Unterstützt wird er hierbei noch durch ein Tutorial, mit dem er das gesamte Funktionsspektrum an einigen Beispieltex-ten nachvollziehen kann. Das BYU Text Retrieval System kann benutzt werden, um

einen umfangreichen Index zu Textdokumenten zu erstellen,

eine Konkordanz zu erzeugen, in der die Wörter mit ihren Kontexten enthalten sind. Diese kann gedruckt oder mit einem Textverarbeitungsprogramm modifiziert werden.

Dokumente nach falsch geschriebenen Wörtern oder Namen zu durchsuchen,

bestimmte Informationseinheiten aus beliebigen Texten zu gewinnen,

bibliographische Angaben oder Abstracts aus einem Textarchiv zu recherchieren,

Informationsmaterialien zu einem Thema aus unterschiedlichen Quellen (z.B. Zeitschriften, Lexika, Enzyklopädien, Kataloge) zusammenzustellen,

die Wort- und Satzverteilung innerhalb literarischer Texte zu untersuchen,

verschiedene Textversionen (z.B. Texte aus unterschiedlichen Quellen oder Zeit-

- abschnitten) miteinander zu vergleichen,
- verschiedene Textmaterialien thematisch miteinander zu vergleichen,
- die Wortumgebung in beliebigem Textmaterial zu analysieren,
- Beispiele der Wortverwendung aus fremdsprachlichen Texten herauszusuchen, die von Übersetzern oder Sprachwissenschaftlern benutzt werden können,
- eine Cross Reference Tabelle der Variablen in einem Computerprogramm zu erstellen.

2. Wo und Wie läßt sich etwas finden?

Nachdem im Eingangsmenue (Abb. 1) ein Text (Book) zur Bearbeitung ausgewählt ist, werden von BYUVIEW die folgenden Suchmöglichkeiten angeboten:

In der alphabetischen Liste der Wörter eines Textes wird durch Blättern oder durch Tippen des Wortanfangs eine Wortform als Suchbegriff ausgewählt. Die Anzahl der Belegstellen im Text wird zu jeder Wortform angezeigt (Abb. 2).

Die Teilstringsuche erlaubt das Auffinden aller Wortformen im Text mit einer bestimmten Wortendung oder einer bestimmten Zeichenkette als Teil der Wortform. Wörter, die im Text untersucht werden sollen, können aus dem Wortregister in eine Suchliste übernommen werden. Auch aus einem angeschlossenen Lexikon können Wörter in die Suchliste aufgenommen werden.

Wortformen können nach verschiedenen Kriterien miteinander kombiniert werden. Es lassen sich Bedingungen über die Aufeinanderfolge der kombinierten Wortformen angeben, oder daß sie innerhalb eines bestimmten Zeichen- bzw. Textbereichs zusammen vorkommen sollen. In ähnlicher Weise lassen sich auch Wortformen aus ganzen Wortlisten miteinander verknüpfen.

Fachp
Z
b
e
A
d
v
e
C
e
d
t
e
h
E
f
v

Computer BookShelf

Press ESPRITDK
ENTER to use ESPRITDT
HUCKFJNN
SHAKESPR
TESTLONG

Book Information

Number of:

Text	1
Kapitel	64
Abschnitt	492
Words	22477
Unique Words	4417
Characters	151499
File Size	188032
Stop Words	787

BYUVIEW
(C)Copyright
1985, 1986
Brigham Young
University
C/O
I K S
Popp. Allee 47
53 Bonn 1

c:\DONCORD\BYU\ESPRITDT.BYB 1st 10 lines

.OP

!E_ESPRIT-TEXT_(deutsch)

IU_1 Vorschlag fuer einen Beschluss des Rates
zur Annahme des ersten europaeischen strategischen
Programms
fuer Forschung und Entwicklung auf dem Gebiet
der Informationstechnologie[

IU_2 ESPRIT [

IU_3 Einleitung[

Acht von zehn der in Europa verkauften Personalcomputer

Abb. 1: Eingangsmenue zur Textauswahl

BYUVIEW

		Select Word
(F10) ReView: View last seen list	industriezweigen	3
(F20) Combine: Use with other word(s)	industry	1
	infoige	3
	informatik	2
Press Enter to see (or type word) ---	information	13
	informationen	29
	informations	1
(Shift F4) Options Menu	informations-	7

Select Word HELP 1

Type in a word. Press ENTER to view the highlighted word.
Press [+], [-] to highlight the next/previous word.
[PgUp], [PgDn] to see the next/previous 8 words.
[Home], [End] to see the start/end of the list of words.
C / I

Abb. 2: Bildschirm zur Wortauswahl und Belegstellensuche

zum Auffinden von Fundstellen zu einem bestimmten Themenbereich kann ein zuvor erstellter Thesaurus benutzt werden. Durch Auswahl eines bestimmten Oberbegriffs werden sämtliche Fundstellen der unter ihm bezeichneten Begriffe angezeigt. Der Thesaurus kann aber außer nach thematischen Gesichtspunkten auch nach semantischen oder syntaktischen Gesichtspunkten erstellt sein, so daß z.B. ein Thesaurus, der die Kategorie "Pronomen" enthält, dazu benutzt werden kann, Fundstellen der entsprechend aufgeführten Pronomina zu extrahieren. In ähnlicher Weise können für eine gegebene Grundform die Fundstellen auch für all ihre flektierten Formen ermittelt werden.

Wenn die Wörter spezifiziert sind, die im Text gesucht werden sollen, werden die Textausschnitte zu allen Fundstellen in eigenen Bildschirmfenstern gezeigt. Die Größe der Fenster und damit die Anzahl der Kontextzeilen kann bis zur gesamten Bildschirmgröße erweitert werden. Der Text kann in jedem Fenster geblättert werden. Gibt es mehr Fundstellen, als auf einem Bildschirm gezeigt werden können, dann sind Fortsetzungsseiten abrufbar (Abb. 3).

3. Nutzung der Retrieval-Ergebnisse

Die Fundstellen können auf unterschiedliche Weisen mit dem System weiterbearbeitet werden. Doppelte Fundstellen können gelöscht oder besonders interessierende Belegbeispiele sofort gedruckt werden. Die Seitenformatierung und der Umfang des Kontextes wird dabei individuell festgelegt. Der Druckvorgang erfolgt unabhängig von den übrigen Systemfunktionen.

Die ausgewählten Beispiele lassen sich auch in einer Datei zur weiteren Bearbeitung zwischenspeichern. Dadurch können Belegbeispiele in andere Bearbeitungssysteme, z.B. Datenbanken oder Textverarbeitungsprogramme, transferiert werden. Die Ergebnisse einer Textrecherche können auch in Form der Referenzen auf die Textstellen in einer Datei gespeichert und zu einem späteren Zeitpunkt mit BYUVIEW weiter analysiert werden.

Durch Zuladen einer Word Processor Shell (z.B. WordPerfect) ist eine gekoppelte Bearbeitung der Texte mit beiden Systemen möglich, ohne daß ein System verlassen werden muß.

4. Statistische Auswertungen des Systems

Mittels der Programmfunktionen zur Bearbeitung der Frequenzdistribution lassen sich statistische Werte über die Verteilung von Wörtern in spezifischen Teilen eines Textes oder eines Korpus ermitteln. So lassen sich z.B. gattungsspezifische Abweichungen der Häufigkeiten von Prono-

men, Konjunktivformen o.Ä. leicht sichtbar machen.

5. Die Indexierung

Die Indexierung des Textes erfolgt mit einer eigenen Programmkomponente (BYU-INDEX), die es erst ermöglicht, daß die mit BYUVIEW gesuchten Belegstellen so schnell gefunden und angezeigt werden können. Durch diese vorbereitende Komponente des Systems werden aber auch viele zusätzliche Such- und Verbindungsmöglichkeiten geschaffen. Im einzelnen sind dies:

- Spezifizierung von Referenzebenen
- Anpassung an nationale Zeichensätze
- Festlegung der Sortierreihenfolge
- Zuweisen eines Zeichentyps
- Verwendung einer Stopwortliste
- Generierung von Lexika und Thesauri

5.1 Referenzebenen im Text

BYU erlaubt eine dreigliedrige Unterteilung eines Textes (z.B. Buch, Kapitel, Abschnitt). Diese Unterteilung kann mit VIEW zur Suche referenzbezogener Fundstellen benutzt werden. Wird ein Text mit einem Textverarbeitungsprogramm für BYU erfaßt, so kann der Benutzer die verschiedenen Referenzebenen mit gesonderten Kodierungen in seinen Text einbauen. Einen bereits vorliegenden maschinenlesbaren Text nachträglich mit solchen Referenzangaben zu versehen, wäre aber u.U. eine sehr mühsame und zeitraubende Arbeit. Aus diesem Grund kann eine solche Untergliederung des Textes von BYUINDEX auch automatisch vorgenommen werden, wobei als Kriterien der Seitenumbruch, der Zeilenumbruch und/oder die Satzzeichen benutzt werden. Zur Vorbereitung des zu verarbeitenden Textes ist sonst nur noch zu beachten, daß er keine Kontrollzeichen des Textverarbeitungsprogramms enthält, sondern ein 'reiner' ASCII-File ist.

5.2 Zeichen und was BYU damit anstellt

Das BYU Text Retrieval System ist so konzipiert, daß Texte aller Sprachen (soweit sie lateinische Buchstaben benutzen bzw. in einer Transkription vorliegen) verarbeitet werden. Dies wird durch eine externe Zeichensatztafel erreicht, die den text- und sprachabhängigen Erfordernissen angepaßt werden kann. Die Zeichensatztafel kann in ihren folgenden Bestandteilen modifiziert werden: Spezifikation der Zeichen des Alphabets, Spezifikation der gewünschten Sortierreihenfolge, Klassifizierung weiterer Zeichen (Satzzeichen,

ESPRIT-TEXT (deutsche 7:2 Reference List
 (zu Preisen von 1980) zu den groessten Fertigungsbereichen der Welt gehoeren
 * P 54 * P* (7.) * Der Umgang mit \$FTN Information wird zum wichtigsten
 Einzelbereich der Beschaeftigung * Das Statistische Amt der Vereinigten
 ESPRIT-TEXT (deutsche 10:Heading
 zur Ausarbeitung eines FuE-Programms mit langen Vorlaufzeiten namens ESPRIT -
 European Strategy Programme for R&D in Information Technology (Europaeisches
 Strategisches Programm fuer Forschung und Entwicklung auf dem Gebiet der
 ESPRIT-TEXT (deutsche 25:9
 und industrielle Software-Erzeugung * P 168 * P* (3.) * Fortgeschrittene
 Informationsverarbeitung (AIP - advanced information processing) [Ziel ist
 hierbei die Schaffung der Grundlage fuer die industrielle Nutzung des
 ESPRIT-TEXT (deutsche 26:Heading
 von ESPRIT und ihrer Beziehung zu nationalen und internationalen Normen \$EA
 einem Informationsaustauschsystem (IES - Information Exchange System), um
 durch Erleichterung wechselseitiger Kommunikation die gute technische
 ESPRIT-TEXT (deutsche 28:11
 Dollar (zu Preisen von 1980) weltweit der groesste Fertigungssektor sein *
 P 216 * P* (22.) * Umgang mit \$FTN Information wird zum wichtigsten
 Einzelbereich der Beschaeftigung * Das statistische Amt der Vereinigten
 ESPRIT-TEXT (deutsche 28:11
 Informationsbereich taetig waren * Dabei sind neben der Fertigung auch alle
 anderen Beschaeftigungen, die sich mit Information befassen, eingerechnet *
 Die europaeischen Zahlen sind aehnlich * In der IT- Herstellung alleine sind
 Pg 1 of 3 F1 FreqDist ShiftDel: PrefHead+DupRef

Abb. 3 Referenzliste zum Suchbegriff 'Information'

0 Preface:1100
 P1100 IT SOM EN FAKTOR VED ØKONOMISK
 ÆNDRING
 P1101 14. Informationsteknologi er
 vel nok det område, der
 P1102 udvikler
 P1103 sig hurtigst inden for

Dictionary HELP

Move Words: [+], [-], Type, [Esc]

Definition: [PgUp], [PgDn]

BookShelf: [F5]

Definition of Highlighted Word

nok 1 => schon 1
 :SEMF # = CONCESSIVE
 Det skal nok passe.
 Das wird schon stimmen.
 IDzu:dtok

Dictionary Name: ESPRITDK

Words in Dictionary

mulighed
 nogen
 --) nok 1
 nok 2
 nu

PgUp/PgDn] 1/4 page UP/DOWN

page 1 | 103 Return

Abb. 4 Verbindung zwischen Referenzliste und Lexikon

onderzeichen, Delimiter). Standardzeitensatztabellen für die Sprachen in der europäischen Gemeinschaft liegen vor, insbesondere der Zeichensatz für eine Transkription des Griechischen.

3. Ein 'Muß' für jedes Konkordanzprogramm: die Stopwortliste

Eine Stopwortliste erlaubt die Spezifikation von Wörtern, die nicht in die Konkordanz aufgenommen werden sollen. Eine kleine Liste hochfrequenter Wortformen für die Sprachen in der europäischen Gemeinschaft sind dem System bereits beigefügt. Die Einschränkung auf einzelne Wörter, für die eine Konkordanz erstellt werden soll, wird durch einen zusätzlichen Bearbeitungsschritt ebenfalls erreicht.

4. Was tun mit Sekundärinformationen zum Text?

Mit BYUVIEW kann eine Verbindung zwischen Textstellen und Informationen aus einem Lexikon oder einem Thesaurus hergestellt werden. Hierzu muß das Lexikon bzw. der Thesaurus genauso wie der Text durch BYUINDEX bearbeitet werden. Ein Thesaurus im BYU-System benutzt zwei Ebenen zur Darstellung der Information. Die primäre Ebene wird aus einer Liste von Wörtern gebildet, die in den Texten gesucht werden sollen oder die z.B. grammatische Kategorien oder Themenbereiche bezeichnen. Jedem Wort der primären Ebene können auf untergeordneter sekundärer Ebene eine Liste von Wörtern und zusätzlich ein Definitionsteil zugeordnet werden. Durch die Möglichkeit, ein Wort der sekundären Ebene auch auf der primären Ebene zu finden, können verketterte Listen dargestellt werden. Ein Lexikon im BYU-System besteht aus einer Wortliste über welche ein Lexikoneintrag ausgewählt wird, und einem Definitionsteil, in dem jede Art von Information zum Worteintrag aufgenommen werden kann. Der Aufruf des Lexikons bewirkt das Anzeigen des Definitionsteils zum aktuell in einem Text ausgewählten Wort, mit der Möglichkeit im Lexikon zu blättern (Abb. 4). Gegebenenfalls kann auch auf weitere Lexika zugegriffen werden.

Als Beispiel soll kurz die Nutzung dieser Schnittstelle in einem Forschungsprojekt am Institut für Kommunikationsforschung und Phonetik (IKP) beschrieben werden. Dort werden im Rahmen des Projekts zur Entwicklung eines maschinellen Übersetzungssystems (EUROTRA) Transferlexika für verschiedene Sprachpaare auf der Grundlage eines ausgewählten Textkorpus erstellt. Die lexikalischen Einträge werden durch ein eigenes Lexikonsystem (MULI) bearbeitet, das eine Schnittstelle enthält, die es ermöglicht, die lexikalischen Informationen aus den Transferlexika in das BYU-System zu integrieren. Das BYU Text Retrieval System gibt dem Lexikographen wichtige Hilfsmittel in die Hand, um Text und Lexika abzugleichen und aneinan-

der anzupassen. Er kann nicht nur die Lesarten der verschiedenen Wortformen in den Texten feststellen, sondern er hat auch einen direkten Zugriff auf die lexikalischen Einheiten jedes Transferlexikons und so die Möglichkeit, alle vorliegenden Einheiten miteinander zu vergleichen und auf Vollständigkeit oder Konsistenzen zu prüfen. Liegt das Textkorpus in den verschiedenen Sprachen vor, so kann er diese Versionen untereinander vergleichen.

6. Zusammenfassung

Das BYU Text Retrieval System unterscheidet sich von anderen Konkordanz-Programmen für den Personal Computer durch seinen breiten Funktionsumfang. Es ist auf eine schnelle interaktive Bearbeitung von Texten ausgerichtet. CORDA [1] z.B. hat seinen wesentlichen Anwendungsbereich in der Generierung und Bearbeitung von Druckvorlagen für Textindizes und Konkordanzen. FATRAS [2] und IAC [3] beschränken sich auf die Fundstellensuche. Interaktive Verbindungen zu Lexika oder Thesauri lassen sich mit diesen Systemen nicht herstellen. Die Einbindung von externen Wortlisten erlaubt es, das Text Retrieval System zu einem Textanalyse-System im linguistischen Sinne auszubauen. Am Institut für angewandte Kommunikations- und Sprachforschung (IKS) [4] werden Schnittstellen zu einer automatischen Lemmatisierung und Lexikonerstellung ausgebaut. Utilities zur Zeichen- und Wortlängenstatistik oder zur Sortierung nach Wortendungen sind dort bereits verfügbar. Besonders hervorzuhebende Eigenschaften des BYU Retrieval Systems sind die umfangreichen Such- und Analysemöglichkeiten, die das System mit einer komfortablen und leicht verständlichen Benutzeroberfläche anbietet. Es ist insgesamt ein Instrument, das in dieser oder ähnlicher Form in Zukunft ein unverzichtbarer Bestandteil eines rechnergestützten Arbeitsplatzes für Lexikographen und Sprachwissenschaftler sein wird. Aber auch jede andere Tätigkeit, bei der man wesentlich darauf angewiesen ist, Text zu analysieren und zu erschließen, wird durch dieses System hilfreich unterstützt.

Anmerkungen:

- [1] CORDA ist ein Konkordanzprogramm des ECO Instituts Pielmühle, vgl. hierzu auch den Beitrag von R.v. Ammon im LDV-Forum 2/85
- [2] FATRAS ist ein Text Retrieval System, das an der Universität von Montreal, Kanada von L. Boivin und P. Bratley entwickelt wurde.
- [3] IAC ist ein Konkordanz-System, das von E. Gildenstein an der Universität Göttingen aufgebaut wurde.
- [4] Informationen und Bezugsunterlagen zum BYU Text Retrieval System erhalten sie über IKS, Poppelsdorfer Allee 47, 53 Bonn 1, Tel. 0228/735645

Anze

GRUNDLAGEN DER KOMMUNIKATION FOUNDATIONS OF COMMUNICATION

MAX J. CRESSWELL

Die Sprachen der Logik und die Logik der Sprache

Oktav. XII. 451 Seiten. 1979. Kartoniert DM 56,- ISBN 3 11 004923 6 de Gruyter Studienbuch

Das Buch zeigt, daß natürliche Sprachen wie Deutsch und Englisch genauso beschrieben werden können wie formale Sprachen vom Typ der Kategorialsprachen. Semantische Erscheinungen wie Vagheit, Kontextabhängigkeit der Bedeutung und Metaphernbildung lassen sich auf diese Weise präzise erfassen. Standardlehrbuch, zum Selbststudium geeignet.

PIRMIN STEKELER-WEITHOFER

Grundprobleme der Logik

Elemente einer Kritik der formalen Vernunft

Oktav. XXX. 578 Seiten. 1986. Kartoniert DM 48,- ISBN 3 11 010491 1 de Gruyter Studienbuch
Gebunden DM 92,- ISBN 3 11 010492 X

Aus dem Inhalt: Platonisch-aristotelische Logik: Begriff und Satz, terminologisches System und Syllogistik. – Freges Wahrheitsfunktionale Logik: Atomismus/Satzholismus, Konstruktion

Analyse, Wahrheitsbegriff und Antinomien, Gleichheit, Gegenstand und Kennzeichnung, inhaltlicher und formaler Schluß. – Formale Argumentationslogik: Konstruktive, Operative und Dialogische Logik, Reallogik.

SHALOM LAPPIN

Sorts, Ontology, and Metaphor

The Semantics of Sortal Structure

1981. Octavo. X. 175 pages. Bound DM 72,- ISBN 3 11 008309 4

Examination of semantic anomalies and their relationship to the structure of ontological categories in the speaker's world of discourse. Application of the results to natural logic and the theory of metaphor.

Causal Theories of Mind

Action, Knowledge, Memory, Perception, and Reference

Edited by Steven Davis

1985. Large-octavo. X. 421 pages. Cloth DM 178,- ISBN 3 11 007730 2 Library Edition

Contents: Papers by D. Davidson, A. I. Goldman, M. Swain, C. B. Martin and M. Deutscher, H. P. Grice, S. Kripke, K. S. Donnellan, and G. Evans, on action, knowledge, memory, perception and reference. Introduction contains explanatory background for Goldman on action and Kripke on reference. Name and subject indices and extensive bibliography.

Preisänderungen vorbehalten

Walter de Gruyter



Berlin · New York

GRUNDLAGEN DER KOMMUNIKATION FOUNDATIONS OF COMMUNICATION

Sublanguage

Studies of Language in Restricted Semantic Domains

1982. Large-octavo. VI, 240 pages. Cloth DM 96,- ISBN 3 11 008244 6 Library Edition

The often distinctive use of language which characterizes certain fields of technology and science has given rise to the notion of specialized sublanguages within a natural language. This book explores various aspects of the notion of sublanguages - including the historical development of the concept, its use in applied fields such as machine translation and information retrieval, and its role in the description of the structure of natural languages. Although most of the investigations described have been carried out in the U.S. and Canada, there is also a description of the considerable research conducted in the Soviet Union involving both applied and theoretical work on sublanguages. The book concludes with a paper by Zellig Harris, who pioneered the investigation of sublanguages, and who here examines the relation between sublanguages and discourse. The editor, who have been engaged in sublanguage research for a number of years, have written an introduction summarizing work in this relatively undeveloped field of linguistics and tying the various articles together.

Maschinelle Sprachanalyse

Beiträge zur automatischen Sprachbearbeitung I

Herausgegeben und eingeleitet von Peter Eisenberg

Oktav. 272 Seiten. 1976. Kartoniert DM 36,- ISBN 3 11 005722 0 de Gruyter Studienbuch

Anhand der Originalbeiträge und Übersetzungen wichtiger auf Englisch und Schwedisch erschienener Arbeiten wird ein Überblick über die neuere Entwicklung und den gegenwärtigen Stand der maschinellen Sprachanalyse gegeben. Schwerpunkt liegt bei Arbeiten, die linguistisch fundiert sind oder selbst einen Beitrag zur Entfaltung linguistischer Theorien geben.

Semantik und künstliche Intelligenz

Beiträge zur automatischen Sprachbearbeitung II

Herausgegeben von Peter Eisenberg

Oktav. VIII, 244 Seiten. 1977. Kartoniert DM 36,- ISBN 3 11 005721 2 de Gruyter Studienbuch

Die neuere Entwicklung der künstlichen Intelligenz wird anhand bisher überwiegend unveröffentlichter Beiträge aus dem angelsächsischen Raum dokumentiert. Hauptgesichtspunkt für die

Auswahl: Unter welchen Bedingungen ist davon zu sprechen, daß eine Maschine Sprache versteht? Wie können Sprachverstehensprozesse simuliert werden? Was leisten linguistische Theorien für diesen Zweck?

MANFRED WETTLER

Sprache, Gedächtnis, Verstehen

Oktav. X, 331 Seiten. 1980. Kartoniert DM 36,- ISBN 3 11 007971 2 de Gruyter Studienbuch

Das Buch gibt eine kritische Einführung in die Ergebnisse der kognitiven Psychologie, linguistischen Semantik und Erforschung künstlicher Intelligenz. Es untersucht die Frage, welche Rolle begriffliches Wissen beim Verstehen von Sprache spielt, und stellt die wichtigsten semantischen Modelle zur Beschreibung von Wissen dar: assoziative Strukturen, semantische Netze, konzeptuelle Abhängigkeitssysteme, frame-Theorie. Diese Modelle werden fortentwickelt, so daß sie in die vom Autor entworfene Theorie des Sprachverstehens passen.

Preisänderungen vorbehalten

Walter de Gruyter



Berlin · New York

GLDV-Jahrestagung

DAS PROGRAMM DER GLDV-JAHRESTAGUNG 1986 IN GÖTTINGEN

Folgende Vorträge wurden auf der GLDV-Jahrestagung 1986 (25.-27. Februar) gehalten: Mart Ren
"Die mathematischen Mittel für die Formalisierung der historischen Linguistik"; C. Schneider und C. Wor
Hacker: "PADOK. Test und Vergleich von Texterschließungssystemen für das Deutsche Patent- und Fa
formationssystem"; Jürgen Krause: "Informationslinguistische Komponenten von Informationssystemen
Rahmstorf: "Graphische Darstellung binärer semantischer Netze"; Reinhard Baukowitz: "Algorithmen
Datenstrukturen für den Entwurf eines natürlich-sprachlichen Frage-Antwort- und Übersetzungssysteme
Hoepelmann und J. Machate: "Dialogtableaux. Theorembeweisen; DB/Abfrage und natürliche Spra
Rainer Gimnich: "Yur formalen Definition der Übergangnetzwerke"; H. Rüggeberg: "Die GPSG mit
neuen Merkmalstheorie der 'FOOT features', angewandt auf einige Fälle des Spanischen"; Jürgen Ro
ven: "Eine selbstlernende Grammatik"; Klaus Netter: "Verbalkomplex und Wortstellung im Deutsch
Karin Dierks und Christian Zinko: "Computerunterstützte Emendierung des Hethitischen"; Bernhard
"Retrieval in mundartlichem Sprachmaterial ohne feste Wortgrenzen"; Heiny Schay: "Lemmatisierung ph
phischer Texte des 18. Jahrhunderts"; Gerhard Jaritz: "Datenbankdienst für terminologisch frei beschrie
mittelalterliche Bildquellen"; Andras Kornai: "Finite State Semantics"; Manfred Thiel: "Determinismus
Instantiierung"; Heinz-Dirk Luckhardt: "Normalisierung Deutscher Oberflächenstrukturen mit Controlle
tive Procedures"; Alois Baumer: "Ein Algorithmus zur Verberkennung im Englischen"; Günther Thurn
"Linguistische Verfahren für ein speech understanding System"; Dietmar Roesner: "Von Titeln zu Te
Zur Entwicklung des Textgenerators SEMTEX"; Andreas Blumenthal: "Vorüberlegungen zur Möglichkeit
Computereinsatzes bei der Abfassung lexikographischer Bedeutungserläuterungen"; Gert Frackenpohl: "A
Aufbau einer Lexikondatenbank des Vocabulari Rumantsch Grischun"; H. Huonker: "Reflexionen zum
griff 'Wortdatenbank' in der Linguistischen Datenverarbeitung"; Heinz J. Weber: "Maschinelle Überset
von EG-Nachrichten (Deutsch- Englisch/Französisch). Vorschläge für einen text-orientierten Transfer"; D
Schmidt: "Syntaktische Valenz in einem MÜ-System"; Erich H. Steiner: "Die Zuweisung satzsemantische
Rollen im Rahmen des MÜ-Projekts EUROTRA-D"; Cornelia Zelinsky-Wibbelt: "Entwurf eines System
mantischer Merkmale für die maschinelle Übersetzung"; Karin Haenelt et al.: "Ein sprachtheoretisches M
geisteswissenschaftlicher Texterschließungsverfahren als Grundlage einer Softwareentwicklung. Ein Be
des Projekts PRO TEXT"; Barbara Schäfer et al.: "Der Einsatz der EDV im Bereich der bibliographis
Dokumentation am Beispiel der Romanischen Bibliographie (RB)"; Birkenhauer: "Datenbankprobleme e
Übersetzerzentrums"; Thomas Stadler: "EDV/gestützte Editionsprojekte im SFB 'Wissen des Mittelalt
Peter Schreiner und Renate Soehnen: "Stilistische Analyse von Sanskrit/Texten".

GLDV-JAHRESTAGUNG 1986 - PERSÖNLICHE EINDRÜCKE

Vom 25. bis 27. Februar 1986 fand im Zentralen Hörsaalgebäude des Geisteswissenschaftlichen Zentrums der Universität Göttingen die Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung (GLDV) statt. Titel der Tagung: "Linguistische Datenverarbeitung in den Geisteswissenschaften"

"Die Geisteswissenschaften haben endlich ein Recht auf angemessene technische Hilfsmittel". In diesen Satz gipfelte die Begrüßungsansprache der GLDV- Vorsitzenden Brigitte Endres-Niggemeyer. Daß sich darüber Computerlinguisten mit Geisteswissenschaftlern gemeinsam auseinandersetzen, ist sicherlich begrüßenswert. Eine interdisziplinäre Tagung also - gleichzeitig Experiment und Notwendigkeit.

Experimente haben das Recht, nur teilweise zu gelingen. Nicht vollständig gelungen war für mich die zeitliche Organisation des Programms. Fast durchweg liefen Vorträge aus verschiedenen Disziplinen parallel und die Wahl verfiel sicherlich manchen Teilnehmer dazu, in seiner eigenen Schublade stecken zu bleiben.

Die leider sehr knapp bemessene Vortragsdauer (20 Minuten plus 10 Minuten Diskussion) hat sich sehr negativ auf die inhaltliche Qualität der Vorträge ausgedrückt. Oftmals blieb nach einem Vortrag der Eindruck zurück, daß das Wesentliche eigentlich nicht kommen mußte, daß der Vortrag noch nicht fertig geworden war.

"Ich fühle mich jetzt als Klassiker". Dieser Ausspruch eines Computerlinguisten, nachdem er festgestellt hatte, daß sein vor zehn Jahren entwickeltes System diesmal von einem Geisteswissenschaftler erfunden worden war, steht wohl einerseits für, daß nicht die Zeit fuer Revolutionen war, ergibt aber andererseits die Notwendigkeit zur interdisziplinären Auseinandersetzung.

Leider konnten inhaltliche Schwächen, so vorhanden, nur selten durch brillante Darstellungen überspielt werden. Die Konkurrenz im Wettbewerb um die schlechteste Folie war auf dieser Tagung ebenfalls sehr groß. Mir ist es schon sehr unverständlich, warum ein Referent überhaupt eine Folie auflegt, von der er mit der größten Selbstverständlichkeit sagt, daß man sie wahrscheinlich nicht lesen und man deshalb auch gar nicht sehen könne, was er zeigen wolle. Vielleicht sollten die Leute, denen dieser Schuh passt, sowohl ihre Sache als auch ihr Publikum etwas ernster nehmen.

ZWEI ROUND TABLES ZU LDV, PHILOLOGISCHE DATENVERARBEITUNG UND LEXIKOGRAPHIE

Vom 25. bis 27. Februar hat in Göttingen die Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung (GLDV) stattgefunden. Dem Gesamthema der Tagung entsprechend ("Linguistische Datenverarbeitung in den Geisteswissenschaften") wurden zwei Round Tables durchgeführt: "Computerlinguistik und Philologische Datenverarbeitung: Kontroversen und Perspektiven" und "Brauchen Lexikographen wirklich die Werkzeuge der Lexikographischen Datenverarbeitung?".

Die Teilnehmer der Tagung gehörten einem breiten Spektrum an, das von Sprachwissenschaftlern und Philologen bis hin zu Übersetzern und Informatikern reichte. Die erste Diskussion ließ sich in drei Phasen teilen: Zuerst nahmen die Diskussionsteilnehmer gegenüber der Datenverarbeitung als Werkzeug für den Philologen Stellung. Dann wurden diese Stellungnahmen konträr diskutiert. Zum Abschluß wurde versucht, Perspektiven aufzuzeigen.

Nach den Aussagen der Diskussionsteilnehmer leistet die EDV dem Philologen eine bedeutende Hilfe, indem sie ihm die Routine- oder Trivialarbeiten (die sogenannten "niederen" Arbeiten) erleichtert. Zum Beispiel lassen sich das Kollationieren von Texten bei Editionsarbeiten sowie der Vergleich überlieferten Varianten computerunterstützt wesentlich leichter handhaben. Große Datenmengen, die vielleicht ohne die Möglichkeit einer maschinellen Bearbeitung nicht zu bewältigen wären, stellen mit EDV kein unüberwindliches Problem mehr dar. Das heißt aber nicht, daß für den Philologen dadurch auch die eigentlich wissenschaftliche Arbeit

Die also insgesamt eher ruhige Atmosphäre der Tagung wurde ausgesprochen positiv "gestört" von der lebhaften Round-Table-Diskussion unter dem Leitthema

"Computerlinguistik und philologische Datenverarbeitung: Kontroversen und Perspektiven" geleitet von Peter Hellwig. Nach den wie üblich zu langen Eingangs-Statements war es besonders Herrn Schweisstal zu verdanken, daß sie sich nach durchaus nicht polemikfreien Angriffen auf die einzelnen Teilnehmer zu einer engagiert und kontrovers geführten Diskussion entwickelte. Während die Philologie bereit ist, "ihr Recht auf angemessene technische Hilfsmittel" mehr als in Anspruch zu nehmen, d.h. eigene Aktivitäten entwickelt, rufen die Computerlinguisten verstärkt nach einer eigenen Standortbestimmung. Angst vor dem Verlust einer Domäne oder Ausdruck des Willens zur Auseinandersetzung? Die Diskussion sollte fortgesetzt werden.

Barbara Kienitz-Vollmer, Technische Hochschule Darmstadt

erleichtert wird. Für die Fragestellung und die Behandlung des Themas und damit die Qualität des erzeugten Werkes trägt nur er die Verantwortung. Der Zeitaspekt spielt nach Meinung der Referenten eine große Rolle bei der Entscheidung für die EDV. Jedoch liegen ihre Dienste nicht nur in den "gewonnenen" Tagen oder Stunden Arbeit, sondern vielmehr in der Verbesserungsmöglichkeit der bestehenden Arbeit, welche der Computer anbietet, sowie in der totalen Zugänglichkeit und der universellen Kombinierbarkeit der maschinell erfaßten Daten.

Der aus dem Plenum heraus geäußerte Verdacht, daß die völlige Hinwendung der Geisteswissenschaften an den Computer eine Leistungsprinzipforderung widerspiegeln könne sowie eine Konkurrenz zu den Naturwissenschaften darstelle, wurde zurückgewiesen. Nur durch die Kenntnis der Arbeitsweise anderer Zeiten im gleichen Arbeitsgebiet, z.B. bei der Lexikographie, kann man die tatsächliche Leistung der Datenverarbeitung für den Philologen abschätzen. Diese erste Diskussion wurde ausdrücklich mit dem allgemeinen Wunsch abgeschlossen, daß Philologen, Sprachwissenschaftler, Informatiker und Programmierer zum gegenseitigen Vorteil enger zusammenarbeiten sollten.

Der zweite Diskussionstag verlief, im Vergleich zum ersten, viel pragmatischer. Hier wurden die praktischen Ergebnisse des Computereinsatzes bei der philologischen-lexikographischen und sprachwissenschaftlichen Arbeit dargestellt, sowie die daraus resultierenden Probleme. Der kommerzielle Faktor

hat sich für den Einsatz der DV als ausschlaggebend erwiesen. Das wird besonders deutlich am Beispiel des Dudenverlags, wo, wegen der enormen Kosten der Datenerfassung aller vorhandenen Einträge, der traditionelle Karteikasten noch relevant ist. Eine wesentliche Hilfestellung bietet die DV dem Übersetzer, wenn dieser seinen zu übersetzenden Text automatisch mit einem Wörterbuch abgleicht. Dabei wird ihm selbst bei seiner Arbeit geholfen, ohne daß seine persönliche Kreativität am Arbeitsplatz verloren geht. Denn von ihm eingeführte neue Wörter bleiben gespeichert und erweitern das Wörterbuch. Auch in der zweiten Diskussion hob man den zeitlichen Aspekt als besonders wesentlich für den Computereinsatz hervor und ebenso plädierte man für

eine intensivere Kommunikation der Lexikographen mit den Informatikern.

In beiden Diskussionen kam mehrfach Ausdruck, daß die EDV für die philologischen und sprachwissenschaftlichen Arbeiten eine Dienstleistung darstellt. Diese kann aber wegen des menschlichen Faktor ersetzt, sondern die menschliche Intelligenz simulieren und dazu Ansporn für weitere Forschungen sein.

Charola Argyriadis, Max-Planck-Institut für Europäische Rechtsgeschichte, D-6000 Frankfurt am Main, Freiherr-vom-Stein-Straße 7

INFORMATIONSMARKT AUF DER GLDV-TAGUNG

Der auf der GLDV-Jahrestagung '86 eingereichte Informationsmarkt und dessen "Nachbereitung" hat für Studenten eine Reihe wesentlicher Informationen an den Tag gebracht. In der Rubrik "Arbeitskreise" ist unter dem Titel "WISSENSWERTES VOM INFORMATIONSMARKT IN GÖTTINGEN", vom Arbeitskreis "Ausbildung und Berufsperspektiven" zusammengestellt, das Wichtigste nachzulesen.

Podiumsdiskussion: Brauchen Lexikographen wirklich die Werkzeuge der Linguistischen Datenverarbeitung

- Teilnehmer:
- Frau Birkenhauer, Straelen (NRh.), Europäisches Uebersetzerkollegium
 - Herr Buksch, Bruehl, Freier Woerterbuchautor
 - Herr Frackenpohl, Universitaet Koeln
 - Herr Drewek, Stuttgart (Verlag Klett)
 - Herr Hellmann, Mannheim, (Institut fuer deutsche Sprache, IdS, Mannheim)
 - Herr Scholze-Stubenrecht, Mannheim (Verlag Bibliographisches Institut, Dudenredaktion)
 - Frau Storrer, Heidelberg, Universitaet und Firma IBM, Projekt LEX

Buksch), bis zur Erstellung von Hilfsgrammatischen, Konkordanzen, Indizes und Wörterdatenbanken (Frackenpohl). Im Rahmen der Arbeiten am LEX-Projekt werden auch Wörterbucheinträge für Verben, Adjektive und Substantive erstellt, die die Eingabe eines sprachverarbeitenden Systems darstellen; diesen Blickwinkel könnte man statt "Lexikographie mit dem Computer" eher "Lexikographie für den Computer" nennen (Storrer).

Die Leitfrage der Diskussion lässt sich in zweifacher Hinsicht stellen. Zum einen kann man danach fragen, ob und wie Lexikographen die Hilfsmittel 'gebrauchen', die ihnen die linguistische Datenverarbeitung anbietet, zum anderen aber auch danach, ob sie überhaupt solcher Werkzeuge bedürfen bzw. welche Art von Werkzeugen sie 'brauchen'. Der Blickwinkel der Diskussionsteilnehmer, die zu den beiden Aspekten ihren Standpunkt vertraten, reicht dabei von der Benutzung der linguistischen Datenverarbeitung als Hilfsmittel für Satz, Umbruch und Korrektur von Wörterbüchern (z.B.

Verfasser von Wörterbüchern und Koordinatoren lexikographischer Projekte nutzen die linguistische Datenverarbeitung bei der Gestaltung des - rein manuell erstellten Produkts Wörterbuch, jedoch fallweise auch bei der Koordination lexikalischer Beschreibungen aus verschiedener Hand, oder zur Verwaltung von Belegen aus verschiedenen Quellen (beim Klett-Verlag werden z.T. die Wörterbuchartikel von 20 bis 30 verschiedenen Autoren gesammelt, Drewek). Vorhandene große Belegsammlungen, die vom Zettelkasten auf Datenbankform umzustellen, ist außerordentlich aufwendig und teuer (Scholze-Stubenrecht: die Mio. Belege der Dudenredaktion umzustellen ist momentan nicht realisierbar), jedoch auf der Basis eines maschinenles-

...fuegbaren Corpus gearbeitet wird (wie
 3. im Rahmen des Ost/West-Wortschatzver-
 eichs am IdS, Mannheim), ist gerade die
 schnelle Belegverwaltung und -verar-
 tung das Hilfsmittel der Wahl bei der
 Stellung von Woerterbuchartikeln
 (Illmann).

...schliesslich die Integration der Woer-
 terbuchkomponente in ein sprachverarbei-
 endes System die Einhaltung eines vorge-
 benen Schemas fuer die Mikrostruktur der
 Woerterbuchartikel vorschreibt, laesst
 sich mithilfe des Rechners die lexikogra-
 phische Arbeit auf Konsistenz und Voll-
 staendigkeit hin ueberpruefen (Storrer).
 ...der Arbeit des Uebersetzers fallen
 Notizen an, die bestimmte Angaben zu Woer-
 tern betreffen, die in den zu bearbeiten-
 den Texten vorkommen; solche Angaben im
 Textzug auf Konnotation, sondersprachliche
 Bedeutung, etc. fehlen oft genug in den
 vorhandenen Woerterbuechern oder sind dort
 nicht hinreichend praezise. Diese Notizen
 koennen am Europaesischen Uebersetzerkolle-
 kium mittels eines Textsystems verwaltet
 werden ueber die Editor-Funktion anderen
 Nutzern zugaenglich gemacht werden
 (Kirkenhauer).

...der Arbeit des Uebersetzers kommt es
 drauf an, dass die Verarbeitung solcher
 Notizen keinen besonderen zusaetzlichen
 Arbeitsaufwand darstellt, der sich auf die
 eigentliche Uebersetzertaetigkeit nachteil-

...lig auswirken wuerde. Wichtig ist hier
 also hoher Bedienungskomfort bei gering-
 ster Einlernzeit. Insgesamt erscheint
 neben der Kontrolle der Gestaltung von
 Woerterbuechern auch die ihrer inhaltli-
 chen Konsistenz als wuensenswert.

Ein weiteres Einsatzgebiet fuer die Daten-
 verarbeitung liegt in der Verwaltung von
 Belegmaterial; die Kombination von Lexikon
 und Textzugriff, interaktive Konkordanz-
 und Indexprogramme u. dgl. erlauben die
 Verwendung von Textkorpora (z.B. Magnet-
 baender von Zeitungsverlagen) als Basis
 der lexikographischen Beschreibung
 (Frackenhohl). Solche Moeglichkeiten wer-
 den bislang nur in geringem Masse genutzt
 und koennten durch den Einbezug von Ver-
 fahren zur Lemmatisierung oder zur Wort-
 artenerkennung und -trennung, die in der
 Entwicklung sind, in Zukunft an Bedeutung
 gewinnen.

Im Augenblick stehen jedoch noch weniger
 die zukuenftigen Moeglichkeiten, die die
 Hilfsmittel der Datenverarbeitung dem
 Lexikographen eroeffnen, im Vordergrund
 des Interesses, als vielmehr technische
 Probleme: am Markt erhaeltliche Textverar-
 beidungsprogramme sind z.B. fuer die
 Zwecke des Lexikonzugriffs ungeeignet. Um
 vorhandene Programme und Geraete fuer
 lexikographische Zwecke nutzen zu koennen,
 bedarf es (noch) umfangreicher technischer
 Entwicklungsarbeiten.

Heid, Universität Stuttgart, Institut für Linguistik

ge

...rogrammsysteme
 ...r die
 ...rtgeschrittene
 ...xtverarbeitung

Ein Microcomputer-Programmsystem für Autoren, Handbuch-Schreiber u.a.

ECOINDEX

für die Herstellung von Stichwortverzeichnissen, Indizes, Sach- und Personenregistern

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

ECOINDEX verarbeitet auch Dateien von Textverarbeitungssystemen wie WordStar (von MicroPro) und MS-WORD (von Microsoft);

ECOINDEX kann auch down loaded Dokumente verarbeiten, die aus einem Informationssystem recherchiert wurden.

ECOINDEX kostet nur 855,- DM

Ein Microcomputer-Programmsystem für die Corpuslinguistik, Textlinguistik und alle Wortschatz- und Textuntersuchungen

CORDA

für die Herstellung von Wortformenwörterbüchern, lemmatisierten Wörterbüchern, Frequenzwörterbüchern, rückläufigen Wörterbüchern, (zentrierten) Zeilen- und Satzkonkordanzen, KWIC, KWAC, KWOC-Ausgabeformate, Wortstatistik (absolut, relativ)

CORDA kostet nur 1.140,- DM

...CO
 ...stitut
 ...r Electronische
 ...ommunication und
 ...rganisation GmbH

für den Textdatei-Transfer vom Microcomputer zum Fotosatz und zur automatischen Übersetzung der typographischen Codierungen.

TRANS/TRANSFILTER

TRANS ist ein Programm zur Übertragung von ASCII-Textdateien von Microcomputern auf Fotosatz-Anlagen.

TRANS ermöglicht auch eine Übertragung zum Microcomputer hin, wie z.B. von einem Cassettengerät oder einem anderen Computer aus. TRANS bedient die wichtigsten Fotosatz-Anlagen (z.B. Berthold, Linotype, Compugraphic), falls eine serielle Schnittstelle unterstützt wird.

TRANS kostet nur 1.368,- DM — TRANSFILTER kostet nur 399,- DM

Arbeitskreise

ARBEITSKREIS "AUSBILDUNG UND BERUFSPERSPEKTIVEN" WISSENSWERTES VOM INFORMATIONSMARKT IN GÖTTINGEN

Betr.: BAFÖG und Auslandsstudium

Zunächst* nur für das europäische Ausland (incl. europ. Türkei) gilt:

- Für BAFÖG-Bezieher ist ein Auslandsstudium möglich.
- Die Höchstdauer des Auslandsstudiums beträgt 1-2 Semester.
- Diese Zeit wird nicht auf die Förderungshöchstdauer angerechnet.
- Ein Auslandsstudium während der beiden letzten Semester der Förderungshöchstdauer ist nicht möglich.
Dasselbe gilt für das erste Studiensemester - es sei denn, es handelt sich um ein Studienfach mit bundeseinheitlichen Zulassungsbeschränkungen.
- Die Antragstellung für ein Auslandsstudium erfolgt über Ihr zuständiges BAFÖG-Amt und sollte ca. 1/2 Jahr vor dem fälligen Semesterbeginn liegen.
- Sie benötigen ein formloses Schreiben eines Ihrer Dozenten, in dem er Ihr Auslandsstudium an der gewählten Universität befürwortet - und zwar für den entsprechenden Fachbereich.
- Jedes Bundesland ist für eine Anzahl von bestimmten europäischen Ländern "zuständig". Das Auslandsamt Ihrer Universität kann Ihnen mitteilen, welches Bundesland für das von Ihnen gewählte europäische Land "zuständig" ist. Auf diese Weise können Sie Näheres über "Ihre" Auslandsuniversität erfahren.
- Während Ihres Auslandsstudiums erhalten Sie Ihr Inland-BAFÖG als Darlehen, zusätzlich als Zuschuß (also nicht rückzahlbar) eine Ausgleichszahlung für Studiengebühren, Fahrtkosten und einen Kaufkraftausgleich, der je nach dem Land bemessen ist.
- Kenntnisse in der Landessprache sind erforderlich. Sprachaufenthalte werden nicht BAFÖG-finanziert.

TIP: LDV-Studenten, Studenten von Computer Science oder AI-Interessenten würden am besten zuerst erkunden, welche Universitäten in Europa überhaupt für sie in Frage kommen. Auskünfte geben Ihnen Ihre Dozenten und der Supplement-Band 9 vom Dezember 1983 vom American Journal of Computational Linguistics.

Betr.: Auslandsstudium und (Ferien-)Praktika im Ausland

- Über EUROTRA werden Gelegenheit zum Auslandsstudium und für (Ferien-)Praktika, aber auch Hilfskraftstellen im Ausland vermittelt.
Kenntnisse in der Landessprache sind erforderlich.

Es bestehen Kontakte zu den Universitäten in

Essex und Manchester

Nancy und Paris

Pisa und Turin

Gent und Löwen

Kopenhagen

Athen

Barcelona.

* Eine Ausdehnung auf nichteuropäische Länder ist vorgesehen.

Arbeitskreise

Bitte wenden Sie sich an den Projektleiter:

Dr.J.Haller
Projekt EUROTRA D
Institut der Gesellschaft zur Förderung der Angewandten
Informationsforschung e.V. an der Universität des Saarlandes
Martin-Luther-Str.14

6600 SAARBRÜCKEN 3

Tel.0681/39313

Auskünfte über Sprachaufenthalte oder Auslandsstudium in ITALIEN bei:

ITALKONTAKT

Im Brachfelde 14

3400 GÖTTINGEN , Tel. 0551/24718

Bitte geben Sie bei Anfragen die gewünschte Fakultät an!

etr.: Studienortwechsel im Inland und BAFÖG-Bezug

Bei Studienortwechsel kommt es für BAFÖG-Bezieher darauf an, daß sie nicht von einem Studiengang in einen völlig anderen wechseln, sondern daß es sich allenfalls um eine Schwerpunktverlagerung in den Fächern handelt.

D.h. Ihre bisherigen Studienzeiten werden von den Prüfern des gewünschten Studiengangs anerkannt, und beide Studiengänge, der bisherige und der zukünftige, müssen die gleiche Pflichtsemesterzahl aufweisen.

Sonsternfalls bleibt Ihnen nur eine Unterbrechung des Studiums (mit Exmatrikel der Beurlaubung) - worauf bei Rückkehr in den alten Studiengang ein BAFÖG-Neuantrag zu stellen wäre.-

Schlechte Aussichten für den Studienortwechsel im Fach LDV (s.Studienführer LDV), da man sich auf knifflige Diskussionen bezüglich der Schwerpunktverlagerung vorbereiten muß. -

Studiengang	BI	BN	BN	KO	RG	ST	ST	TR
	LIN	INF	KOFO	INF	SP	INF	LIN	LDV
BI LIN	-				?			
BN INF		-		X		X		
BN KOFO			-					
KO INF		X		-		X		
RG SP	?				-			?
ST INF		X		X		-		
ST LIN							-	
TR LDV					?			-

"X" bedeutet , daß ein Studienortwechsel bzgl. des BAFÖG-Bezugs möglich ist.

Die Tabelle bezieht sich auf die Universitäten in Bielefeld, Bonn, Regensburg, Tübingen und Trier, die EWH in Koblenz und die Studiengänge Linguistik, Informatik, Kommunikationsforschung, Sprachwissenschaft, Linguistische Datenverarbeitung (s.dazu Studienführer LDV).

Es bedarf, wie man leicht sieht, der Vervollständigung; dazu aber brauchen wir Informationen von Studierenden, die entsprechende Wechsel vollzogen oder erfolglos versucht haben.

Magdalena Lutz-Hensel, Universität Bonn

ARBEITSKREIS "MASCHINELLE LEXIKOGRAPHIE UND LEXIKOLOGIE Bericht über das zweite, dritte und vierte Zusammentreffen

Im Anschluß an die konstituierende Sitzung unseres AK's im November letzten Jahres in Mannheim (s. hierzu den Bericht von Gert Frackenpohl im LDV-Forum 2/1985, S. 116) traf sich der AK im Berichtszeitraum (Januar bis April 1986) insgesamt dreimal:

- am 14. Januar im IKP (Institut für Kommunikationsforschung und Phonetik) in Bonn,
- am 27. Februar im Anschluß an die GLDV-Jahrestagung in Göttingen und
- am 21. April bei der GID (Gesellschaft für Information und Dokumentation) in Frankfurt.

Alle drei Zusammenkünfte waren erfreulicherweise mit jeweils ca. 20 Teilnehmern gut besucht.

1 Das Treffen im IKP in Bonn

Bei der Bonner Zusammenkunft standen zwei Dinge im Mittelpunkt des Interesses:

- die Lexikonkomponente des Expertensystems LEX sowie
- die am IKP zur Verfügung stehenden maschinellen Hilfsmittel zur Unterstützung lexikographischer Arbeiten.

Die Lexikonkomponente des Expertensystems LEX

Vormittags berichteten Ernst Latniak und Angelika Storrer von den Arbeiten an der Lexikonkomponente für LEX. LEX wird am WZH (Wissenschaftlichen Zentrum Heidelberg) der IBM Deutschland entwickelt und ist ein - auf das Gebiet des Verkehrsrechts spezialisiertes - juristisches Expertensystem. Sowohl beim Wissenserwerb ('knowledge acquisition') als auch beim Dialog mit dem Endbenutzer stützt sich LEX auf eine natürlichsprachliche Schnittstelle. Die Erstellung der Lexikonkomponente für diese natürlichsprachliche Schnittstelle wird geleitet von Brigitte Barnett, Hubert Lehmann und Magdalena Zoeppritz; die Bearbeiter (innen) sind Brigitte Endres (Substantive), Ernst Latniak (Adjektive), Georg Schuelzke (Substantive) und Angelika Storrer (Verben). Mit dem Lexikonaufbau wurde im Sommer 1984 begonnen, und inzwischen umfaßt das Lexikon ca. 8000 Einträge.

Bei der Lemmaauswahl wird von Unfallbeschreibungen sowie von einschlägigen Gesetzes- und Urteilstexten ausgegangen. Bei der lexikographischen Bearbeitung des so gewonnenen Wortschatzes werden

dann aber auch nicht verkehrsspezifische Bedeutungen berücksichtigt.

Vorgetragen wurde in Bonn über die syntaktische Beschreibung der Lexikonkomponente. Die syntaktische Beschreibung folgt valenzorientiert (wobei den syntaktischen Valenzbeschreibungen inhaltlich wesentlichen gefolgt wird) und stützt sich auf diesbezügliche Angaben in Lexikonwörterbüchern und allgemeinen Wörterbüchern sowie auf die Auswertung von Konkordanzen. Das Lexikon ist als relationale Datenbank organisiert, und in dieser Datenbank stehen 15 Attribute ('Spalten') zur Aufnahme der syntaktischen Valenz der Wörter zur Verfügung, wobei für jedes Wort ein endlicher Wertebereich festgelegt ist.

Im Hinblick auf die Gesamtkonzeption des LEX blieb unklar,

- in welcher Weise die übergeordnete Syntaxanalyse-Komponente und die Lexikon zur syntaktischen Valenz im Dialog aufeinander abgestimmt sind
- welche Zusammenhänge zwischen
 - syntaktischer Beschreibung,
 - semantischer Beschreibung und
 - Inferenzkomponente
 bestehen.

Der Nachmittag stand im Zeichen der Einführung von drei verschiedenen Hardwaresystemen, die am IKP zur Unterstützung lexikographischer Arbeiten zur Verfügung stehen: ALEXSYS, MULI und CONCORDANCE. Die Vorführenden waren Brustkern, Maria Efstathiou und Gert Frackenpohl, und die nachfolgenden Wortbeschreibungen stützen sich im wesentlichen auf eine Vorlage, die mit dem Brustkern freundlicherweise zur Verfügung gestellt hat.

- ALEXSYS ist ein Programmsystem für die Verwaltung, Bearbeitung und Benutzung einer ca. 300000 Worteinträge umfassenden Wortdatenbank des Deutschen. Dieses System ist auf dem Univac 11/03 Rechner installiert.
- MULI ist ein multilinguales Lexikon, das den Aufbau der Transflex in EUROTRA-D für die Sprachen Englisch, Französisch, Griechisch, Italienisch, Niederländisch, die am IKP erstinstanzlich unterstützt. Das System wird auf einem IBM PC AT eingesetzt. In MULI werden die folgenden Informationen gespeichert und bearbeitet:

Arbeitskreise

- Wortform in Verbindung mit einer Lesartennummer.
- Definition (zur Desambiguierung der Lesarten für den Bearbeiter).
- Beispielsätze und Phrasen (zur Verdeutlichung des Transfers).
- Angaben einer Geometrie (formaler Aufbau der Wortumgebung, soweit sie für die Transferbedingungen wichtig ist).
- Angabe der Transferbedingungen (im wesentlichen semantische Merkmale).

Diese Informationen werden mehrsprachig in der folgenden Form aufgenommen: Wortform, Lesart und Definitionen stehen in einer einsprachigen Wortliste; zu dieser einsprachigen Wortliste gibt es für jedes Sprachpaar (z.Zt. dt-gr, dk-dt, nl-dt) eine entsprechende Transferliste, in denen die Beispielsätze mit der Übersetzung und die transferabhängigen Informationen stehen.

BYU-CONCORDANCE ist ein Konkordanzpaket, das zur Erstellung von Konkordanzen des EUROTRA-Testkorpus verwendet wird, und läuft ebenfalls auf einem IBM PC AT.

Zusammen mit diesen Systemen wird am IKP eine allgemeine Programmbibliothek der LDV und ein Textarchiv geführt. Diese Systeme werden wie ALEXSYS auf der IBM-Anlage des Regionalen Hochschulrechenzentrums betrieben.

Die Treffen in Göttingen und in Frankfurt bei der GID

Nachdem eine Anregung von Seiten des LDV-Vorstands, ein Tutorial zur maschinellen Lexikographie durchzuführen, vom AK auf seiner zweiten Sitzung in Bonn positiv aufgenommen worden war, fand es bei der dritten und vierten Zusammenkunft in Göttingen und Frankfurt hauptsächlich um die Vorbereitung dieses Tutorials. Im Hinblick auf die Gestaltung des Tutorials ist der Stand der Diskussion im AK z.Zt. m.E. in etwa der folgende:

Zeitlicher Rahmen: Vier Tage.

Ort und Termin: Entweder unmittelbar vor oder nach der GLDV-Jahrestagung 1987 am gleichen Ort, oder aber entkoppelt von der GLDV-Jahrestagung Anfang 1987.

Teilnehmer: Studenten, Linguisten, LDV'er und Informatiker sowie 'praktizierende Lexikographen' aus der akademischen und aus der kommerziell betriebenen Lexikographie.

- Zielsetzung: Die Teilnehmer des Tutorials sollen in Methoden und Probleme der maschinellen Unterstützung des lexikographischen Prozesses eingeführt werden und anhand ausgewählter Beispiele etwas Erfahrung im Umgang mit entsprechender Software erwerben.

- Struktur: Das eigentliche Tutorial untergliedert sich in fünf Sektionen:

- Einführung in EDV und LDV.
- Einführung in die Lexikographie.
- Das Wörterbuch im Computer: Wörterbuchstrukturen und deren Modellierung.
- Vom (Text-)Korpus zum Wörterbuch: Wie macht man ein Wörterbuch?
- Der (elektronische) Arbeitsplatz des Lexikographen: Realitäten und Tendenzen.

Die beiden einführenden Sektionen sollen evtl. parallel stattfinden; ansonsten gehen wir davon aus, daß alle Tutorialsteilnehmer an allen Sektionen - die entsprechend aufeinander abgestimmt sein werden - teilnehmen.

Als Ergänzung zu der in erster Linie auf Wissensvermittlung ausgerichteten Arbeit in den Sektionen ist ein Rahmenprogramm geplant, das noch einige ergänzende und/oder weiterführende Vorträge und/oder Demonstrationen enthalten soll.

- Textgrundlage: Die Arbeit in den Sektionen soll sich auf ein 'Lesebuch' zur (maschinellen) Lexikographie stützen, daß den Tutorialsteilnehmern bereits einige Zeit vor Tutorialbeginn zugänglich gemacht werden soll.

Auf der kommenden AK-Sitzung am 5. Mai in Heidelberg soll die Vorbereitung des Tutorials fortgesetzt werden.

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

ARBEITSKREIS "LDV UND NACHBARN"

Der Arbeitskreis "Ldv und Nachbarn" kann sich leider in keiner Weise großer Taten rühmen. Die Arbeitsbelastung der Mitglieder ist die bedauerliche aber zutreffende Entschuldigung. Dem Erfolgskonto des Arbeitskreises kann ansonsten zugeschlagen werden, daß zur Vorbereitung der COLING '86 eine "Landkarte einschlägiger Forschungsaktivitäten" auf der Grundlage der AK-Unterlagen von einem AK-Mitglied erstellt wurde. Dazu wurde eine Datenbank aufgebaut, über deren sinnvolle Weiterverwendung noch nachgedacht wird. Sofern die Kapazitäten ausreichen, sollen auch die anderen dem AK zugänglichen Informationen zur weiteren Auswertung in eine Datenbank gebracht werden.

ARBEITSKREIS "TEXTANALYSE"

1. Während der GLDV-Frühjahrstagung in Göttingen wurde ein weiterer GLDV-Arbeitskreis gegründet, der AK Textanalyse (). Die erste Arbeitssitzung war am 27.02.86 in Göttingen. Die thematischen Interessen der Teilnehmer dienen dazu, die thematischen Interessen der Teilnehmer kennenzulernen. Ein zweites Treffen wurde für den 19.03.86 in Ulm vorgesehen. Wer Interesse an der Mitarbeit im Arbeitskreis hat, wende sich bitte an:

Sebastian Goesser
Universität Ulm, SFB 129 TP B2
Am Hochsträß 8
D-7900 Ulm

2. Um zur Diskussion über die Ziele des AK, die bislang definiert sind, beizutragen, will ich hier einige Probleme vorstellen, die nach meinem Verständnis für die Textanalyse in der LDV charakteristisch bzw. essentiell sind. Diese zusammengekommen entsprechen sie ungefähr dem, was Ch. Rieser (1983) als 'Realistic Language Comprehension' bezeichnet:

2.1 Ein TA-System sollte natürliche Texte verarbeiten (können), d.h. höchstens schwach aufbereitetes Textmaterial, wie es in alltäglichen Kommunikationssituationen vorkommt. Dazu gehören z.B. Zeitungsartikel, Gesprächsaufzeichnungen, literarische Erzählungen, Inhaltsverzeichnisse, MMK-Dienstreche, Gebrauchsanweisungen, Bürokorrespondenz und vieles andere. Von theoretischer Hinsicht gehören TA-Objekttexte eher zur Linguistik als zur Logik, und sie enthalten zahlreiche Phänomene (z.B. Textsorten-Markierung), die die linguistische Textanalyse (noch) nicht erklären kann.

Natürliche Textmengen sollten nicht künstlich kleiner gemacht werden. Deshalb sind ernsthafte Zweifel an der TA-Eigenschaft eines Systems angebracht, das nur z.B. Zeitungsartikel über einen bestimmten Flugzeugabsturz "verstehen" kann. Insbesondere offene Textmengen (mit fortlaufend hinzukommenden Texten) sind -u.a. aus entwicklungsökonomischen Gründen- gute Gegenstände für die TA-Anwendung.

2.2 Ein wesentliches Merkmal von TA-Systemen ist Robustheit, d.h. die Fähigkeit mit -bezüglich des Systemwissens- nicht wohlgeformter, aber "bedeutungsvoller" Eingabe umzugehen. In der Literatur wird seit einiger Zeit das Teilproblem der Robustheit gegenüber außergrammatischen Eingabe intensiv behandelt. Lösungsansätze sind hier beispielsweise die constraint-Ermäßigung während des Parsings, beidseitige Syntax-Semantik-Interaktion sowie wie das heuristische und/oder partielle Parsen. Andererseits ist insbesondere bei der Kombination solcher Strategien die Robustheit logikorientierter Parser, bei der Integration von semantischer Erwartungen und ähnlichem für die wissenschaftliche LDV noch viel zu tun.

2.3 Ein wichtiger Unterschied zu (den meisten Arten von) computerunterstützter Inhaltsanalyse liegt darin, daß TA-Systeme grammatikbasiert sind, wobei als Grammatik eine hierarchische oder netzwerkorientierte Abbildung von Eingabestrings auf eine gewählte semantische Repräsentation bezeichnet wird. Eine systematische Evaluierung der Textanalyseeigenschaften verschiedener Grammatikmodelle liegt m.W. nicht vor, obwohl die Literatur oftmals Einzelaussagen hierzu enthält, z.B. über die notorische Störanfälligkeit von ATN-Parsern. In diesem Zusammenhang ist es wichtig, daß auch heuristische "Grammatiken" im definierten Sinn Grammatiken sind, auch wenn für sie häufig keine implementationsunabhängige Darstellung angegeben wird.

2.4 Zumindest für TA-Systeme, deren verfahrensinternes Wissen unabhängig von der Anwendungsumgebung ist, gibt es einen Tradeoff von Robustheit und Analysetiefe: Robustes Parsing wird zu einer trivialen Aufgabe, wenn es nur leichtlösliche Analyseprobleme berücksichtigt; andererseits ist es ein extrem hartes Problem, einen Parser, der z.B. kausale Kursstrukturen herausarbeitet, mit der Robustheitseigenschaft auszustatten. Denn eine Lösung setzt voraus, daß die Spezifität des Wissens, das in die Textanalyse eingeht, sorgfältig dosiert wird.

NEUER ARBEITSKREIS "SPRACHERKENNUNG, SPRACHGENERIERUNG UND PHONETISCHE DATENBANK"

Offensichtlich unter dem positiven Eindruck der Mitgliederversammlung der GLDV am 26.2.1986 in Göttingen beschlossen, einen weiteren Schritt in Richtung ihrer ursprünglichen satzungsgemäßen "Clearingstellenaktivitäten" zu tun: Die nächste Jahrestagung in Bonn soll unter dem Schwerpunktthema - Spracherkennung - Wissenschaftler der Fächer zusammenführen, die an der Aufgabe arbeiten, fließende Sprache automatisch so zu analysieren, daß eine Weiterverarbeitung z. B. auf Graphembasis ermöglicht wird.

Da dieses Thema mit Sicherheit nicht allein von Mitgliedern der Gesellschaft getragen werden kann, wurde der neue Arbeitskreis beschlossen. Im ersten Jahr hat er die Aufgabe, die Jahrestagung 1987 vorzubereiten zu helfen und führende Fachvertreter als Referenten zu gewinnen.

Inzwischen hat die konstituierende Sitzung am 4.1986 im Institut für Phonetik und Sprachliche Kommunikation der Universität München stattgefunden.

Nachdem die Aktualität des Rahmenthemas 1987 insbesondere auch vor dem Hintergrund der an diesem Tage zu Ende gehenden Tagung der Nachrichtentechnischen Gesellschaft (NTG) bestätigt wurde, wurden auftragsgemäß folgende Vorschläge dem Vorstand gerichtet:

1. Zur Jahrestagung der GLDV in Bonn im Februar 1987 lädt der Vorstand Fachvertreter ein aus NTG, GI, DGS, GLDV u. a. zu einem Podium mit dem Thema Automatische Segmentation von Sprachsignalen fließender Rede - Stand der Forschung

2. Für Übersichtsvorträge zum Rahmenthema Spracherkennung sollen u. a. folgende Fachvertreter gewonnen werden: Ney (Philips Hamburg), Höge (Siemens München), Tillmann (Uni München), Heß, Ruske (TU München), Altmann (Uni München), Lenders (Uni Bonn), Frau Brietzmann, Görtz (Uni Erlangen) Erfreulicherweise liegen bereits Zusagen von Prof. Altmann, Prof. Tillmann und Prof. Lenders vor.

3. Der "Call for Papers" wurde ausgearbeitet. Siehe dazu dessen Veröffentlichung in dieser Ausgabe des LDV-Forums.

4. Nächstes Treffen des Arbeitskreises findet nach telefonischer Abstimmung während des COLING-Kongresses in Bonn, vermutlich am Freitag, 29.8.1986, statt.

Kontaktadressen:

Klaus Günther Schweisthal, Institut für Phonetik und Sprachliche Kommunikation der Universität, Schellingstraße 3, II. Stock, VG, 8000 München 40
Gerd Willee, Institut für Kommunikationsforschung und Phonetik, Poppelsdorfer Allee 47, 5300 Bonn 1.

ARBEITSKREIS "MASCHINELLE ÜBERSETZUNG"

Die Aktivitäten des Arbeitskreises bestanden hauptsächlich in der Organisation von und der Beteiligung an Veranstaltungen, über die unter der Rubrik "Tagungsberichte" informiert wird.

Tagungsberichte

ERSTER KONGRESS DER INTERNATIONALEN GESELLSCHAFT FÜR PSYCHOLINGUISTIK, BARCELONA

Vom 16. bis 20. Juni 1985 war das Department de Psicologia der Universität Barcelona (Organisationskomitee: Ehrenpräsident: Antoni Badia i Marañón, Präsident: Miguel Siguan, Generalsekretär: Ignasi Vila) Gastgeber des ersten Kongresses der Internationalen Gesellschaft für angewandte Psycholinguistik (International Society of Applied Psycholinguistics -ISAPL-). In den Plenarvorträgen an den Nachmittagen wurden folgende Grundsatzerferate gehalten: Slama-Cazacu: 'Psycholinguistics and the Analysis of Literary Text'; E. Oksaar: 'Bilingualism and International Competence'; I. Kurcz: 'Present Day Research in Psycholinguistics Applied to Cognitive Psychology'; F. V. Manien: 'Mass Media and Psycholinguistics'; K. Butler: 'Speech disturbances: a Psycholinguistic Approach'; R. Titone: 'Metalinguistic Awareness in Children: A Crucial Prerequisite to Reading'; D. Seleskovich: 'La Traducción d'un Point de Vue Psycholinguistique'; und M. Siguan: 'La Internacionalización del Idioma'. - In den Nachmittagsstunden fanden parallele Arbeitssitzungen zu den folgenden Themen bereichen statt: 'Language Acquisition', 'Language Production and Perception', 'Second Language Acquisition', 'Bilingualism', 'Reading and Writing', 'Communication', 'Speech Disorders' und 'Language and Education'. - Eine Sektion war der Diskussion der Besonderen Bedeutung Vigoskys für die angewandte Psycholinguistik gewidmet. - Vier von den Veranstaltern einberufene Symposia beschäftigten sich darüberhinaus mit speziellen Problemen der Angewandten Psycholinguistik: 'Psycholinguistics and Reading Research in Scandinavia: Some Pedagogical Consequences' (Organisator: E. Engberg); 'Introspective Methods in Second Language Acquisition Research' (Organisatoren: C. Faerch und G. Kasper); 'Bilingualism and Aphasia' (Organisator: M. Paradis); 'Politeness Formulas and Social Interaction' (Organisator: T. Slama-Cazacu). Eine weitere Arbeitsgruppe behandelte das Thema: 'The Revelations and Limits of Language Classroom Data'. Auf einer Sitzung des ISAPL Komitees wurde als Veranstalter für den zweiten Internationalen Kongreß die Universität Kassel, Bundesrepublik Deutschland, vorgeschlagen. Dieser Kongreß soll im August/September 1987 stattfinden.

Arbeitskreis für Semiotik, Berlin

SYMPOSIUM: LANGUAGE AS A COGNITIVE SYSTEM

Am 27. Mai 1985 fanden in Santa Barbara zwei Symposien zu Ehren von Edward Sapir statt. Sie waren als Beitrag der Linguistic Society of America zu dem Treffen der American Association for the Advancement of Science in Los Angeles organisiert. Für das erste waren folgende Beiträge angekündigt: Micheal Cole und Peg Griffin: 'Cognition as a Social Process'; Elizabeth Bates: 'Cognitive Sources of Language Development'; Victoria A. Fromkin: 'Language as an Independent Cognitive System';

Paul Kay: 'Universal Cognitive Categories in Human Lexical Systems'; und Charles Ferguson: 'Language Learning and Language Change as Cognitive Processes'. Auf dem zweiten Symposium am Nachmittag desselben Tages gab es folgende Beiträge: Michael Shapiro: 'Sapir's Concept of Drift in Semiotic Perspective'; Michael Silverstein: 'Sapir's Psychological and Psychiatric Perspectives on Culture'; sowie Paul Rubel und Abraham Rosman: 'West Coast Tribes and Their Social Structural Transformation'.

Arbeitskreis für Semiotik, Berlin

INTERNATIONAL SUMMER INSTITUTE FOR SEMIOtic AND STRUCTURAL STUDIES

Vom 27. Mai bis 22. Juni fand, organisiert von Michael Hersfeld, an der Indian University, Bloomington, das sechste International Summer Institute for Semiotic and Structural Studies statt. Es wurden folgende Seminare abgehalten:

1. Language and Cognition in Social Theory (Aaron Cicourel);
2. Theories and Models in Semiotics and Literatures (Robert de Beaugrande);
3. Pierce and Semiotics (Gerard Deledalle);
4. Semiotics and Fictionality (Ludomir Dolezel);
5. Polysystem Theory and Heterogeneity in Language, Literature and Society (Itamar Even-Zohar);
6. Semiotics of the Theater (Erika Fischer-Lichte);
7. The Social Semiotic of Grammar (Michael A.K. Halliday);
8. Semiotics and Human Identity (Michael Herzfeld);
9. Homology of Language and Culture - a Case Study of Japanese Semiotics (Yoshihiko Ikegami);
10. Gesture, Forms and Funktionen (Adam Kendon);
11. Structuralism and Semiotics of History (Valentin Mudimbe);
12. The Coy Science: Semiotics and Art History (Donald Pezosi);
13. The Anthropology of Oral Literature (Erik Schwimmer);
14. Indexicality (Thomas A. Sebeok);
15. Marketing (Jean Umiker-Sebeok);
16. Semiotics and Liturgy (Drid Williams);
17. Semiotics and the Golden Bough: Rereading Frazer (Masao Yamaguchi).

Das Programm wurde ergänzt durch eine Vielzahl von Abendvorträgen, hauptsächlich aus dem Kreis der 'Visiting Scholars', einen mehrtägigen Workshop zum Thema 'Marketing Semiotics' sowie drei Wochenendkolloquien (1. Semiotics and Social Theory, 2. Semiotics and the Curriculum, 3. Semiotics in/of Africa) und das an mehreren Abenden tagende Symposium über 'The Semiotics of the Humanities: a Collective Investigation into the Ideology of the Discourse'. Die durchgängig bestimmende Interdisziplinarität und Internationalität resultierte in fruchtbaren Diskussionen, welche die Bedeutung der Semiotik für die verschiedensten Richtungen der Geistes- und Sozialwissenschaften unter Beweis stellten.

Arbeitskreis für Semiotik, Berlin

KOLLOQUIUM 'PROCESSING REFERENCE' IN KLEVE

Vom 28. bis 30. August 1985 fand das vom Nijmegen Institute of Semantics veranstaltete 3rd International Colloquium 'Interdisciplinary Study of the Semantics of Language' zum Thema 'Processing Reference' statt. R. Rhodes (Michigan) sprach über 'The Mark of Non-Coreference'; L. J. Vandenberg (Antwerpen) setzte sich mit dem Beitrag 'Pronoun Control and Agreement' mit P. Boschs Theorie der Anaphora und der Anapher (P. Bosch: Agreement and Anaphora, London and New York 1983) auseinander, und M. Lucas (Glasgow) befaßte sich mit 'Lexical Constraints on Pronominal Reference: A Problematic Anaphoric Relationship' und Verfahren zu ihrer Auflösung waren Gegenstand der Beiträge von M.M. Woods (London): 'Anaphora and Gapping in Coordinate Structures'; Ch. Hauenschild (Berlin): 'The Russian Noun Phrases' und Th. Aston: 'Processing reference: The Case of 'Invited Texts''. S. Löbner (Düsseldorf) sprach über 'Processing Definite Description', K. Rosen (Berlin) über 'Reference to Events and the Complexity of Sentences', P. Seuren (Nijmegen): 'A Logical Problem With Reference', darüber die sogenannten 'Eselssätze' und Möglichkeiten seiner Lösung mittels analytischer Verfahren. Logik behandelte R. Bäuerle (Konstanz) in seinem Beitrag 'Anaphora, Noun Phrases, and the Logical Problem With Reference', P. Staudacher (Regensburg) stellte die Beziehung dieser Problematik 'A Logic for the Relationship Between Deixis and Anaphora' vor. 'A Conversational Approach to the Relationship Between Deixis and Anaphora' war das Thema des Vortrags von F. Spas (Sussex). G. Houghton (Sussex) sprach über 'The Quantification of NP-Types in Computer-Generated Texts' und S. Garrod (Glasgow) über 'Reference in the Shared Physical Environment: Problems of Composition, Cancellation, and Adaptation' behandelte R. van der Sandt (Nijmegen), H. van Jaarsveld und R. Schreuder (Nijmegen) die Ergebnisse ihrer Informantentests 'The Quantification of Temporal on the Use of Adverbs' vor. 'Automatic Anaphora Resolution: The Use of Focus in the Computation of Referring Expressions' und D. Carter (London): 'Common Sense: Inference in a Focus-based Anaphora Resolver'. Philosophische und sprachtheoretische Probleme der Referenz thematisierten Th. Gyllfason (Rykyjavik): 'Reference and Language', P. Lutzeier (Berlin): 'Reference to Reality' und D. Kaplan (Los Angeles): 'World Beliefs'. - Die Veröffentlichung der Beiträge vom NIS herausgegebenen Journal of Semantics ist geplant.

Arbeitskreis für Semiotik Berlin

KONGRESS
SEMIOTIK, KOMMUNIKATION UND KULTUR*
IN SAO PAULO

Vom 2. bis 6. September 1985 wurde von der Brasilianischen Semiotik-Gesellschaft in Zusammenhang mit der Universidade de Sao Paulo und der Pontificia Universidade Catolica de Sao Paulo der zweite Brasilianische Semiotikkongress veranstaltet. Wie schon beim ersten Kongress 1978 (vgl. Zeitschrift für Semiotik 1 (1979), S.144) war jeder Tag einer besonderen Problematik gewidmet, die am Vortag von einem international bekannten Semiotiker themenformig abgehandelt, anschließend in einer Podiumsdiskussion behandelt wurde, so daß sich am Ende ein umfassendes Bild des Forschungsstandes zum Rahmenthema des Kongresses 'Semiotik, Kommunikation und Kultur' ergab. Der erste Tag galt dem Verhältnis von Semiotik und Einzelwissenschaften. Im Vordergrund des Interesses stand das theoretische Verhältnis zwischen speziellen Zeichenanalysen innerhalb oder außerhalb der traditionellen Disziplinen auf der einen Seite und der allgemeinen semiotischen Rekonstruktion der Fragestellungen und Methoden dieser Disziplinen auf der anderen Seite. Der zweite Tag war der Semiotik der Kunst und Architektur gewidmet und wurde durch Vorträge von Alain Renier (Paris) über 'Die Problematik des Entwerfens' und von Rosa Ravera (Buenos Aires) über 'Die Problematik der Interpretation' eingeleitet. Auf der von Eduardo Penuela Canizal (Sao Paulo) die Übersetzung von einem Medium ins andere - Operationen, die unter dem Gesichtspunkt ihrer Angemessenheitskriterien erörtert wurden. Der dritte Tag war für die Semiotik der Literatur reserviert und wurde durch Vorträge von J. Dines Johansen (Odense) über 'Peirce und die literarische Interpretation' und Wladimir Kryszinski (Montreal) über 'Semiotik und Hermeneutik' eingeleitet. Anschließend diskutierten Diana Luz Pessoa de Barros (Sao Paulo), Heraldo de Campos (Sao Paulo) und Joao Alexandre Barbosa (Sao Paulo) unter der Leitung von Luis Felipe Ribeiro (Rio de Janeiro) Probleme der literarischen Übersetzung. Am vierten Tag wurde die Semiotik der Kommunikation behandelt. Ausgehend von einem Vortrag von Joan Costa (Barcelona) über 'Die Rolle der Photographie in der Massenkommunikation' unterzeichnet Aloisio Trinta (Rio de Janeiro), Arlindo Machado (Sao Paulo) und Aurora Bernardini (Sao Paulo) unter der Leitung von Dacio Pignatari (Sao Paulo) die Möglichkeiten, das Kommunikationsverhalten des Menschen in Familie, Staat und Kirche durch Anwendung semiotischer Modelle zu klären. Alle Diskussionspartner waren sich einig, daß ein solcher Klärungsprozeß tiefgreifende Veränderungen gerade auch der Brasilianer bewirken werde. Der letzte Tag des Kongresses, der die Semiotik der Kultur zum Thema hatte, wurde von Roland Posner (Berlin) mit dem Vortrag über 'Fünf Alternativen der semiotischen Rekonstruktion von Kultur' eingeleitet. David Danow (New York City), Boris Schneiderman (Sao Paulo), Cid Seixas (Salvador) und Luis Fiorin (Araraquara) untersuchten anschließend zwei dieser Alternativen genauer, nämlich Kultur als Menge von Texten im Sinne von Lotman und Kultur als Menge von Zeichenprozessen im Sinne von Bachtin. Wie die Präsidentin des Kongresses, Ana Claudia Mei A. de Oliveira (Sao Paulo) abschließend feststellte, wurde auf diese Weise ausgehend von akademisch anmutenden Fragen zur Analyse der Wissenschaften im Laufe des Kongresses die Thematik immer enger auf die Probleme der Konstitution von Ich und Anderem mit Hilfe von Zeichenprozessen konzentriert, ein Verlauf, den die Organisatoren

zwar angestrebt, aber in dieser Deutlichkeit nicht einzuplanen gewagt hatten. Für nichtbrasilianische Teilnehmer des Kongresses war neben den Plenarveranstaltungen besonders auch die große Zahl und die Vielfalt der semiotischen Untersuchungen bemerkenswert, die in den Sektionsreferaten dargestellt wurden - ein Eindruck, der sich auch in der von Conrado Paschoale (Sao Paulo) geleiteten Informationsveranstaltung über den Stand der Semiotik in den brasilianischen Bundesstaaten mit den Teilnehmern Eduardo Penuela Canizal (Araraquara), Antonia Schwinden (Parana), Hildo Honorio do Couto (Brasilia), M. Lucia A. Garcia (Rio Grande de Norte), Elizabeth Duarte (Rio Grande do Sul) und Milton J. Pinto (Rio de Janeiro) bestätigte. Zur Abrundung dieses Eindrucks trug außerdem das vielseitige Begleitprogramm des Kongresses bei: eine umfangreiche Buchausstellung, Ausstellungen von zeichenbezogener moderner brasilianischer Kunst, von neuen Technologien der Zeichenverarbeitung, von Volksliteratur (Literatur de Cordel), sowie multimediale Vorführungen von nonverbaler Kommunikation. Höhepunkt des Begleitprogramms war eine Veranstaltung zum Gedenken an Roman Jakobson, auf der Boris Schneiderman und Augusto de Campos moderne russische Lyrik zweisprachig (russisch und portugiesisch) vortrugen. Daß es zu den Fragestellungen und Methoden der brasilianischen Semiotik heute heute sehr viele Parallelen außerhalb Brasiliens gibt, zeigte sich in der Informationsveranstaltung über 'Neue Entwicklungen in der Internationalen Semiotik', auf der am Abend des letzten Tages unter der Leitung von Jose A. Seabra (Portugal), Alain Renier (Frankreich), J. Dines Johansen (Dänemark), Joao Costa (Spanien), Andrea Bonomi (Italien), Marlene Landsch (BRD), David Danow -1 (USA) und Wladimir Kryszinski (Kanada) Kurzdarstellungen der wichtigsten Richtungen semiotischer Forschung in ihren Ländern gaben.

Arbeitskreis...

"20. LINGUISTISCHES KOLLOQUIUM"
IN WOLFENBÜTTEL

Von 17. bis 20. September 1985 fand in Wolfenbüttel das von K.-H. Kömer (Braun-Schweig) organisierte 20. Linguistische Kolloquium statt. Gemäß der Tradition der Veranstaltung bestanden keine thematischen Beschränkungen, so daß durch die Vorträge ein breites Spektrum unterschiedlicher linguistischer Themen abgedeckt wurde. Zu lexikologischen und morphologischen Problemen referierten H.U. Boas (Hannoversch-Münden): 'Lexikale und andere systematische Lücken', J. Korhonen (Tübingen): 'Zum Wesen der Polyvalenz im Deutschen', G. Rohdenburg (Münster): 'Eingeschränkt auslaßbare Ergänzungen bei englischen Adjektiven. Vernachlässigte Erscheinungen in der Valenzbeschreibung', E. Leitzke (München): 'Transpositionelle Adjektive. Argumente für eine erneute Auseinandersetzung mit dem Marchandschen Begriff' und K. Sroka (Gdansk): 'Determinants in conjugal forms in Hungarian'. Mit syntaktischen Fragestellungen beschäftigten sich M. Kefer (Liege): 'Ergativität und syntaktische Relationen', E. Löbel (Köln): 'Apposition in der Quantifizierung', P. Eisenberg (Hannover): 'Zum Kontrollproblem im Deutschen', W. Abraham (Groningen): 'Passivdiathese. Typen, Derivation, typologische und diachronische Parametrisierung', B. Geurts (Nijmegen): 'Refering to stereotypes' und U. Schmidt (Bochum): 'Ist die Anordnung der Satzglieder im deutschen Satz tatsächlich keinem konsistenten Schema verpflich-

tet?'. Zur Semantik trugen vor P. Bosch (Nijmegen): 'Pronouns, accent, and lexical meaning', M. Hanowell (Münster): 'Zum Verhältnis von epistemischen und nicht-epistemischen Modalitäten im Englischen und im Deutschen', N. Montulet (Nijmegen): 'Focus, context, and unique reference', A. Nova (Nijmegen): 'Expressions of identity', L. Okon (Minusio): 'Le "Supplement au Voyage de Bougainville" de Denis Diderot, Bronislaw Malinowski et la semantique d'Odgen et Richards', und O. Mori (Münster): 'Spanische Verbalperiphrasen und mögliche deutsche Entsprechungen'. Referate zu Problemen der Pragmatik wurden gehalten von A. Burckhardt (Darmstadt): 'Antönungsarten und "Konversation" in Hoffmannsthal's "Der Schwierige"', U. Hinrichs (Berlin): 'Verschleierung, Vagheit, Gerede - Eine Tendenz in der deutschen öffentlichen Sprache', W. Grabsch (Köln): 'Überlegungen zur Indirektheit von explizit performativ vollzogenen Sprechakten', J. Graffe (Münster): '"How to promise" revisited: Eine Taxonomie kommissiver Sprechakte', J. Maibauer (Köln): 'Rhetorische Aufforderungen und Rhetorizität', H. Weber (Tübingen): 'Zur Klassifikation sprechaktsbezeichnender Verben im Deutschen', M. Kohl (Münster): 'Gesprächsstrukturierung und Rekonstruktion von Sprecherstrategien auf der Basis einer Dialoggrammatik', K. Batke (Köln): 'Sprachliches Verhalten von Student/innen an Themenübergangsstellen in Diskussionen', und S. de Knop (Bruxelles): 'Referenzidentifizierung oder Präzisierung der Gelegenheitsbildungen in Überschriften', Mit sprachdidaktischen Fragen setzten sich auseinander: K. Dorfmueller-Karpus (Bielefeld): 'Texte bilingualer griechischer Kinder', A. Braun (Marburg): 'Ein Modell zur Analyse von Rechtschreibfehlern bei Kindern ausländischer Arbeitnehmer' und G. Iobin (Paderborn): Sprachmodelle als Lehrstoffmodelle im Fremdsprachenunterricht', während Vorträge zu Fach- und Sondersprachen von M. Krenn (München): 'Satz- und textgrammatische Merkmale englischer medizinischer Fachsprache', J. van Leeuwen (Berlin): 'Argot als kulturgeschichtliches Phänomen' und K. Schubert (Utrecht): 'Wo die Syntax im Wörterbuch steht. Esperanto als Brückensprache der Maschinenübersetzung' gehalten wurde. Während sich weitere Vorträge mit sprachhistorischen Aspekten beschäftigten, trugen über Theorie und Modelle in der Sprachwissenschaft E. Feldbusch (Paderborn): 'Geschriebene Sprache. Abriß zu einer Neubestimmung', N. Reiter (Berlin): 'Über die Vermenschlichung der Sprache' und P.R. Lutzeier (Berlin): 'Gedanken zur individuellen/personlichen Mehrsprachigkeit' vor. Schließlich hielt im Rahmenprogramm des Kolloquiums W. Kürschner (Osnabrück/Vechta) in der Augusteerhalle der Herzog August Bibliothek Wolfenbüttel einen öffentlichen Vortrag mit dem Titel 'Zur Geschichte der Sprachkultur in Deutschland'. Wer genauere Information über die Referate wünscht, sei auf die von K.H. Kömer und A. Burckhardt herausgegebenen Kongressakten verwiesen, die im Mai 1986 im Max Niemeyer Verlag Tübingen erscheinen.

Mathias Kohl, Universität Münster (überlassen vom semiotischen Arbeitskreis Berlin; geringfügig gekürzt).

**SYMPOSIUM ÜBER
EMPIRISCHE GRUNDLAGEN DER
INFORMATIONEN- UND SOFTWARE-
WISSENSCHAFTEN, DÄNEMARK**

Vom 21. bis 24. Oktober 1985 fand in Risø bei Roskilde, Dänemark, das 3. Symposium über die empirischen Grundlagen der Informations- und Softwarewissenschaften statt. Es wurde unterstützt vom Risø National Laboratory und von der Königlichen Schule für Bibliothekswesen Dänemarks. Ziel der Veranstaltung war es, Themen und Methoden wissenschaftlicher Forschung zusammenzutragen, die sowohl für die Informations- als

auch für die Softwarewissenschaften von Interesse sind. Dabei sollte auch ein Forschungsweg vorgezeichnet werden, der beiden Bereichen von Nutzen sein kann. Die Veranstaltung war die erste ihrer Art, die in Europa stattfand. Die ersten zwei Symposien zu demselben Thema waren im Georgia Institute of Technology, Atlanta, USA, organisiert worden. Die nächsten Symposien sollen jährlich und abwechselnd in den USA und in Europa stattfinden. Sinn dieser Regelung ist, die Förderung des Austausches zwischen den europäischen und den amerikanischen Wissenschaftlern, die in den Gebieten von Informations- und Softwarewissenschaften arbeiten.

(Semiotischer Arbeitskreis Berlin)

**BERICHT ZUM WORKSHOP
"WORD ORDER AND PARSING IN UNIFIKATION GRAMMARS"
VOM 7. BIS 11.4.86 IN FRIEDENWEILER (SCHWARZWALD)**

Der Workshop, der nur im kleinen Kreis als Treffen der EUROTRA-Begleitforschung gedacht war, fand in großem Rahmen statt (ca. 60 Teilnehmer, 21 Vorträge). Hauptgewicht lag auf linguistischen Fragestellungen. CHRISTIAN ROHRER begann in seiner Einleitung die Wichtigkeit der Kommunikation zwischen Linguisten und Computerlinguisten. Aber auch die rein informatische Seite kam nicht zu kurz, z.B. GERG GAZDARS Vortrag über die Mächtigkeit von Indexed Grammars und ihre Anwendbarkeit auf natürliche Sprache. Anhand von Belgisch, Norwegisch und Englisch zeigt er, daß kontextsensitive Beispielsprachen, wie etwa die "String Copying Language", nur theoretischen Betrachtungen dienen.

Damit ist der Bogen vom ersten zum letzten Vortrag des Workshops gespannt. komme nun zu den dazwischenliegenden Vorträgen.

MARTIN KAY begann mit der provokanten These, daß alle zur Sprache kommenden Methoden von GB über GPSG bis LFG nur Formalismen seien, aber keine linguistischen Theorien, was zu spontanem Beifall führte.

HENK ZEWAT stellte den Formalismus der Unification Categorical Grammar vor. eine gut lesbare semantische Repräsentation komplexer NPs erzeugt, wobei allerdings die Referentenidentifikation nicht behandelt wird.

DIETER METZING sprach über Strategien zur Auflösung von Ambiguitäten auf verschiedenen Ebenen (phrase-, sentence- und discourse-level) und schlug als eine Lösung die Definition von Defaults, als eine andere die Ausnutzung von mehr linguistischer Information zur Steuerung der Analyse vor. Problematisch ist hier, daß aus Effizienzgründen die Unifikation unterlaufen wird.

Als nächstes stellte JÜRGEN WEDEKIND eine formale Repräsentation des Unifikationsprozesses vor. Dieser Vortrag wurde stark diskutiert, da in dem angegebenen Formalismus keine explizite C-Struktur mehr zu erkennen ist, die die Linguisten für unentbehrlich halten. Meines Erachtens wird durch die Formulierung des Ableitungsbegriffes, wie er analog aus den formalen Sprachen bekannt ist, der Vorgang der Unifikation klarer und die Argumentation über einen solchen Ableitungsschritt als Darstellung des Unifikationsschrittes vor allem beweistechnisch einfacher.

Am Ende des ersten Tages stellte *BRAN BOGURAEV* sein Konzept eines "Natural Language Toolkit"s für den GPSG'85 Formalismus vor. Fraglich ist, ob die Gruppe die hohen Ansprüche bzgl. zur Verfügung stehender Features bzw. Benutzeroberfläche, in der kurzen Zeitspanne, die sie sich gesetzt hat, realisieren kann.

Am Mittwoch ging *HANS USZKOREIT* auf die Probleme von Wortstellung, ihrer Variation, Permutationsmöglichkeiten und Diskontinuitäten. Welche Regeln in welchen Formalismen werden diesen Phänomenen am besten gerecht?

KLAUS NETTER stellte einen LFG-Formalismus zur Analyse von Infinitivkonstruktionen im Deutschen vor. Die Frage ist hier, ob durch seine elegante rekursive Definition nicht eine linguistisch zu flache Struktur aufgebaut wird.

CHRISTOPH SCHWARZE machte an einer Reihe von Beispielen die Schwierigkeiten der Behandlung des 'adjectif detache' im Französischen klar.

Nachdem bisher das Gewicht mehr auf LFG-Vorträgen wechselte am Nachmittag der Schwerpunkt zu GPSG:

CHRISTA HAUENSCHILD schlug in Anlehnung an *HANS USZKOREIT*s Dissertation einige Erweiterungen der LP-Regeln dahingehend vor, daß ein feinmaschiges Netz aus Disjunktionen über Reihenfolgeregeln im Satz entsteht.

JIM KILBURY stellte die "Category Cooccurrence Restriction (CCR)" vor als äquivalentes Beschreibungsmittel zu ID- bzw. Metaregeln zur Darstellung von Constraints für Cooccurrence von Kategorien in lokalen Bäumen. Diese Erweiterung dient zur Generalisierung bzw. macht Metaregeln überflüssig.

Im letzten Vortrag des Tages stellte *WILLIAM KELLER* den "Nested Cooper Storage" vor, eine Methode zur semantischen Repräsentation von NP's. In seinem Vortrag machte er die Methode im Kontrast zu *MONTAGUE*s Vorgehen klar.

Am Donnerstag sprach *STUART SHIEBER* über die Notwendigkeit der Trennung zwischen linguistischer Theorie und linguistischer Analyse aufgrund unterschiedlicher, zum Teil widersprüchlicher Zielsetzungen. Detailliert ging er auf die Probleme ein, die bei der Transformation zwischen Grammatikformalismen auftreten können.

Als nächstes stellte *PETE WHITELOCK* seine feature-based morpho-syntaktische Komponente vor, die in einem shift-reduce-Parser für die Übersetzung Englisch-Japanisch benutzt wird.

STEFAN BUSEMANN gab einen Überblick über die Komponenten seines noch zu implementierenden Systems zur Generierung natürlicher Sprache. Zur sauberen Definition seines System benutzt er als Basisoperation nur die Unifikation und arbeitet topdown, breadth-first, left-to-right auf DAG's.

SASCHA FELIX plädierte in seinem Vortrag dafür, daß eine Analyse mit reinem Regelwissen (Ergebnis sind Bäume) noch nicht ausreichend sei, sondern um einen zweiten Schritt erweitert werden müsse, der diese Bäume auf die Erfüllung gewisser Konventionen hin untersucht, wobei die Aufgabenstellung auch darauf zielt, einen unvollständigen bzw. grammatikalisch falschen Satz so zu vervollständigen, wie es ein Mensch am ehesten tun würde. Die Frage ist, ob diese Zweistufigkeit den "human sentence parser" wirklich adäquat modelliert.

HANS-ULRICH BLOCK und HANS HAUGENEDER stellten ihre Arbeiten zur Effizienzsteigerung von LFG-Parsern vor, die im Rahmen der WISBER-Kooperation Siemens angesiedelt sind.

JÖRG KINDERMANN und JUSTUS MEYER sprachen über ihre Überlegungen zur Erweiterung von Parsern, die aus dem Bereich der formalen Sprachen bekannt sind, zu Unifikationsparsern. Sie stellten einen shift-reduce-Parser vor, der von Hand transkribierte LFG-Regeln (einer eingeschränkten Klasse) verarbeiten kann.

Der letzte Vormittag begann mit zwei linguistischen Vorträgen: DIETER WUNDERLICH sprach über Koordinationsprobleme im Deutschen, die er eindrucksvoll am Beispiel "Den Korb die Tante in das Haus stellte und den Sack in den Hof der G... warf".

GISBERT FANSELOW behandelte Wortstellungsfragen in verschiedenen Sprachen. Mir bleibt abschliessend noch einmal Herrn ROHRER und seinen Mitarbeitern die Organisation und den reibungslosen Ablauf des Workshops zu gratulieren und dazu hinzuweisen, daß laut vorläufigen Informationen ein Tagungsband im REIDEL-Verlag, Amsterdam, in der Reihe "Studies in Linguistic and Philosophy Theories" voraussichtlich im Frühjahr '87 erscheinen wird.

Karin Harbusch, FB 10 - Informatik 4 - Lehrstuhl Prof. Dr. W. Wahlster, Universität des Saarlandes, Stadtwald 15, D-6600 Saarbrücken 11

Treffen des Arbeitskreises "KI und Informationswissenschaft", TH Darmstadt

Am Freitag, dem 18.4.86, traf sich von 9:15 bis 17:00 Uhr der Arbeitskreis "KI und Informationswissenschaft" im Fachgebiet Datenverwaltungssysteme II des Fachbereichs Informatik der TH Darmstadt. Dieser Arbeitskreis hatte sich 1981 auf der ersten KI-Frühjahrsschule (KIFS) als loser Zusammenschluß von Wissenschaftlern konstituiert, die an interdisziplinärer Arbeit interessiert und dafür auch engagiert sind. Als erstes konkretes Ergebnis der Arbeit des AK's darf das z.Z. an der TH Darmstadt laufende Projekt AUTOCAT zur Entwicklung eines Systems für eine wissensbasierte Formalkatalogisierung von Physikdokumenten nach INIS-Regeln angesehen werden, dessen Planung, Formulierung und Beantragung der AK aktiv begleitet hat. Insofern ist es nicht überraschend, daß der Vormittag der Veranstaltung von Mitarbeitern des AUTOCAT-Projektes gestaltet wurde. Zunächst berichteten Ulrike Rauth und Christina Schnellbach vor sehr interessierten Teilnehmern über Erfahrungen mit einem optischen Lesesystem (KURZWEIL) sowie über die Situation im Verlagsbereich hinsichtlich der Nutzung von Zeitschriftendaten für den elektronischen Satz. Damit war die Schnittstelle des Projektes auf der Inputseite beschrieben. Amelie von Below stellte anschließend die datenbankgestützte Auswertung des Aufbaus der ca. 400 sogenannten "Kernzeitschriften" (core journals) im

Fachgebiet Physik vor, die die empirische Grundlage für das im Projekt zu modellierenden über diese zu bearbeitenden Zeitschriften darstellt. Zu der Frage, welche Anforderungen eine solche Modellierung an eine adäquate und gabenbezogene Wissensrepräsentation stellt, wickelte Brigitte Endres-Niggemeyer konstruktive Überlegungen, die auf einen erweiterten Framework zielen. Abgeschlossen wurde dieser Schwerpunkt des Vormittagsprogramms durch die Demonstration der (PROLOG-) Implementierung eines ersten Prototypes für eine AUTOCAT-Wissensrepräsentation, die wesentlich auf der FRL-Sicht aufbaut und in Zusammenarbeit mit KEE und LOOPS integriert (Prof. Dr. Mann).

Am Nachmittag trugen Lothar Rostek und Dittrich Fischer (GID) vor: "Entwurfsskizzen für ein Thesaurusverwaltungssystem (am Beispiel des Thesaurus)". In das Zentrum ihrer Betrachtung stellten sie den von Nelson bereits 1967 eingeführten Begriff eines "Hypertextsystems". Als entscheidende Frage - und hier war der Berührungspunkt zu der Diskussion des Vormittags gegeben - war auch hier die nach einer geeigneten verfügbaren Wissensrepräsentationssprache zu beantworten. SMALLTALK als objektorientierte Programmiersprache mit einer komfortablen

Programmierungsumgebung und einem reichen Inventar an System-Klassen und -Methoden wird z.Z. genauer untersucht - nicht zuletzt deswegen, weil viele Alternativen in der Literatur leider nicht ausreichend detailliert beschrieben sind. Dies gilt allerdings gleichermaßen auch für PIE, eine auf SMALLTALK aufbauende Wissensrepräsentation.

Den Abschluß der Veranstaltung bildete ein Vortrag von Karin Ammersbach, die über ihre Diplomarbeit "Benutzermodelle für Information-Retrieval-Systeme" vortrug, in der sie Dialoge zwischen Endbenutzer und professionellem Rechercheur bei der Online-Recherche aufzeichnete und auswertete. Die Entwicklung eines "kooperativen Retrieval-Systems", das wesentliche Komponenten der sprachlichen, kommunikativen und fachlichen Fähigkeiten eines Informationsvermittlers simuliert, bildet den Hintergrund dieser Arbeit. Wenngleich sicherlich viele Fragen nicht abschließend geklärt sind (z.B.

die nach der Repräsentativität), stellt die Arbeit dennoch in einem Feld, in dem breite empirische Grundlagen nicht anders als durch Akkumulation vieler und bei aller Beschränktheit dennoch aufwendiger Einzeluntersuchungen zustandekommen, einen notwendigen und wichtigen Schritt dar.

Die Veranstaltung wurde von ca. 25 bis 30 Teilnehmern besucht, wobei neben einer starken "Hausmannschaft" der THD auch Vertreter der Fachinformatiker, anderer Universitäten (Informationswissenschaft, Linguistik) und der Industrieforschung sich sehr rege an den Diskussionen beteiligten. Ein Anschlußthema wurde auch bereits identifiziert: Die Problem- und Anwendungsabhängigkeit von Wissensrepräsentationskonzepten. Bedarf und Interesse an einer sorgfältigen Diskussion dieses Themas, das hat sich klar gezeigt, sind gegeben.

Gerhard Knorz, Technische Hochschule Darmstadt

LINGUISTIC TOOLS AND SOFTWARE TOOLS FOR MT

Kurzbericht vom "Second Saarbrücker Expert Talk on Machine Translation", 3. bis 4.4.1986

"Experience in the field of MT has led to a growing awareness of a lack of feasible tools available to linguists and computer scientists..." (aus der Einladung). Das IAI hatte geladen, und trotz der etwas problematischen Terminwahl (z.B. "konkurrierende" Veranstaltungen in der Woche später) hatten sich etwa ein Dutzend (Computer-)Linguisten und Informatiker auf den Weg nach Saarbrücken gemacht, so daß sich zusammen mit den ortsansässigen Mitarbeitern von IAI und/oder SFB 100 ein Auditorium von ca. 25 Teilnehmern ergab. (Für EUROTRA-Verhältnisse fehlte etwas die Internationalität: ein Vertreter von EUROTRAF, ein amerikanischer Referent, ansonsten "hiesige" bzw. in "hiesigen" Projekten arbeitende Teilnehmer).

Die Programmplaner hatten zwei je anderthalbstündige Vorträge vorgesehen (Martin Kay: "Linguistic Tools for MT", Jan Messerschmidt: "Software Tools for MT"), ansonsten Raum für Diskussionen und - am Freitagnachmittag - Vorführungen gelassen. Martin Kay (XEROX Parc, derzeit Universität Genf) faszinierte durch inhaltliche Brillanz und humorvollen Vortragstil. Er organisierte seine Ausführungen um drei "intellectual tools" und ihre aktuelle bzw. potentielle Relevanz für die MU: i) phrase structure grammars, ii) unification, iii) finite state automata. Besonders

ausführlich die Ausführungen zu Unifikation / motiviert war er hierzu wohl auch aufgrund seiner Lektüre von EUROTRA-Dokumenten: "the term is simply not understood in EUROTRA" (M. Kay).

Nach diesem Vordredner hätte jeder einen schweren Stand gehabt, aber auch inhaltlich konnte der Redner des Nachmittags (mich) nicht so recht begeistern: die Möglichkeit zur Verarbeitung von Strings beliebiger Länge in COMSKEE als Software-Lösung für die MÜ? Am Freitagmorgen dann noch einmal intensive - und wohl auch klärende - Diskussionen anhand aktuellen EUROTRA-Materials: so lieferten schon die ersten vier Zeilen der gezeigten Lexikoneinträge M. Kay den Beleg seiner Behauptung vom Vortrag frei Haus. Eine weitere Erkenntnis: Vieles ist bei EUROTRA noch offen - so gibt es z. B. bisher kaum mehr als Ansätze dazu, welche Inhalte die "EURO-versale" semantische Repräsentation wirklich enthalten sollte - und bedarf weiterer Parsing- und Generierungs-Software-Umgebungen unabdingbar (wenn auch, zumindest auf den ersten Blick, nicht unbedingt "billig zu haben) sind, wurde deutlich. Bleibt nur zu hoffen, daß sich davon auch die Geldgeber überzeugen lassen?

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

ARBEITSTAGUNG BEGRIFFSANALYSE, TH DARMSTADT

Die Forschungsgruppe Begriffsanalyse im Fachbereich Mathematik der Technischen Hochschule Darmstadt veranstaltete vom 17. bis 19. Januar 1986 eine Arbeitstagung "Begriffsanalyse", die von der Stiftung Volkswagenwerk gefördert wurde. Das Programm verlangte dem Teilnehmer einiges ab: Als Indikator mag die Tatsache dienen, daß das Tagungsprogramm z.B. am zweiten Tag der Veranstaltung, einem Samstag, vom 8:00 bis 19:20 angesetzt war! Im einzelnen wurden folgende Vorträge gehalten: B. Seiler (TH Darmstadt): Begriffe von Begriff: Analysen und Konzeptionen von Begriffen in der psychologischen Forschung; I. Dahlberg (INDEKS, Frankfurt): Die analytische, gegenstandsbezogene Begriffstheorie und ihre Definitionsarten; R. Ungvary (Nationalbibliothek, Budapest): Der Begriff der Relation; R. Wille (TH Darmstadt): Bedeutungen von Begriffsverbänden; V. Takacs (Nat. Center of Educational Technology, Veszprem): Concept lattices in pedagogical research; B. Monjardet (EHESS Paris): An ordinal generalisation of Guttman scales; W. Nedobity (Infoterm, Wien): G. Frege's Begriffsschrift aus terminologischer Sicht; G.

Knorz (TH Darmstadt): Begriffsstrukturen in Information Retrieval; G. Rahmstorf (IBM, Wiesbaden): Semantik und Wissensrepräsentation; L. Sieber (TH Darmstadt): Konnotative Wortfelder; H. von Petzinger (Univ. Paderborn): Begriffsanalyse; J. Zelger (Univ. Innsbruck): Philosophische Konzepte der Begriffsanalyse; J.-P. Doignon (Univ. Libre Brüssel): Knowledge assessment: a theoretic framework; H. Mönke (Stein a. d. Traße, Darmstadt): Grundzüge einer allgemeinen linguistischen Definitionstheorie; M. Kresic (Markt Schwaben): Intuition/Thesaurus mit einer Achsen-Begrifflichekeitsanalyse; V. Duquenne (Sorbonne, Paris): Applications between attributes; B. Ganter (TH Darmstadt): Algorithmen zur formalen Begriffsanalyse.

Wie Prof. Dr. R. Wille, der die sehr fruchtbare Arbeitstagung engagiert und souverän leitete, teilt, wird eine Auswahl der gehaltenen Vorträge in einen Tagungsband aufgenommen werden.

Gerhard Knorz, Technische Hochschule Darmstadt

Nachrichten

EIN SOFTWAREPAKET FÜR MORPHOLOGISCHE ANALYSE

Im SFB 100 wurde für das MÜ-System SUSY das Software-Paket WOBUSU entwickelt, das Wörter des Deutschen, Russischen, Englischen, Französischen und des Esperanto morphologisch zerlegt und ihnen eine morpho-syntaktische Beschreibung zuweist. Die Fehlerquote bzw. die Quote nicht erkannter Wörter ist sehr gering, sodaß man (zumindestens fürs Deutsche) sagen kann, daß jedem Wort der Sprache eine Beschreibung zugewiesen wird (Ausnahme: Namen). Für die einzelnen Sprachen stehen die folgenden Lexika zur Verfügung:

Deutsch:	SADAW	142.000 Einträge
Russisch:	RAWO	15.000 Einträge
Englisch:	ENGWOSYN	9.000 Einträge
Französisch:	DCTFRA	9.400 Einträge
Esperanto:	EWOBU	5.000 Einträge

Die eingegebenen Wörter werden einer Flexionsanalyse und - wenn der Stamm nicht im Lexikon gefunden wird - einer Derivations- und Kompositionanalyse unterzogen. Output ist pro Wort ein Wort-Record (bei Homographen ein Record pro Deutigkeit) mit dem folgenden Inhalt:

- dem (u.U. in Morpheme zerlegten) Lemmanamen
- einer morphologischen Beschreibung (entsprechend den 18 flektierenden (verbalen und nominalen) Kategorien)
- einer syntaktischen Beschreibung (aus dem entsprechenden Lexikoneintrag)

WOBUSU läuft auf einer Siemens-Anlage unter BS2000 und steht (wie im übrigen das ganze SUSY-System) für nicht-kommerzielle Zwecke allgemein zur Verfügung. Für kommerzielle Anwendungen können besondere Absprachen getroffen werden.

Leistung: ca. 10-15 Wörter/Sekunde CPU-Zeit

Kontaktadresse: Heinz-Dirk Luckhardt
SFB 100/A2
Universität des Saarlandes
D-6600 Saarbrücken
Tel.: 0681/302 2050 oder 3204

Die Adresse gilt bis Ende 1986, danach gilt die des Rechtsnachfolgers des SFB 100, ansonsten die der Universität des Saarlandes.

Heinz-Dirk Luckhardt, Universität Saarbrücken

Nachricht

THESAURUS-DATENBASIS BEI DER GESELLSCHAFT FÜR INFORMATION UND DOKUMENTATION (GID) VERFÜGBAR

Über tausend Fachvokabularien aus allen Wissensgebieten sind in der neuen Thesaurus-Datenbasis der GID beschrieben.

Die Gesellschaft für Information und Dokumentation mbH (GID), Frankfurt/Main, hat im Auftrag der Kommission der Europäischen Gemeinschaften (KEG) in Luxemburg, eine umfangreiche Datenbasis mit bibliographischen und inhaltlichen Angaben über Thesauri, die in wenigstens einer der offiziellen Sprachen der Gemeinschaft vorliegen, erstellt. Außer den EG-Ländern, insbesondere Frankreich, Großbritannien und der Bundesrepublik Deutschland, sind die USA und Kanada stark vertreten. Darüber hinaus enthält die Datensammlung englisch- und französischsprachige Thesauri aus aller Welt.

Ende 1983 wurden Sammlung und Analyse zunächst abgeschlossen. Nach der Korrespondenz mit mehreren hundert Thesaurusherstellern, die die Thesaurusbeschreibungen prüften und ergänzten, übergab die GID-Projektgruppe im Dezember 1984 der KEG die Lichtsatzvorlage für die Publikation "THESAURUS GUIDE" und ein Magnetband für den Aufbau einer entsprechenden Datenbank.

Die Datenbasis wurde 1985 von der GID weiter aktualisiert und hat im Dezember 1985 folgenden Stand:

Thesaurus-Beschreibungen insgesamt:	1 009
- ausführliche Beschreibungen:	498
- Kurzbeschreibungen:	511
- mehrsprachige Thesauri:	141
- englischsprachige Thesauri:	441
- französischsprachige Thesauri:	180
- deutschsprachige Thesauri:	226

Die ausführlichen Beschreibungen (gegliedert nach 10 Haupt-Fachgebieten, 51 Untergruppen) umfassen außer aktuellen bibliographischen Details vor allem Angaben über Inhalt und Struktur eines Thesaurus (z.B. Sachgebiete, Wortmaterial, Relationen, Darstellungsarten) sowie über dessen Anwendung, Zugänglichkeit u.a.

Die Kurzbeschreibungen betreffen neben einigen für eine ausführliche Analyse nicht erreichbaren Werken u.a. auch nicht mehr benutzte Thesauri aus "alten" Fachgebieten und sonstige dokumentarische Hilfsmittel (Subject Headings Lists, u.ä.), die immerhin zu einer Thesauruserstellung herangezogen werden könnten oder sonstige Informationen bieten.

Die gedruckte Version "THESAURUS GUIDE" wird durch drei jeweils dreisprachige Sachregister (deutsch-französisch-englisch) sowie ein Institutionen- und Personenregister ergänzt.

Die Thesaurus-Datenbasis dient u.a.

- der Wahl einer geeigneten Dokumentationssprache
- der Vermeidung von Doppelarbeit bei der Thesaurusneuerstellung
- der Standardisierung von Thesauri
- Forschung und Lehre.

Nachrichten

zugänglich sind jetzt:

1. THESAURUS GUIDE.

Analytical Directory of Selected Vocabularies for Information Retrieval, 1985.

Amsterdam, NL: North Holland 1985. XXXVI + 748 pages.

ISBN 0-444-87736-3. US \$ 74.00 / Dfl. 200.00

Das Werk kann beim Verlag direkt bestellt oder über den Buchhandel bezogen werden.

2. Database THESAURI

ECHO - European Commission Host Organization

Customer Service, 177, route d'Esch, L-1471 Luxembourg

Tel.: 352 48 80 41, Telex: 2 181

3. Thesaurussammlung

Gesellschaft für Information und Dokumentation mbH (GID)

Informationszentrum (IZ), Bibliothek

Lyoner Straße 44-48, D-6000 Frankfurt am Main 71,

Tel.: 49 (0 69) 66 87-374

Ein großer Teil der im "THESAURUS GUIDE" und in der Datenbank nachgelesenen Thesauri ist in der Bibliothek des GID-Informationszentrums (GID-IZ) auch verfügbar. Entsprechende Hinweise finden sich in den Thesaurusbeschreibungen.

Die ursprünglich schon vorhandene Thesaurussammlung des IZ hat sich im Laufe des Projekts zu einer der größten dieser Art entwickelt.

Die GID bleibt weiterhin zuständig für die Pflege der Datenbasis und versorgt Verlag, Host und IZ-Bibliothek mit aktuellem Material. Darüber hinaus leistet sie auch Beratung zu Thesaurusfragen unterschiedlichster Art.

Mit dem Dank an all jene, die das Projekt bisher unterstützt haben, verbindet sich die Bitte an alle - ehemalige und zukünftige - Thesaurushersteller, der GID-Projektgruppe (weiterhin) ihre Thesauri (auch leihweise) bzw. relevante Kopien daraus (Einführung, mehrere typische Seiten der Thesaurusteile) sowie weitere Angaben im Sinne der Beschreibungen im "THESAURUS GUIDE" bzw. in der Datenbank zu übermitteln.

Kontaktadresse für Thesaurusfragen und Aktualisierung der Datenbasis:

Dr. Alfred Gerstenkorn

Gesellschaft für Information und Dokumentation mbH (GID)

Herriotstraße 5, D-6000 Frankfurt am Main 71

Tel.: 49 (0 69) 66 87-299

Pressemitteilung der GID

BOCHUMER SEMIOTISCHES COLLOQUIUM 1985/86

Das Bochumer Semiotische Colloquium (BSC) veranstaltete im Wintersemester 1985/86 folgende Vorträge: Robert E. Innis (Lowell, Mass.): "Articulation as Emendation. Philipp Wegener's Anti-formalist Theory of Language.". Erhardt Güttgemanns (Bonn): "Semiotik und Theologie". Jan Wind (Amsterdam): "Die Soziobiologie und die Humanwissenschaften". Ann Shuckman (Oxford): "The Semiosphere". Herta Schmid (München): "Otakar Zich und der Prager Linguistische Kreis". Die Vortragsreihe wird fortgesetzt.

Semiotischer Arbeitskreis Berlin

Im

Studiengang Angewandte Informatik/Schwerpunkt Linguistik

der

Erziehungswissenschaftlichen Hochschule Rheinland-Pfalz
Abteilung Koblenz

sind im

WS 1986/87 (13.10.1986 - 14.02.1987)

vorbehaltlich der Mittelbereitstellung

LEHRAUFTRÄGE

für die folgenden Veranstaltungen zu vergeben:

- Formale Grammatiken (V/Ü, 4std)
- Automatische Spracherkennung und -erzeugung V/S,
- Textverarbeitung und Büroautomation (V/S, 2std)
- Codier- und Editionstechniken (V/S, 2std)
- Computerunterstützter Unterricht (V/S, 2std)

Wer Interesse hat (möglichst in der Nähe von Koblenz), einen der ge-
ten Lehraufträge zu übernehmen, setze sich bitte wegen weiterer In-
formation in Verbindung mit

Prof. Dr. I. Bátor
Studiengang Angewandte Informatik
Schwerpunkt Linguistik
EWH RLP, Abt. Koblenz
Rheinau 3 - 4
5400 Koblenz-Oberwerth
Tel. 0261/1 21 56

KOMMUNIKATORMODELLE UND IHRE FORMALISIERUNG RINGVORLESUNG AN DER TU BERLIN

Im Wintersemester 1986/86 fand am Insti-
tut für Linguistik der TU Berlin in Zusammenar-
beit mit dem Semiotischen Arbeitskreis Berlin
(SAB) eine Ringvorlesung über "Kommunikator-
modelle und ihre Formalisierung" statt. Folgende
Vorträge wurden gehalten: Roland Posner (TU
Berlin): "Signal, Symptom, Geste, Kommuni-
kation" (21.10.85). "Wissen Bewirken, Intendie-
ren" (28.10.85). "Kommunikationsversuche verste-
hen: ohne Kode, mit Kode, mit Sprache" (4.11.85).
Pavel Tichy (Otago, Neuseeland): "The Notation
of Language" (11.11.85). "A Fallacy in Intatio-
nal Logic" (18.11.85). Klaus Robering (TU Berlin):
Darstellung von Gesprächsverläufen mittels des An-
satzes von Sneed und Steegmüller (25.11.85). Mar-

celo Dascal (Tel Aviv): "The Process of Under-
standing and the Possibilities of its failure" (28.11.85).
Maria Lucia Santaella Braba (Sao Paulo): "Pe-
ripheral Signs and his Rational Psychology"
(2.11.85). Konrad Burdach (München): "Ge-
stalt als Zeichen und die Prozesse ihrer Verarbeit-
ung" (5.12.85). Rolf Haberbeck (FU Berlin): "Die ge-
schichtliche Darstellung von Modellen in der Zeichenthe-
orie" (9.12.85). "Vorschläge für neue Anwendungs-
modelle der Katastrophentheorie" (16.12.85). Peter Sch-
äfer (Münster): "Semasiologische Konzeptionen der
Zeichenverarbeitung und Zeichenwandel" (9.1.86).
Manfred Krause (TU Berlin): "Modelle der inte-
rioren Repräsentation von Denkobjekten" (13.1.86).
"Wissensbasen und Konventionen in der Künstlichen
Intelligenz-Forschung" (20.1.86). Die Vorträge
wird fortgesetzt.

Semiotischer Arbeitskreis Berlin.

SKRIPTENFOLGE ZUR INFORMATIONSLINGUISTIK AM INFORMATIONSWISSENSCHAFTLICHEN LEHRSTUHL DER UNIVERSITÄT KONSTANZ

Am Lehrstuhl für Informationswissenschaft der Universität Konstanz wurde von 1982-1985 das vom BMT geförderte Projekt "Informationsvermittlung" durchgeführt. Neben der Beschäftigung mit genuin informationswissenschaftlichen Forschungsthemen - u.a. ist das TOPIC-Projekt in dieser Umgebung angesiedelt, das die Entwicklung wissensbasierter Verfahren zur Volltextanalyse und Textkondensierung (Abstracting) zum Thema hat - war dieses Projekt durch einen besonderen curricular-infrastrukturellen Forschungsauftrag charakterisiert. Die Planung und Implementierung eines eigenständigen Aufbau- und Diplom-Studiengangs Informationswissenschaft umfaßte dabei auch ganz gezielt die Entwicklung geeigneter Lehrmaterialien (Skripten, Aufgabensammlungen, Lehr-Software usw.).

Die Ergebnisse dieser curricularen Aufgabenstellung des Projekts "Informationsvermittlung" sind für das Teilgebiet der Informationswissenschaft, das sich mit natürlichsprachlichen Problemen der rechnergestützten Informationsvermittlung befaßt, in einer dreiteiligen Skriptenfolge zur Informationslinguistik dokumentiert:

- (1) HAHN, Udo (1985a)
 Informationslinguistik I. Einführung in das linguistische Information Retrieval. Konstanz: Univ. Konstanz, Informationswissenschaft, Projekt Informationsvermittlung, (März 1983) Januar 1985, 180pp. (= Bericht CURR-4/85).
- (2) HAHN, Udo (1985b)
 Informationslinguistik II. Linguistische und statistische Verfahren im experimentellen Information Retrieval. Konstanz: Univ. Konstanz, Informationswissenschaft, Projekt Informationsvermittlung, (Juli 1983), Januar 1985, 175pp. (= Bericht CURR-5/85).
- (3) HAHN, Udo (1985c)
 Intelligente Informationssysteme: Verfahren der Künstlichen Intelligenz im experimentellen Information Retrieval. Konstanz: Univ. Konstanz, Informationswissenschaft, Projekt Informationsvermittlung, (Februar 1984) Februar 1985, 381pp. (= Bericht CURR-6/85).
 (Kursabschnitt zu Wissensrepräsentationsverfahren und natürlichsprachlichen Systemen entspricht "Informationslinguistik III")

Während diese Skripten eine detaillierte Übersicht über informationslinguistische Lehrinhalte geben, wird in

- (4) HAHN, Udo (1986)
 Curriculum Informationslinguistik. Zur inhaltlichen und didaktischen Konzeption informationeller Sprachanalyse am Beispiel des Diplom-Studiengangs Informationswissenschaft an der Universität Konstanz. Konstanz: Univ. Konstanz, Informationswissenschaft, März 1986, 46pp. (= Bericht CURR-12/86).

der Versuch einer Bestandaufnahme des im Verlauf dieses Projekts entwickelten curricularen Konzepts zur Informationslinguistik unternommen.

In der Anlage sind zur Informierung über die Inhalte der informationslinguistischen Kurse die Inhaltsverzeichnisse der drei Skripten wiedergegeben. Die Skripten (1)-(3) sowie Bericht (4) können angefordert werden

- | | |
|---|---------|
| (1) Skript "Informationslinguistik I" | DM 25.- |
| (2) Skript "Informationslinguistik II" | DM 25.- |
| (3) Skript "Intelligente Informationssysteme" | DM 50.- |

Die genannten Anlagen konnten leider aus Platzgründen nicht in die vorliegende Ausgabe aufgenommen werden, was die Redaktion außerordentlich bedauert. Die Skriptgliederungen werden in der nächsten Ausgabe abgedruckt und können zwischenzeitlich bei der Redaktion bzw. direkt bei U. Hahn angefordert werden.

**PROJEKT AUTOCAT AN DER TH DARMSTADT
PROJEKT ZUR WISSENSBASIERTEN
BIBLIOGRAPHISCHEN DOKUMENTERSCHLISSUNG**



Fachbereich Informatik
Forschungsgruppe
Datenverwaltungssysteme

1. Die Aufgabe: Automatische Formalkatalogisierung

AUTOCAT soll automatisch Titelaufnahmen von Aufsätzen aus Physikzeitschriften nach INIS-Regeln ableiten. Das Originaldokument muß in seinen bibliographisch relevanten Teilen maschinenlesbar verfügbar sein. Prinzipiell soll AUTOCAT allen im Anwendungsgebiet vorkommenden Formen der Datenübergabe (vor allem Daten vom Satzband, voll kategorisierte Daten bei elektronisch publizierten Dokumenten, optisch gelesene Daten) gewachsen sein; es setzt nur eine minimale Strukturkennzeichnung voraus. AUTOCAT übernimmt die Daten und zerlegt sie anhand der Strukturinformationen in Informationsblöcke. Das System benutzt seine Wissensbasis über die Informationsstrukturen in Zeitschriften, um die Informationsrollen der Informationsblöcke zu erkennen. Die Daten werden nach INIS-Regeln transformiert und in das INIS-Datenschema eingetragen. Die von AUTOCAT produzierte Dokumentationseinheit wird in eine Datenbank geschrieben, in der sie einem Bearbeiter zwecks Kontrolle und Ergänzung (Indexierung, Überarbeitung des Autorenreferates usw.) zur Verfügung steht.

2. Das AUTOCAT-System

Formalerfassung ist ein stark strukturorientierter Prozeß, der sich in einzelne in einem hohen Grad voneinander unabhängige Arbeitsgänge mit gut definierten Schnittstellen zerlegen läßt.

Die Systementwickler finden in den Arbeitsmitteln der intellektuellen Formalerfassung (Regelwerke, Kategorien, Schemata u.s.w.) einen Stand der Modellierung vor, der die Konzeption eines wissensbasierten Systems erleichtert. Das AUTOCAT-System vollzieht den in Prinzip zweistufigen Arbeitsablauf: Informationselemente erkennen - Informationselemente transformieren in seiner Systemstruktur nach.

Da AUTOCAT ein übertragbares und erweiterungsfähiges System werden soll, wird es in einen aufgabenunabhängigen Teil (AUTOCAT-SHELL) und einen aufgabenabhängigen Teil (Wissensbasen und Datenbanken) untergliedert. Die weitere Systemorganisation soll sich an ein "Expertenmodell" anlehnen.

AUTOCAT wird insgesamt als Produktionssystem in der Programmiersprache PROLOG realisiert.

Das System braucht zusätzliche Softwarewerkzeuge, insbesondere ein Datenbanksystem zur Verwaltung von Massendaten.

AUTOCAT greift ein für wissensbasierte Systeme neues Anwendungsgebiet auf. Seine Systemarchitektur muß den Anforderungen der Massenproduktion von Dokumentationseinheiten, der Anpassbarkeit an verschiedene Formalerfassungsaufgaben und der Interaktion mit Menschen (Lern- und Erklärungs-fähigkeit) gerecht werden.

Projektleitung:

Prof. Dr. Brigitte Endres-Niggemeyer, Fachhochschule Hannover
Dr. Gernard Knorz, Technische Hochschule Darmstadt

Mitarbeiter:

Peter Müller, Dipl.-Inform.
Ulrike Rauth, Dipl.-Inform.
Christina Schnellbach, Dipl.-Dokum.

Adresse:

Technische Hochschule Darmstadt
Fachbereich Informatik
Fachgebiet Datenverwaltungssysteme II
Projekt AUTOCAT
Karolinenplatz 5
6100 Darmstadt

Tel. (06151) 16-3952, 16-5417

Förderungsprogramm:

BMFT-Projekt Nr. 10200170

Projektpartner:

Fachinformationszentrum Energie Physik Mathematik, (FIZ 4),
Eggenstein, Leopoldshafen

Universität Saarbrücken, Informatik IV
Fachhochschule Hannover, Fachbereich BID

Projektdauer:

01.10.86 bis 30.09.87

Publikationen

die Bibliographie "Neuere Veröffentlichungen im Bereich Linguistische Datenverarbeitung" werden im LDV-FORUM vom

LDV-Informationszentrum
Lyoner Strasse 44-48
D-6000 Frankfurt 71

Standardprofile **Linguistische Datenverarbeitung** und **Künstliche Intelligenz** der Datenbasis INFO-TA regelmäßig zur Verfügung stellt. Die weitere Verarbeitung dieser Standardprofile umfaßt eine inhaltliche Selektion, um Arbeiten ohne Bezug auf die Verarbeitung natürlicher Sprache auszuschließen (die Grenze ist fließend und subjectiv gezogen).
eine formale Selektion (nur deutsch- und englischsprachige Arbeiten, nur Arbeiten, die nach 1983 erschienen sind).
eine radikale Verdichtung auf wesentliche bibliographische Angaben.

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

Wer, die hier aufgeführte Literatur bestellen wollen, wenden sich das GID/IZ:

Herrn A. Stelzer, GID-IZ, Lyonerstr. 44-48, F-71
Tel.: (069) 66 87-378 oder 66 87-362

verarbeitete Ausschnitte der INFODATA-Datenbank: Standardprofil vom 22.11.1985 (Erf-Jahr 8509, Erf-Nr. 3301-3900 bis Erf-Jahr 8510, Erf-Nr. 3901-4200); Standardprofil vom 30.1.1986 (Erf-Jahr 8511, Erf-Nr. 4101-4700).

Aufbau einer Dokumentationseinheit:

Beispiel: **Der Titel in Fettdruck, mit einem Punkt abgeschlossen.** Die Quelle, mit Seitenangabe, in Klammern: *Deskriptor1; Deskriptor2; ...; Deskriptor-n.* Ref.-Nummer

van Berghem, C. J. van (Hrsg.): **Research and development in information retrieval.** Proceedings of the 1984 joint BCS and ACM symposium, King's College, Cambridge (1984) 433 S., *Information Retrieval; Information Retrieval System; Datenbank; Künstliche Intelligenz; Computerlinguistik; Software; Mensch-Maschine-Kommunikation; Büroorganisation; Maschinelles Indexierungsverfahren.* 85-03328

Christodoulakis, S.: **Framework for the development of an experimental mixed-mode message system.** S. 1-20. *Kommunikationssystem; Büroorganisation; Mensch-Maschine-Kommunikation; Entwurf; Prototyp; Test.* 85-03329

Laender, A. H. F.; Stocker, P. M.: **An interactive database end user facility for the definition and manipulation of forms.** S. 41-54. *Mensch-Maschine-Kommunikation; Datenbank; Zugriff; Datenbankaufbau; Information Retrieval System; Kommandosprache; Suchfrage; Formular; Benutzerführung; Formaldialog* 85-03331

Hollaar, L. A.: **The Utah text retrieval project. A status report.** S. 123-132. *Information Retrieval System; Software; Hardware; Mensch-Maschine-Kommunikation; Projekt; USA; Utah, Systemarchitektur, Benutzerschnittstelle.* 85-03336

Kwok, K. L.: **A document-document similarity measure based on cited titles and probability theory, and its application to relevance feedback retrieval.** S. 221-231. *Maschinelles Indexierungsverfahren; Zitat; Information Retrieval; Ranking; Relevanz; Feedback; Rechercheergebnis; Relevanzfeedback, probabilistisches Retrieval.* 85-03342

Cooper, W. S.: **Bridging the gap between AI and IR.** S. 259-265. *Information Retrieval; Künstliche Intelligenz; Forschung; Logik; Linguistik.* 85-03345

Defude, B.: **Knowledge based systems versus thesaurus. An architecture problem about expert systems design.** S. 267-280. *Expertensystem; Integriert; Information Retrieval System; Thesaurus; Entwurf; Systemarchitektur.* 85-03346

Metzler, D. P.; Noreault, T.; Richey, L.; Heidorn, B.: **Dependency parsing for information retrieval.** S. 313-324. *Computerlinguistik; Information Retrieval; Syntax; Dependenzgrammatik; Parser.* 85-03349

Hahn, U.; Reimer, U.: **Computing text constituency. An algorithmic approach to the generation of text graphs.** S. 343-368. *Maschinelles Referieren; Textanalyse; Verdichtung; Graphische Darstellung; Information Retrieval System; Algorithmus.* 85-03351

Niedermair, G. T.; Thurmaier, G.; Buettel, I.: **MARS: A retrieval tool on the basis of morphological analysis.** S. 369-381. *Computerlinguistik; Morphologie; Software; Datenbank; Recherche; Volltextsuche.* 85-03352

- Frakes, W. B.: **Term conflation for information retrieval.** S. 383-389. *Volltextsuche; Recher-
tegie; Maskierung; Information Retrieval System; Morphologie; Methode; Test.* 85-03353
- Fuhr, N.; Knorz, G. E.: **Retrieval test evaluation of a rule based automatic indexing (AIR).**
S. 391-408. *Maschinelles Indexierungsverfahren; Bewertung; Vergleich; Indexierungsverfahren;
Test; Information Retrieval.* 85-03354
- Jones, K. P.; Bell, C. L. M.: **The automatic extraction of words from texts especially for in-
formation retrieval systems based on inverted files.** S. 409-420. *Maschinelles Indexierungs-
ren; Auswahl; Maschinell; Dateneingabe; Information Retrieval; Datei; Invertiert; Wörterbuchera-
Textinvertierung.* 85-03355
- Stratton MacAllister, A.: **Systems design for a turnkey local library system DOBIS/LIBIS.** In
A. H. Weiss, J. W. (Hrsg.): Essen symposium September 1984. (Festschrift in honour of F. G. Kilgou-
70th birthday. Essen, 1984). S. 259-266. ISBN 3-902602-08-8. *Bibliotheksautomation; Mensch-Ma-
Kommunikation; Software; Benutzerfreundlich; Systemforschung.* 85-03370
- Sokolova, E. N.: **Reduced text and automated information systems.** Automatic documentat-
mathematical linguistics. New York, US 17(1983)5, S. 8-10. *Algorithmus; Textanalyse; Bewertung*
03376
- Shemakin, Yu. I.; Kuznetsov, V. E.: **A model for representation of a conceptual system in
translation.** S. 31-37. *Maschinelle Übersetzung; Künstliche Intelligenz; Mathematisches Mod*
03379
- Ovchinnikova, N. O.: **Factors affecting the functioning of recurring nouns as true
Intersentence relations.** S. 38-47. *Textanalyse; Englisch; Wortbedeutung.* 85-03380
- Afonichkin, I. A.: **Information analysis in automated data systems.** Automatic documentat-
mathematical linguistics. New York 17(1983)5, S. 87-90. *Informationssystem; Sprache; Semantik*
Künstliche Intelligenz; Bewertung. 85-03385
- Kaloshin, V. V.: **Analysis of the development trends of information science through logic
semantic modeling of the subjects of dissertations.** Scientific and technical information processin-
York 17(1983)4, S. 56-68. *Informationswissenschaft; Textanalyse; Semantisches Netz; Logik; Algor*
85-03605
- Manukyan, G. G.: **Method of statistical automatic coding of natural language words.** Automat-
mentation and mathematical linguistics. New York 17(1983)6, S. 30-35. *Sprachstatistik; Russisch*
Morphologie; Wortbildung; Information Retrieval System; Algorithmus. 85-03619
- Maassen, B. (Projl.): Deutsches Bibliotheksinst. (DBI). Berlin: **PRECIS.** Berlin. (1984) 339
materialien 35. *Syntaktisches Indexieren; Maschinelles Indexierungsverfahren; Richtlinie; Bibliothek.* 85-03620
- Michalski, R. S.; Carbonell, J. G. (Hrsg.): Mitchell, T. M.: **Machine learning. An artificial Intel-
approach.** Palo Alto: Tioga (1983) 583 S., ISBN 0-935382-05-4. *Künstliche Intelligenz; Lernen; Sys-
schung; Heuristik; Inferenz; Modell; Expertensystem.* 85-03673
- Carbonell, J. G.; Michalski, R. S.; Mitchell, T. M.: **An overview of machine learning.** S. 1-10.
Künstliche Intelligenz; Lernen; Simulation; Maschinell; Expertensystem. 85-03674
- Dietterich, T.; Michalski, R. S.: **A comparative review of selected methods for learning
examples.** S. 41-81. *Künstliche Intelligenz; Lernen; Induktion; Frame; Vergleich.* 85-03675
- Michalski, R. S.: **A theory and methodology of inductive learning.** S. 83-134. *Künstliche Int-
Lernen; Induktion; Heuristik; Datenanalyse; Methode.* 85-03676
- Carbonell, J. G.: **Learning by analogy. Formulating and generalizing plans from past exper-
S. 137-161.** *Künstliche Intelligenz; Lernen; Analog; Inferenz; Modell.* 85-03677
- Mitchell, T. M.; Utgoff, P. E.; Banerji, R.: **Learning by experimentation. Acquiring and
problem-solving heuristics.** S. 163-190. *Künstliche Intelligenz; Lernen; Heuristik; Software; So-
sches Netz.* 85-03678
- Lenat, D. B.: **The role of heuristics in learning by discovery. Three case studies.** S. 21-30.
Künstliche Intelligenz; Lernen; Heuristik; Expertensystem; Fallstudie. 85-03679
- Haas, N.; Hendrix, G. G.: **Learning by being told. Acquiring knowledge for information fo-
gement.** S. 405-427. *Mensch-Maschine-Kommunikation; Lernen; Frage-Antwort-System; Humaf*
85-03682
- Sleeman, D. H.: **Inferring student models for intelligent computer-aided instruction.** S. 43-52.
Künstliche Intelligenz; Lernen; Inferenz; Modell; Systemforschung; Rechnerunterstütztes Lerne
03684
- Lee, R. M.: **On information system semantics. Expert vs. decision support systems.** Social
information studies. Guildford, 5(1985)1, S. 3-10. *Expertensystem; MIS; Systemforschung; Semant*
03759
- Wahlster, W.: **Cooperative access systems. Future generations computer systems.** Amsterdam 11
S. 103-111. *Verarbeitung Gesprochener Sprache; Mensch-Maschine-Kommunikation; Expertensystem*
03785

- kamura, K.; Sage, A. P.; Iwai, S.: **Associative learning using similarity knowledge bases for relational database search.** Future generations computer systems, Amsterdam 1(1984)2, S. 123-133. *Systemforschung; Datenbank; Entwurf; Lernen; Frage-Antwort-System; Semantisches Netz.* 85-03787
- Richard, J. A. T.: **Electronic indexing and hard-copy management.** Manchester, GB (1983) 197 S., *Büroorganisation; Indexieren; Maschinelles Indexierungsverfahren; Dokumentart.* 85-03811
- Anderson, J. R.; Reiser, B. J.: **The LISP tutor. It approaches the effectiveness of a human tutor.** Byte, Scarborough, 10(1985)4, S. 159-175. *Programmiersprache; Künstliche Intelligenz; Rechnerunterstütztes Lernen; Fallstudie.* 85-03812
- Ingert, F.: **Automated indexing based on SNOMED.** Methods of information in medicine, Stuttgart, (1985)1, S. 27-34. *Maschinelles Indexierungsverfahren; Code; Medizin.* 85-03822
- Weyfus, H. L.: **Die Grenzen künstlicher Intelligenz. Was Computer nicht können.** Königstein, Athenäum (1985) 373 S., ISBN 3-7610-8369-6 *Künstliche Intelligenz; Forschung und Entwicklung; Kritik; Geschichte; Humanfaktor; Lernen.* 85-03873
- Robidity, W.: **Terminology and artificial intelligence.** International classification, 12(1985)1, S. 17-19. *Künstliche Intelligenz; Mensch-Maschine-Kommunikation; Terminologie; Maschinelle Übersetzung.* 85-03892
- Arth, W.; Nirschl, H.: **Sichere sinnentsprechende Silbentrennung für die deutsche Sprache.** Angewandte Informatik, (1985)4, S. 152-159. *Textverarbeitung; Wort; Algorithmus; Deutsch; Silbentrennung.* 85-03935
- Engelmann, U.; Meinzer, H. P.: **Bessere Mensch/Maschine-Schnittstellen durch Beachtung von Benutzerfehlern.** Angewandte Informatik, (1985)5, S. 191-196. *Mensch-Maschine-Kommunikation; Humanfaktor; Fehler; Datenausgabe; Bildschirm.* 85-03936
- Behrich, J.: **Fehlerlose automatische Silbentrennung in fachsprachlichen Texten.** Angewandte Informatik, (1985)5, S. 198-206. *Textverarbeitung; Wort; Algorithmus; Deutsch; Silbentrennung.* 85-03937
- Imprecht, H.: **Worterkennung auf der Basis partieller Übereinstimmung von Wortfragmenten.** Informatik, Berlin, DD 32(1985)1, S. 32-34. *Indexierungsverfahren; Rechnerunterstützt; Morphologie; Wortbildung; Mustererkennung.* 85-03944
- Wermann, A. W.; Rodman, R. D.; Rubin, D. C.; Heidlage, J. F.: **Natural language with discrete speech mode for human-to-machine communication.** Communications of the ACM, 28(1985)6, S. 628-636. *Textverarbeitung; Gesprochener Sprache; Mensch-Maschine-Kommunikation; Bewertung* 85-03980
- Wuft, A. L.: **Prädikation, Regulation und Abstraktion. Ein Beitrag zu den philosophischen und terminologischen Grundlagen der Wissensrepräsentation.** Angewandte Informatik, (1985)3, S. 91-100. *Künstliche Intelligenz; Linguistik; Softwaretechnologie; Terminologie; Wissensrepräsentation.* 85-03984
- Wmmon, R. von: **Das Evaluationskonzept des automatischen Übersetzungsprojekts SUSY-DJT (Deutsch-Japanische Titelübersetzung).** Multilingua, Amsterdam, 3 u. 4(1984 u. 1985)4 u. 1, S. 189-195, 27-33. *Maschinelle Übersetzung; Deutsch; Japanisch; Bewertung; Test.* 85-03988
- Walt, D. L.; Pollack, J. B.: **Massively parallel parsing. A strongly interactive model of natural language interpretation.** Cognitive science, 9(1985)1, S. 51-74. *Computerlinguistik; Textanalyse; Semantisches Netz; Wissensrepräsentation.* 85-03990
- Robertson, I. T.: **Human information-processing strategies and style.** Behaviour and information technology, London, 4(1985)1, S. 19-29. *Informationsverhalten; Lernen; Recherchestrategie; Humanfaktor; Mensch-Maschine-Kommunikation; Typologie.* 85-03998
- Greenberg, S.; Witten, I. H.: **Adaptive personalized interfaces. A question of viability.** Behaviour and information technology, London, 4(1985)1, S. 31-45. *Mensch-Maschine-Kommunikation; Informationsverhalten; Dialogsystem; Benutzerführung; Kommandosprache; Empirische Untersuchung.* 85-03999
- Wertelsmann: **Datenbankdienste.** München: **Proceedings of the International Database Convention INFOBASE '85 held in Frankfurt, May 7-9, 1985.** München, (1985) 529 S. *Dialogdienst; Datenbankaufbau; Datendokumentation; Patentedokumentation; Graphische Datenverarbeitung; Verarbeitung Gesprochener Sprache; Informationstechnologie; Informationsindustrie; Informationspolitik; International; Entwicklungstendenzen.* 85-04001
- Schumacher, G.: **Entwicklungsstand neuer Datenbank-Retrieval-Systeme für Endbenutzer.** S. 285-306. *Mensch-Maschine-Kommunikation; Datenbank; Benutzerführung; Entwicklungstendenz.* 85-04021
- Tetschner, W.: **Voice-synthesis units for databases.** S. 377-385. *Verarbeitung Gesprochener Sprache; Mensch-Maschine-Kommunikation; Büroorganisation.* 85-04026

- Langley, P.: **Learning to search. From weak methods to domain-specific heuristics.** Cognitive Science, Norwood, NJ, US 9(1985)2, S. 217-260. *Künstliche Intelligenz; Lernen; Recherchestrategie; Heuristiken*. 85-04040
- Gevarter, W. B.: **Artificial Intelligence, expert systems, computer vision, and natural language processing.** Park Ridge, NJ: Noyes Publ. (1984) 240 S., ISBN 0-8155-0994-4 *Künstliche Intelligenz; Expertensystem; Mustererkennung; Verarbeitung Gesprochener Sprache; Automation; Wissensrepräsentation*. 85-04054
- Forsyth, R. (Hrsg.): **Expert systems. Principles and case studies.** London: Chapman and Hall (1984) 243 S., ISBN 0-412-26280-0, *Expertensystem; Systemforschung; Inferenz; Software; Fallstudie*. 85-04085
- Forsyth, R.: **Machine learning strategies.** S. 153-167. *Künstliche Intelligenz; Computerprogrammierung*. 85-04085
- Smith, S. F.: **Adaptive learning systems.** S. 168-189. *Expertensystem; Algorithmus; Künstliche Intelligenz; Heuristik*. 85-04086
- Rada, R.: **Automating knowledge acquisition.** S. 190-210. *Expertensystem; Künstliche Intelligenz; Systemforschung; Algorithmus*. 85-04087
- Ford, F. N.: **Decision support systems and expert systems. A comparison.** Information Management, Amsterdam, 8(1985)1, S. 21-26. *Dialogsystem; Entwicklungshilfe; Expertensystem; Wissensrepräsentation*. 85-04101
- Stall, D. S.: **Vollsynthese deutscher Sprache.** In: Ehrich, H. D. (Hrsg.): Gesellschaft für Informatik, Bonn: GI-14. Jahrestagung, Proceedings, Braunschweig, Oktober 1984. Berlin: Springer (1984) S. 1-14. *Verarbeitung Gesprochener Sprache; Computerlinguistik; Algorithmus*. 85-04116
- Blanning, R. W.: **Expert systems for management. Research and applications.** Journal of Information Science, 9(1984)4, S. 153-162. *Expertensystem; Entscheidungshilfe; Manager; Verwaltung; Informationsmanagement; Knowledge engineering* 85-04119
- Nievergelt, J.: **Die Gestaltung der Mensch-Maschine-Schnittstelle.** In: Kupka, I. (Hrsg.): GI-14. Jahrestagung, Proceedings, Hamburg, Oktober 1983. Berlin: Springer (1983) S. 41-50. *Informatik-Fachberichte; Systemforschung; Mensch-Maschine-Kommunikation; Benutzerführung; Humanfaktor; Kritik; Darstellung*. 85-04127
- Crawford, W.: **The RLIN Command Analysis System. Measuring use and performance of an expert system.** Information technology and libraries, Chicago 4(1985)1, S. 29-51. *Dialogsystem; Kommunikation; Mensch-Maschine-Kommunikation; Software; Bewertung; Feedback*. 85-04149
- Gardarin, G.; Gelenbe, E. (Hrsg.): **New applications of data bases. Based on proceedings of a workshop held at Cambridge, England on September 1983.** London: Academic Press (1984) 285 S. *Systemforschung; Mensch-Maschine-Kommunikation; Dialogsystem; Abbildung; Text; Expertensystem; Software; CAD*. 85-04165
- Crehange, M.; Haddou, A. A.; Boukakiou, M.; David, J. M.; Foucaut, O.; Maroldt, J.: **EXPERT: an expert system to aid in progressive retrieval from a pictorial and descriptive database.** *Expertensystem; Information Retrieval; Abbildung; Systemumgebung; Software; Prototyp*. 85-04165
- Bancilhon, F.; Richard, P.: **Managing texts and facts in a mixed data base environment.** *Systemforschung; Datenbanksystem; Fakten; Text; Integriert; Schnittstelle*. 85-04169
- Merrett, T. H.: **First steps to algebraic processing of text.** S. 109-127. *Computerlinguistik; Textanalyse; Mathematische Methode; Konkordanz; Wort; Statistik*. 85-04170
- Boguraev, B. K.; Sparck Jones, K.: **A natural language front end to databases with evaluation and user feedback.** S. 159-182. *Frage-Antwort-System; Mensch-Maschine-Kommunikation; Feedback; Benutzerführung; Methode*. 85-04173
- Nagao, M. (Hrsg.): **(Natürliche Sprache in Datenverarbeitungssystemen). Japan. Original: T. A. Dokumentationshinweise zum 1. Oktober 1984. Ergebnisse der japanischen Forschung (Ergebnisbericht).** Kybernetik, (1984) 342 S., *Maschinelle Übersetzung; Computerlinguistik; Japanisch; Englisch; Französisch; Projekt*. 85-04195
- Ibuki, J.; Kume, M.; Nagov, M.; Tsujii, J.: **Japanese-to-English title translation system. Title translation: its outline and the handling of special expressions in titles.** S. 1-35. *Maschinelle Übersetzung; Zeitschrift; Japanisch; Englisch*. 85-04196
- Hubert, J.: **Japanese-French automatic translation system for paper titles.** S. 36-86. *Maschinelle Übersetzung; Japanisch; Französisch; Englisch; Vergleich; Titel; Zeitschrift; Software*. 85-04197
- Trapp, R. (Hrsg.): **Cybernetics and systems research 2. Proceedings of the seventh European meeting, held at the University of Vienna, Austria, April 1984.** Amsterdam: North-Holland (1984) 243 S., *Kybernetik; Künstliche Intelligenz; Systemforschung; Biologie; Medizin; Operations Research*. 85-04237
- Leeming, A. M. C.: **A systems study of the design of a user interface.** S. 615-620. *Mensch-Maschine-Kommunikation; Benutzerbedarf; Entwurf; Software; Mikrocomputer; Informationssystem; Benutzerführung*. 85-04237

Publikationen

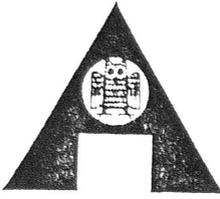
- Keravnou, E. T.; Johnson, L.: **Design of expert systems from the perspective of conversation theory methodology.** S. 651-654. *Expertensystem; Frame; Projekt* 85-04239
- Petkoff, B.: **A cybernetic model of scientific research and cognition.** S. 721-726. *Expertensystem; Kybernetik; Systemforschung; Wissenschaft; Modell; Frame* 85-04240
- Fum, D.; Guida, G.; Tasso, C.: **A rule-based approach to natural language text representation and comprehension.** S. 727-732. *Computerlinguistik; Künstliche Sprache; Textanalyse.* 85-04241
- Faribault, M.; Leon, J.; Meissonnier, V.; Memmi, D.; Zari, G. P.: **From natural language to a canonical representation of the corresponding semantic relationships.** S. 733-738. *Maschinelle Übersetzung; Semantik; Code; Künstliche Sprache.* 85-04242
- Kobsa, A.; Trost, H.: **Representing belief models in semantic networks.** S. 753-757. *Frage-Antwort-System; Mensch-Maschine-Kommunikation; Semantisches Netz; Modell.* 85-04243
- Horacek, H.; Buchberger, E.: **On generation of anaphora and gapping in German.** S. 759-765. *Computerlinguistik; Mensch-Maschine-Kommunikation; Semantisches Netz; Generierung; Methode.* 85-04244
- Cordier, M. O.; Rousset, M. C.: **Propagation: Another way for matching patterns in KBS.** S. 787-791. *Computerlinguistik; Mustererkennung; Algorithmus; Semantisches Netz.* 85-04245
- Ganascia, J. G.: **Explanation facilities for diagnosis systems.** S. 805-810. *Systemforschung; Entwurf; Expertensystem; Medizin.* 85-04246
- Idd, A. L.; Cooper, M. B.: **Man-machine interface issues in the construction and use of an expert system.** *International journal of man-machine studies*, 22(1985)1, S. 91-102. *Expertensystem; Effektivität; Mensch-Maschine-Kommunikation.* 85-04254
- Lehrich, K. P.; Hanne, K. H.; Rigoll, G.: **Maschinelle Sprachverarbeitung. Spracheingabe, Sprachausgabe, Speech-Filling; neue Wege der Mensch-Maschine-Kommunikation.** *Datenverarbeitung. Steuerwirtschaft.* München, 14(1985)3, S. 57-61. *Verarbeitung Gesprochener Sprache; Sprache; Mensch-Maschine-Kommunikation; Entwicklungstendenz* 85-04268
- Branscomb, L. M.: **The computer's debt to science.** *Library Hi Tech, Ann Arbor, MI, US* 2(1984)3, S. 7-18. *Computerindustrie; Informationstechnologie; Wissenschaft und Technik; Speicherung; Programmierung; Mensch-Maschine-Kommunikation.* 85-04276
- Shade, G. A.: **Speech systems for your microcomputer.** Peterborough, NH, US: WG Books (1984) 269 S., ISBN 0-88006-073-5. *Verarbeitung Gesprochener Sprache; Mikrocomputer; Hardware; Software; Marktübersicht; Hersteller.* 85-04304
- Neue Wege der Mensch/Computer-Interaktion: Fortschritte der Software-Ergonomie (Symposium M).** Online. Velbert: Deutscher Telematik-Kongress 2 u. Deutsche Kommunikations-Fachmesse 2. Karlsruhe: 1985. ISBN 3-89077-022-3. *Mensch-Maschine-Kommunikation; Software; Arbeitswissenschaften; Humanfaktor; Künstliche Intelligenz; Expertensystem; Normung; Entwicklungstendenz; Software-Ergonomie.* 85-04321
- Gerald, J.: **Voice response systems technology.** *Library Hi Tech, Ann Arbor*, 2(1984)4, S. 33-38. *Verarbeitung Gesprochener Sprache; Technologie; Sprachausgabe.* 85-04340
- Gooding, F.: **Software for hardware. A control program for electronic speech synthesis.** *ALLC bulletin, Norwich, GB* 13(1985)1, S. 1-4. *Verarbeitung Gesprochener Sprache; Gerät; Mikrocomputer; Kontrolle; Computerprogramm; Sprachsynthese.* 85-04383
- Veiner, E.: **The new Oxford English Dictionary.** *ALLC bulletin, Norwich*, 13(1985)1, S. 8-10. *Wörterbuch; Erstellung; Pflege; Maschinell; Projekt.* 85-04385
- Nicholas, D. E.: **The digital future. User needs.** *Transnational data report, Amsterdam*, 8(1985)3, S. 159-161. *ISDN; Benutzerbedarf; Mensch-Maschine-Kommunikation; Entwicklungstendenz.* 85-04418
- Lein, H. W.: **Sprachsynthese. Systeme zur automatischen Erzeugung von gesprochenem Text.** *Handbuch der modernen Datenverarbeitung*, 22(1985)123, S. 85-91. *Verarbeitung Gesprochener Sprache; Mensch-Maschine-Kommunikation; Sprachsynthese.* 85-04421
- Orkio, H.: **Artificial Intelligence and expert systems research and their possible impact on information science education.** *Education for information*, 3(1985)2, S. 103-114. *Expertensystem; Auswirkung; Ausbildung Bibliothekswesen; Informationsdienst; Bibliotheksdienst.* 85-04469
- Knapp-Lohmann, S.: **Frischer Wind für GOLEM und PASSAT. Siemens PC vereinfachen Zugriff auf Informationen aus GOLEM-Datenbanken.** *COM, München*, 20(1985)2, S. 25-30. *Information Retrieval System; Mikrocomputer; Maschinelles Indexierungsverfahren; SDI; Entwicklungstendenz.* 85-04542
- Limsova, S.: **Multilingual Information and the computer.** *Information development*, 1(1985)2, S. 103-108. *Text; Mehrsprachig; Sprachbarriere; Zeichen; Zeichenvorrat; Transliteration; Kritik; Computerprogramm; Generierung; Pilotanwendung; Tibetisch.* 85-04549
- Vittmann, A.; Schikarski, L.: **Effects of language barrier in searching free-text data bases.** *World Patent information, Oxford*, 7(1985)1/2, S. 145-150. *Patentdokumentation; Information Retrieval; Volltextdatenbank; Sprachbarriere; Deutsch; Anforderung; Projekt; Test; BRD.* 85-04579

Publikationen

- Shepherd, M.: **Fifth-generation computers and information processing.** Canadian journal of information science, Calgary, 9(1984)6, S. 9-19. *Künstliche Intelligenz; Informationstechnologie; Telekommunikationsgesellschaft; Auswirkung; Graphische Darstellung; Entwicklungstendenz.* 85-04589
- Habel, C. (Hrsg.): **Künstliche Intelligenz. Repräsentation von Wissen und natürlichsprachliche Systeme, Frühjahrsschule, Dassel (Solling) März 1984.** Berlin: Springer (1985) 327 S., Informationsfachberichte. *Künstliche Intelligenz; Inferenz; Rechnerunterstütztes Lernen; Semantik; Textanalyse; sensrepräsentation, Verarbeitung natürlicher Sprache.* 85-04612
- O'Leary, M.: **EeasyNet. Doing it all for the end-user.** Online, Weston, CT, US 9(1985)4, S. 106-110. *Dialogdienst; Datenbank; Mensch-Maschine-Kommunikation; Benutzerführung; Benutzer.* 85-04618
- Hoffmann, L.: **On the place of LSP research in applied linguistics.** Fachsprache, Wien, 7(1985)1, S. 2-11. *Terminologieforschung; Linguistik; Wissenschaftstheorie.* 85-04698
- Parton, D.; Huffmann, K.; Pridgen, P.; Norman, K.; Shneiderman, B.: **Learning a menu selection. Training methods compared.** Behaviour and information technology, 4(1985)2, S. 81-91. *Mensch-Maschine-Kommunikation; Benutzerführung; Auswahl; Methode; Vergleich; Empirische Untersuchung.* 85-04715
- Schwarz, H.: **Neue technische Entwicklungen und ihr Einfluss auf die Zukunft der Bibliotheken.** Bibliothek Forschung und Praxis, München, 9(1985)2, S. 127-140. *Informationstechnologie; Telekommunikation; Informationswesen; Bibliothekswesen; Gerät; Dienstleistung; Mikrocomputer; Terminal; Information Retrieval; On-demand-publishing; Elektronische Dokumentlieferung; Künstliche Intelligenz; Auswirkung.* 85-04725
- Eibl, G.: **Current work on expert systems and natural language processing at Siemens.** In: Bernold, Albers, G. (Hrsg.): Artificial intelligence. Towards practical application. Proceedings of the joint technical assessment conference, Zuerich, Switzerland, April 1984. Amsterdam: North-Holland (1985) S. 97-106. *Expertensystem; Projekt; Test; Fakten; Information Retrieval; Textanalyse; Volltextsuche.* 85-04746
- Wiig, K.: **Market trends in artificial intelligence in the United States and Japan.** S. 169-180. *Künstliche Intelligenz; Entwicklungstendenz; Prognose; USA; Japan.* 85-04752
- Huesch, H. W.; Staudt, E.: **The influence of artificial intelligence on organizational structure and rationalization.** S. 189-200. *Künstliche Intelligenz; Rationalisierung; Automation; Industrie; Auswirkung.* 85-04753
- National Research Council, Committee on Human Factors. Washington: **Research needs on the interaction between information systems and their users. Report of a workshop.** Washington, DC, (1984) 43 S. (Vertrieb: NTIS: Springfield, VA, US). *Mensch-Maschine-Kommunikation; Informationssystem; Humanfaktoren; Künstliche Intelligenz; Forschung; Bedarf.* 85-04853
- Hughes, S.: **The communicative power of expert systems.** R and D management, Oxford, GB 15(1985)1, S. 119-124. *Expertensystem; Frage-Antwort-System; Auswirkung; Entscheidungshilfe; Bewertung; Entwicklungstendenz.* 85-04874
- Jacobson, H. E.: **The US and Japan cooperative AI research and development.** Expert systems, Addison-Wesley, 2(1985)3, S. 128-130. *Künstliche Intelligenz; Forschung und Entwicklung; Relation; Zusammenhänge; Entwicklungstendenz; USA; Japan.* 85-04878
- Standera, O. L.: **Electronic publishing. The predicament of occasional users in the editorial process.** Journal of the American Society for Information Science, 36(1985)4, S. 230-240. *Elektronisches Publizieren; Zeitschrift; Kommunikationsprozess; Benutzer; Informationsverhalten; Mensch-Maschine-Kommunikation; Bewertung; Fehler; Projekt; Kanada.* 85-04882
- Wingert, F.: **Morphologic analysis of compound words.** Methods of information in medicine, 24(1985)2, S. 155-162. *Computerlinguistik; Morphologie; Medizin; Begriffssystem; Wörterbucharstellung; Wort; Morphologie.* 85-04899
- Bergeijk, D. van: **Scientific literature in foreign languages and library users.** International Association of Technological Univ. Libraries (IATUL), Goeteborg, (proceedings), 17(1985) S. 187-196. *Schriftliche Sprachbarriere; Bibliothek; Benutzer; Datenbank; Übersetzung.* 85-04914
- Smith, L. C.: **A guide to information sources in artificial intelligence.** Science and technology libraries, New York, 5(1985)3, S. 79-100. *Künstliche Intelligenz; Informationsquelle.* 85-04930
- Faloutsos, C.: **Access methods for text.** ACM computing surveys, 17(1985)1, S. 49-74. *Informationsretrieval; Datenbank; Text; Zugriff; Speicherorganisation* 85-05011
- Jayant, N. S.: **Variable rate speech coding. A review.** In: Dewilde, P.; Nay, C. A. (Hrsg.): Links to the future. Science, systems and services for communications. Proceedings of the International Conference on Communications - ICC '84, Amsterdam, May 1984. Amsterdam: North-Holland (1984) S. 1614-1618. *Datenübertragung; Verarbeitung Gesprochener Sprache; Code; Digital.* 85-05065

- K. D. (Hrsg.): **Studien zur Klassifikation, Systematik und Terminologie. Theorie und Praxis.** In der 6. Arbeitstagung, Münster, September 1984. (Münsteraner Arbeitskreis für Semiotik). Münster: MAK Publ. (1985) 475 S. *Klassifikationsforschung; Terminologieforschung; Linguistik.* 85-05085
- Nedobity, W.: **Die Bedeutung der systematischen Terminologearbeit für den Aufbau von Wissensbanken und anderen Expertensystemen.** S. 37-58. *Terminologearbeit; Expertensystem; Mensch-Maschine-Kommunikation; Künstliche Intelligenz.* 85-05086
- Greiner, G.: **Grundlagen der bibliothekarisch-dokumentarischen Ordnungslehre** S. 73-90. *Ordnungslehre; Clustering.* 85-05088
- Dutz, K. D.: **Klassifikation und Systematik eines wissenschaftlichen Bereichs. Ein Beispiel.** S. 145-190. *Systematischer Katalog; Erstellung; Fachordnung; Klassifikationsforschung; Linguistik.* 85-05091
- Franke, H.: **Die sprachlichen Varietäten des Deutschen Schwierigkeiten einer Klassifikation.** S. 241-254. *Linguistik; Fachordnung; Systematischer Katalog; DK; Klassifikationsforschung; Deutsch.* 85-05093
- Zirkel, U.: **Varietäten der deutschen Gegenwartssprache. Versuch einer Klassifikation.** S. 257-282. *Linguistik; Klassifikation; Wortbildung; Deutsch.* 85-05094
- Merten, K.: **Syntax, Semantik und Pragmatik. Zur Ausdifferenzierung der Morrißschen Semiotik.** S. 385-402. *Semiotik; Syntax; Semantik; Pragmatik; Relation; Modell; Klassifikation; Logik; Kritik; Textanalyse.* 85-05098
- Muller, H. J.: **'Klassifikation' im Studiengang Sprachwissenschaft? Thesen und Beispiel.** S. 437-448. *Linguistik; Klassifikation; Hochschulausbildung.* 85-05099
- F.: **IBM/Unl Tuebingen entwickeln juristisches Expertensystem. Mehr durch weniger. Computerevolutionen Rechtspraxis- und verdrängen gedruckte Texte.** Börsenblatt, Frankfurt. 41(1985)70. 229-2233. *Expertensystem; Jura; Textanalyse; Dialogdienst; Verlagswesen; Auswirkung.* 85-05102
- Staff, J. (Hrsg.): **Proceedings of the third British national conference on databases (BNCOD3).** July 1984. Cambridge: University Press (1984) 271 S. *Datenbank; Datenstruktur; Mensch-Maschine-Kommunikation; Software; Verteiltes System; Benutzerausbildung.* 85-05104
- Johnson R. G.: **Integrating data and metadata to enhance the user interface.** S. 29-39. *Mensch-Maschine-Kommunikation; Information Retrieval System; Semantisches Netz; Verteiltes System; Hochschule.* 85-05107
- Cuff, R. N.: **HERCULES: Database query using natural language fragments.** S. 133-149. *Mensch-Maschine-Kommunikation; Frage-Antwort-System; Datenstruktur; Syntax; Prototyp.* 85-05113
- Martin, N. J.: **The construction of interfaces to triple based databases.** S. 151-172. *Speicherorganisation; Datenstruktur; Modell; Software; Mensch-Maschine-Kommunikation.* 85-05114
- Laender, A. H. F.: **An approach to interactive definition of database views.** S. 173-185. *Mensch-Maschine-Kommunikation; Benutzerfreundlich; Datenstruktur; Prototyp.* 85-05115
- V. K.: **Information systems and problems of upgrading their intelligence level. Automatic annotation and mathematical linguistics.** 18(1984)1, S. 1-24. *Informationssystem; Künstliche Intelligenz; Mensch-Maschine-Kommunikation; Simulation; Expertensystem.* 85-05142
- Osyka, A. F.: **Studies in development of a user-friendly interface.** S. 25-29. *Mensch-Maschine-Kommunikation; Benutzerfreundlich; Sprache; Benutzerführung; Entwicklungstendenz.* 85-05143
- Korolev, E. I.; Korsakova, I. I.; Safronova, M. V.: **The frequency of words in text and their lexical characteristics.** S. 55-66. *Sprachstatistik; Lexikologie; Thesauruserstellung; Wörterbucherstellung; Test.* 85-05146
- Medzhlumyan, A. G.: **Testing the methods of adjustment of documentary systems. With special reference to Ripsme automated IRS.** S. 82-94. *Information Retrieval System; Effektivität; Maschinelle Klassifikation; Mathematische Methode; Test.* 85-05149
- Thomas, J. C.; Schneider, M. L. (Hrsg.): **Human factors in computer systems.** Norwood: Ablex (1984) 271 S. *Systemforschung; Mensch-Maschine-Kommunikation; Humanfaktor; Lernen; Benutzerführung; Ent- 85-05170*
- Carroll, J. M.; Mack, R. L.: **Learning to use a word processor. By doing, by thinking, and by knowing.** S. 13-51. *Mensch-Maschine-Kommunikation; Textverarbeitung; Humanfaktor; Lernen; Benutzerausbildung.* 85-05171
- Reisner, P.: **Formal grammar as a tool for analyzing ease of use. Some fundamental concepts.** S. 53-78. *Systemforschung; Mensch-Maschine-Kommunikation; Kommandosprache; Grammatik; Lernen; Recherchestrategie.* 85-05172
- Savage, R. E.; Habinek, J. K.: **A multilevel menu-driven user interface. Design and evaluation through simulation.** S. 165-186. *Mensch-Maschine-Kommunikation; Benutzerführung; Bildschirm; Entwurf; Informationsverhalten.* 85-05176
- Furnas, G. W.; Landauer, T. K.; Gomez, L. M.; Dumais, S. T.: **Statistical semantics. Analysis of the potential performance of keyword information systems.** S. 187-242. *Mensch-Maschine-Kommunikation; Benennung; Stichwort; Ranking; Statistische Untersuchung.* 85-05177
- Goldsmith, T. E.; Schvaneveldt, R. W.: **Facilitating multiple-cue judgments with integral information displays.** S. 243-270. *Mensch-Maschine-Kommunikation; Entscheidungshilfe; Rechnerunterstützt; Lernen; Test; Humanfaktor; Datenausgabe; Entwurf. Entscheidungssystem* 85-05178

Sonderforschungsbereich 314
Künstliche Intelligenz -
Wissensbasierte Systeme



KI - LABOR
am Lehrstuhl für Informatik IV

Leitung: Prof. Dr. W. Wahlster

Universität des Saarlandes
FR 10.2 Informatik IV
Im Städtwald 15
D-6600 Saarbrücken 11
Fed. Rep. of Germany
Tel. 0681 / 302 - 2363

NEUE BERICHTE ... PAPERS ... MEMOS ...

Einzelne Exemplare der folgenden Berichte und Memos können kostenlos bei den Autoren unter der o.g. Adresse angefordert werden:

Bericht Nr. 3

Wahlster, W.; Kobsa, A.:
Dialog-Based User Models

Erscheint in: Ferrari, G. (Hrsg.): Proceedings of the IEEE, Special Issue on Language Processing, July 1986

Abstract: The paper investigates several approaches to user modeling in natural language dialog systems. First, reasons are pointed out why user modeling has become so important in the last few years, and definitions are proposed for the notions of 'model' and 'user modeling component'. Then, techniques for constructing user models in the course of a dialog are presented and recent proposals for representing a range of assumptions about a user's beliefs and goals in a system's knowledge base are surveyed. Examples for the application of user models in systems developed to cope with presented, and some social implications are discussed. Finally, unsolved problems in coping with collective beliefs or resource-limited processes are investigated, and proposals for application-oriented research are outlined.

Bericht Nr. 4

Fendler, M.; Wichlacz, R.:
SYCON - Ein Rahmensystem zur Constraint-Propagierung auf Netzwerken mit beliebigen symbolischen Constraints

In: Stoyan, H. (Hrsg.): GWAI-85, 9th German Workshop on Artificial Intelligence, Dassel/Solling, September 1985. Heidelberg: Springer, 1986

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 propagiert werden können. Nach einer kurzen Einführung beschreiben wir, wie Restriktionen von Variablen auf Constraint-Netzen propagiert werden. Dabei werden spezielle Algorithmen zur Behandlung von Zyklen und zur Überprüfung der Konsistenz eines Netzwerks vorgestellt. In einem Vergleich mit den Systemen von Steele, Gosling und Freuder zeigen wir die wesentlichen Unterschiede zu SYCON hervor. SYCON ist in Französisch auf einer VAX 11/780 implementiert.

Bericht Nr. 5

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

Gekürzte Version in: LDV-Forum Nr. 2, Dezember 1985, S. 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 Aufforderungen, dem Benutzer Ratschläge bei der Durchführung von Aufgaben erteilen. SC verfügt hierzu sprachliches Wissen, Planungswissen, Wissen über die SINIX-Domäne und über den Benutzer, das in Form separater Wissensquellen gespeichert ist. In vorliegendem Bericht werden nach einer kurzen Einführung über intelligente Hilfesysteme die Systeme UCC und WIZARD vorgestellt, deren Domäne das Betriebssystem UNIX bzw. SINIX 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 des momentanen Projektstatus angeführt.

Bericht Nr. 6

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

Erscheint in: Winter, H. (Hrsg.): Artificial Intelligence and Man-Machine Systems, Heidelberg: Springer, 1986

Abstract: Natural language processing is a prerequisite for advanced knowledge-based systems since the ability to acquire, retrieve, exploit, and present knowledge crucially depends on natural language comprehension and production. Natural language can also guide 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 systems.

Publikationen

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 CraftTM by Carnegie Group Inc., NLMenu by Texas Instruments Inc., and Q & ATM 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

Erscheint in: Proceedings of the COLING-86

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.

Memo Nr. 2

Grasmück, R.; Guldner, A.:

Wissensbasierte Fertigungsplanung in Stanzerien mit FERPLAN: Ein Systemüberblick

Abstract: Unter Fertigungsplanung versteht man den Prozeß des Erstellens von Plänen zur Erstellung von Werkstücken. FERPLAN ist ein wissensbasiertes System, das zur Fertigungsplanung in Stanzerien dienen soll. Ausgehend von einer qualitativen Werkstückbeschreibung, einer Rohmaterialspezifikation und einer Planungsstückzahl wird vom System ein Handlungsplan mit dazugehörigen Vorgabezeiten erstellt.

Nach einem Überblick zur Systemarchitektur wird die Plangenerierung, beginnend mit einer qualitativen Werkstückbeschreibung, an einem Verarbeitungsbeispiel erläutert. Schließlich wird FERPLAN einem tabellarischen Vergleich mit anderen Systemen unterzogen, die sich mit ähnlichen Aufgabenstellungen beschäftigen. Das System ist in FranzLISP unter Verwendung des Produktionssystem-Interpreters OPS5 auf einer VAX 11/780 implementiert.

Memo Nr. 3

Baltes, H.:

GABI - Ein wissensbasiertes Geldanlageberatungsprogramm

Abstract: Dieses Papier stellt das Expertensystem GABI vor (Geldanlage-Beratungs- und Informationssystem), das aus kundenspezifischen Angaben, die das System im Dialog ermittelt, für den Kunden geeignete Anlagevorschläge ableitet. Nach einer kurzen Beschreibung des Leistungsumfanges, einer Darstellung der Entwurfskriterien und Einsatzmöglichkeiten des Systems, wird die Systemarchitektur erläutert, die neben einer Dialog-, einer einfachen Erklärungs- und einer Inferenzkomponente eine stark partitionierte Wissensbasis umfaßt. Der Aufbau der Wissensbasis und die Repräsentation des notwendigen Konditionen-, Welt- und Regelwissens wird anhand von einigen Beispielen erläutert. Das System ist in FranzLISP, Version 38.79, und der eingebetteten PLANNER-artigen KI-Programmiersprache PEARL auf einer VAX 11/780 unter dem Betriebssystem UNIX 4.2 BSD implementiert. Ein Hauptmerkmal des Systems besteht darin, daß sowohl das vom System benötigte Wissen als auch die Inferenzregeln in Frameartigen Strukturen kodiert sind.

Memo Nr. 4

Schmauks, D.:

Formulardeixis und ihre Simulation auf dem Bildschirm. Ein Überblick aus linguistischer Sicht

Abstract: Bis Ende der 70er Jahre wurde lokale Deixis eher in sehr allgemeinem Rahmen behandelt, ferner vermißt man auch heute noch in den meisten linguistischen Arbeiten eine eingehendere Behandlung von Zeigegesten. Der vorliegende Artikel hingegen beschäftigt sich ausschließlich mit deiktischen Operationen auf ein Formular, also einen sehr begrenzten Zeigerumraum. In diesem Diskursbereich wird auf Gegenstände relativ häufig und effektiv mit Zeigegesten referiert. Formulardeixis bietet aus linguistischer

Sicht den Vorteil, daß einige Teilprobleme der lokalen Deixis reduziert sind, eine unnatürliche Situation konstruiert werden muß. Im ersten Kapitel werden Termini definiert und die verschiedenen sprachlichen und außersprachlichen Deixis vorgestellt. Die Kapitel 2-4 listen die wesentlichsten Besonderheiten der Formalardeixis auf. Der Schwerpunkt liegt jeweils auf der Frage, inwieweit Mittel der Deixis leichter beschreibbar geworden sind. Kapitel 5 beleuchtet das Thema aus der Perspektive der Künstlichen Intelligenz. Motivation hierfür ist, daß mittlerweile die außersprachliche Komponente der Formalardeixis auf einem Bildschirm darstellen kann. An der Universität Saarbrücken wird derzeit das natürlichsprachliche System XTRA entwickelt, das dem Benutzer die Möglichkeit bietet, auf Teilbeurteilungen des Formulars nicht nur sprachlich, sondern auch mittels Zeigegesten zu referieren.

Memo Nr. 5

Andre, E.; Bosch, G.; Herzog, G.; Rist, T.:

Characterizing Trajectories of Moving Objects Using Natural Language Partitions

Abstract: The topic of this paper is the analysis of the semantics of the partitioning relations "along" and "past", which are used to characterize the path of moving objects. The German dialog system CITYTOUR, which answers questions about the trajectories of objects in a scene, is presented by means of a simple example. The representational prerequisites of the static and dynamic objects, required for natural analysis are presented. Two concrete predicate functions testing whether an object moves along or moves past a static object are described in detail.

NEUE BERICHTE ... PAPERS ... MEMOS

TOPIC-12/84

Hahn, U. / R. Hammwoehner / R. Kuhlen / U. Reimer

TOPIC II / TOPOGRAPHIC II.

Automatische Textkondensierung und text-orientiertes

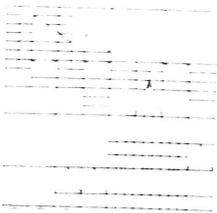
Informationsmanagement.

Projektziele - State-of-the-Art.

A concise overview is given of the aims of the present phase of the text condensation (automatic abstracting) and information retrieval projects at the Information Systems Department at the University of Constance (FRG). The report provides a state-of-the-art report of current research in text processing, statistical, linguistic, knowledge-based and cognitive approaches to full text analysis and retrieval in information systems. The conceptual design of a knowledge-based full text information system is outlined. Combining both, sophisticated text analysis as well as information retrieval procedures, text analysis is achieved with applying a word expert system to text coherence and text coherence phenomena for text parsing, a frame-based representation model for text representation, and appropriate text condensation mechanisms for text summarization (project TOPIC II). Text information management on the other hand is based on various forms of graphical man-machine interaction and alternative ways of graphics-based text information retrieval (project TOPOGRAPHIC II).

(= Bericht TOPIC-12/84 & TOPOGRAPHIC-3/84, in German, 1984, 72pp.)

Universität Konstanz
Informationswissenschaft



Projekt TOPIC II*
Postfach 5560
D-7750 Konstanz 1

*Bestellungen von Non-Profit-Organisationen (Universitäten, öffentlichen Stellen usw.) werden kostenfrei abgewickelt, ansonsten pro Bericht ein Unkostenbeitrag von DM 5,- erhoben.

Publikationen

TOPIC-14/85

Hahn, Udo / Ulrich Reimer

Entwurfsprinzipien und Architektur des Textkondensierungssystems TOPIC

Based on a motivation for full text processing in information systems from an information science point of view the major methodological concepts underlying the automatic text condensation (automatic abstracting) system TOPIC are outlined: partial parsing as a text-specific parsing strategy functionally adequate to indicative abstracting, integrity of the knowledge base under the scope of a formally specified frame data model, recognition of text cohesion and coherence phenomena of expository texts by an appropriately tuned text parser, and variable degrees of thematic abstraction of the condensate (summary) derived from the original text. Subsequently, a structured description of the whole TOPIC system is given which distinguishes three major processes: text parsing based on the close interaction between a frame knowledge base and a word expert parser, text condensation on the basis of dominant concept patterns of the text representation structures, and finally, the construction of the text graph which represents the knowledge structures of text condensates on different levels of thematic abstraction which may be accessed in a flexible manner in the course of an interactive graphical retrieval dialogue.

(= Bericht TOPIC-14/85, in German, April 1985, 17pp.)

TOPIC-17/85

Hahn, Udo / Ulrich Reimer

The TOPIC Project: Text-Oriented Procedures for Information Management and Condensation of Expository Texts. Final Report.

After considering requirements for advanced full text information systems the major methodological concepts underlying the automatic text condensation (automatic abstracting) system TOPIC are outlined: partial parsing as a text-specific parsing strategy, validity and reliability of the knowledge base under the scope of a formally specified frame data model, recognition of text cohesion and coherence phenomena of expository texts by an appropriately tuned text parser, and variable degrees of thematic aggregation of an input text, basically on a level of text summarization comparable to that of structured abstracts (indicative abstracts plus facts acquired from the original text). Subsequently, a coherent formal description of the whole TOPIC system is given which distinguishes three major processes: text parsing based on the close interaction between a frame knowledge base and a word expert parser, text condensation which utilizes various activation, property assignment, and connectivity patterns of the text representation structures, and finally, the construction of the text graph which represents the knowledge structures of text condensates on different levels of thematic abstraction accessible to appropriate information retrieval procedures.

(= Bericht TOPIC-17/85, in English, May 1985, 114pp.)