# LDV - FORUM

Gesellschaft für Linguistische Datenverarbeitung

GLDV

INHALT:

Nr. 2 August 1983

ISSN 0172-992

Editorial ..... 3

Brigitte Endres-Niggemeyer:
LDV-Förderung - Förderung der Innovation ..... 4

Manfred Thiel: Maschinelle Übersetzung und
Perspektiven ihrer Anwendung ..... 10

Arbeitskreis maschinelle Übersetzung ..... 23

Berichte

Informationswissenschaftliches Kolloquium in Konstanz..... 24

Jahrestagung der LDV-Fittings 1983 in Trier..... 28

Projekte

Japanisch/deutsche Titelübersetzung: SEMSYN ..... 31

Automatisches Textkondensierungssystem TOPIC ..... 33

Publikationen ..... 34

Veranstaltungen ..... 38

Presseschau ..... 44

Mitteilungen aus der GLDV ..... 50

**Redaktion**

Dietrich Fischer
Günter Schulz
Lothar Rostek

GID-SfK

Herriotstr. 5

D-6000 Frankfurt/Main 71

**Redaktionsschluß: 15.11.1983**

# Editorial

Und es erscheint doch noch ...

In der Zeit der langen Informationspause hat sich manch einer - und auch uns - gefragt, ob denn das 'LDV-Fittings INFO' mit dem Wechsel des Vereinsnamens aufgehört hat zu existieren. Doch nach einer kleinen Metamorphose können wir es endlich als 'LDV-FORUM' auf den Weg schicken. Wir hoffen, daß wir diesmal so viel dazu gelernt haben, daß wir demnächst das Erscheinungsdatum näher an den den offiziell angekündigten Redaktionsschluß heranrücken können. Zumindest haben wir erfahren, welch gute Arbeit unsere Vorgängerin Brigitte Endres-Niggemeyer in ihrer langjährigen Aufgabe als Redakteurin der INFO's jedesmal leisten mußte.
In diesem Heft kommt sie nun als Autorin zu Wort, und zwar mit ihrem auf der Jahrestagung in Trier gehaltenen Vortrag zur Situation der Förderpolitik. Ein weiterer Vortrag dieser Tagung steht im Mittelpunkt dieser Ausgabe, der Aufsatz von Manfred Thiel zur maschinellen Übersetzung.
Wir glauben, daß auch die weiteren Beiträge in den anderen Rubriken nützliche Informationen für Sie bereithalten.
Damit auch die folgenden Hefte interessante Beiträge enthalten, sind wir auf Ihre Einsendungen - möglichst vor Redaktionsschluß - angewiesen. Wir warten nun auf zahlreiche Aufsätze, Kritiken, Anregungen, Informationen, Thesen, Diskussionsbeiträge ... zur LDV allgemein und zum FORUM speziell, so daß diese Zeitschrift nicht nur 'LDV-FORUM' heißt, sondern auch zu einem Forum der LDV wird.

## Dank an Frau Endres-Niggemeyer

Die frühere LDV-Fittings INFO trägt heute erstmals den Namen 'LDV-Forum', nachdem die jetzige Redaktion die Nr. 1 1983 nochmals unter altem Namen hatte erscheinen lassen. Damit geht ein Abschnitt der GLDV zu Ende, der eng mit der Arbeit und dem Engagement von Frau Endres-Niggemeyer verbunden ist. Sie hatte die seit 1978 als meist zweiseitiges Informationsblatt erscheinende LDV-Fittings INFO Ende 1979 übernommen und sowohl vom Umfang her stark erweitert als auch vom Inhalt her deutlich verbessert. Sie mußte das Informationsblatt Ende 1982 abgeben, weil sie als Professorin an die Fachhochschule Hannover berufen wurde.
Neben der Arbeit an der LDV-Fittings INFO war Frau Endres-Niggemeyer von November 1978 bis Dezember 1980 als Schriftführerin Mitglied des Vorstands.

Wir danken Frau Endres-Niggemeyer - stellvertretend für die Mitglieder der GLDV - für Ihre engagierte und fruchtbare Arbeit im Verein, dessen Erscheinungsbild sie in den letzten Jahren deutlich mitgeprägt hat. Wir sind froh, daß sie der GLDV als Mitglied des Beirats weiterhin mit Rat und Tat zur Verfügung stehen wird.

## Wechsel im Vorstand

Am 3. März 1983 schied Herr Arno Schmidt aus dem Vorstand der GLDV aus, der er in den letzten beiden Jahren als Schriftführer angehört hatte. Beirat und Vorstand danken ihm herzlich für die von ihm geleistete Arbeit.

Brigitte Endres-Niggemeyer

Förderung von Linguistischer Datenverarbeitung -
Förderung von Innovation

Die Situation in der die Linguistische Datenverarbeitung irgendwie betreffenden Förderung von Forschung und Entwicklung ist - soweit ich sehe - nicht so sehr von offenstehenden Fördermöglichkeiten bestimmt; den Akzent setzen im Moment anlaufende Förderaktivitäten. Beispiele dafür sind:

- EUROTRA (am 4.11.82 von der EG beschlossen)
- ESPRIT (am 21.2.83 von der EG beschlossen)
- Fördermaßnahmen zur japanisch-deutschen computerunterstützten Übersetzung, wo die Verhandlungen schon so weit fortgeschritten sind, daß in absehbarer Zeit mit dem Beginn der Entwicklungsarbeiten zu rechnen ist.

Dieser Vorrang der Taten vor Möglichkeiten und Programmen erscheint mir kennzeichnend für eine Veränderung, die ich im Laufe der letzten zwei Jahre in der Förderlandschaft beobachtet und zum Teil auch mitgemacht habe: Nach einer Zeit der Ratlosigkeit und des Zweifelns angesichts der sich abzeichnenden Problemlage von Rezession, Arbeitslosigkeit, Strukturverschiebungen in der Wirtschaft, leeren öffentlichen Kassen und rapider technischer Weiterentwicklung wird nun die überdeutliche Problemsituation als Aufforderung zu politischem Handeln genommen. Man wartet nicht mehr zweifelnd ab, sondern man handelt mit den Mitteln, die man zur Verfügung zu haben glaubt. In der Forschungsförderung könnte das Motto heißen:

"Wir sind technologisch im Rückstand - das muß sich wieder ändern."

und

"Was gut ist für die Konkurrenzfähigkeit der Wirtschaft, ist gut für alle."

oder noch weiter vergröbert

"Wir haben geschlafen, die Japaner nicht. Nichts wie hinterher!"

Vor dem offensichtlichen Handlungsbedarf treten Zweifel beispielsweise am Sinn einer forcierten Automation in allen möglichen Lebensbereichen zurück. Sie werden in die Programme zur Beschleunigung der technologischen Weiterentwicklung etwa unter dem Stichwort "Erforschung der Humanfaktoren" integriert - so etwa bei ESPRIT. Auf der ganzen Linie wird entschlossen Technologieförderung gemacht, ob im BMFT oder bei der EG.

In dieser Situation verwischt sich zumindest in der anwendungsorientierten Förderung der Unterschied zwischen Wirtschaftsförderung und Forschungsförderung immer mehr. Soweit die LDV betroffen ist, läßt sich die Denkrichtung etwa so beschreiben:

Wirtschaftliche Probleme erfordern die Förderung der Wirtschaft. Ein Teil der Wirtschaftsförderung ist Technologieförderung. Ein Teil davon ist die Förderung von Forschung und Entwicklung, z.B. durch die Förderung der LDV.

Förderung nähert sich immer mehr der Vergabe von Forschungsaufträgen. Leistungsanforderungen aus dem gesellschafts- und besonders wirtschaftspolitischen Raum werden durch die Förderinstitutionen an die Wissenschaft weitergegeben und durch Förderung bezahlt. In wirtschaftlich schwierigen Zeiten versucht man stärker als bisher, die Ressource Wissenschaft gezielt zu nutzen. So ist ESPRIT Teil einer wirtschafts- und technologiepolitischen Strategie der EG. Die Leistungsanforderungen an die LDV gelten in erster Linie

- der computerunterstützten oder automatischen Übersetzung
- der fortgeschrittenen Sprach- und Textverarbeitung vor allem im Büro.

Erwartet werden Beiträge zur technischen Innovation.

Eine weitere Veränderung, die ich in der Förderpraxis der letzten beiden Jahre beobachten konnte, betrifft die Zeitvorstellungen bei der Projektförderung. Wenn es nicht eigene Überlegungen waren, dann hat spätestens die Bekanntschaft mit dem japanischen Förderprogramm "Computer der 5. Generation" die Bereitschaft zu mittel- und langfristigen Förderstrategien deutlich verstärkt.

ESPRIT - Europäisches Strategieprogramm für Forschung und Entwicklung auf dem Gebiet der Informationstechnologien

Das aktuellste Beispiel für eine langfristige, großangelegte und LDV-relevante Förderstrategie ist ESPRIT. ESPRIT ist auf einen Förderzeitraum von 10 Jahren angelegt. Die Förderung für 1983/84 im Umfang von 11,5 Millionen Rechnungseinheiten wurde am 21.2.83 beschlossen. Momentan werden Projektvorschläge für die 16 fünfjährigen Pilotprojekte geprüft. Man rechnet damit, daß ESPRIT Mitte 1983 anlaufen kann. Das Programm orientiert sich in Größenordnung und Inhalt an amerikanischen, noch mehr aber an japanischen Vorbildern, besonders am schon genannten "Computer der 5. Generation". ESPRIT ist nicht gerade ein bescheidenes Programm:
"ESPRIT wird sich aus einer Reihe langfristiger FuE-Programme zusammensetzen, die so grundlegend und umfassend sind, daß sich eine Gesamtkapazität ergibt, die von der gesamten europäischen Industrie ausgenutzt werden kann, um einen Wettbewerb mit den amerikanischen und japanischen Bemühungen auf diesem Gebiet erfolgreich bestehen zu können."
Als industrieorientiertes Programm soll ESPRIT die europäische Informationstechnologie auf dem Binnenmarkt und auf dem Weltmarkt konkurrenzfähig machen. Die einzelnen Projekte werden von der EG und den beteiligten Unternehmen im Verhältnis 1:1 finanziert. Konzipiert worden ist das Programm durch Expertengruppen mit starker Beteiligung der Industrie. Bedingung für die Förderung ist zumindest bei den Pilotprojekten die internationale Zusammenarbeit von Partnern aus der EG, wobei zumindest ein Partner ein Industrieunternehmen sein muß. ESPRIT ist im japanischen Sinn "precompetitive". Im Förderprogramm sollen die technischen Grundlagen entwickelt werden, die die einzelnen Unternehmen dann zur Entwicklung marktreifer Produkte benutzen können. Die Grundlagen werden gemeinsam mit öffentlicher Förderung erarbeitet, die Produkte in Konkurrenz. In einer ersten Phase von ESPRIT soll durch die von den vorbereitenden Expertengruppen vorgeschlagenen Pilotprojekte das europäische Forschungspotential in der Informationsverarbeitung aktiviert werden. Anschließend soll nach

etwa einjährigem Vorlauf das "Hauptprogramm" beginnen.
Forschungsschwerpunkte von ESPRIT sind:
- Fortgeschrittene Mikroelektronik
- Softwaretechnologie
- Fortgeschrittene Informationsverarbeitung
- Büroautomation
- Computer integrated manufacturing.

Für die LDV sind die Punkte
- Fortgeschrittene Informationsverarbeitung und
- Büroautomation

die einschlägigsten.

Fortgeschrittene Informationsverarbeitung

Fortgeschrittene Informationsverarbeitung unterscheidet sich von der bisherigen durch eine entwickeltere Intelligenz. Die Systeme nehmen Informationen durch Sensoren aktiv auf, verarbeiten sie unter Benutzung des bereits gespeicherten Wissens. Sie erzeugen neues Wissen durch Anwendung von Inferenzregeln. Es handelt sich bei ihnen um KI-Systeme, besonders um Expertensysteme. Die Anwendung von Techniken der Künstlichen Intelligenz wird forciert, weil eine verbesserte Technologie dies möglich erscheinen läßt und weil das Vorpreschen des japanischen Computers der 5. Generation es geraten erscheinen läßt.
Forschungsthemen im Programmpunkt "Fortgeschrittene Informationsverarbeitung" sind:
- Informations- und Wissensengineering
- Signalverarbeitung und externe Schnittstellen
- Speicherung und Gebrauch von Information und Wissen
- Rechnerarchitektur
- Entwurfsmethoden und -ziele.

Wissensbasen im Sinn von ESPRIT unterscheiden sich von herkömmlichen Informationsbasen dadurch, daß sie Urteile über den Wert der in ihnen gespeicherten Informationen enthalten und Interpretationsregeln mit den zugehörigen Metaregeln zur Steuerung der Regelanwendung. Bis aus den Datenbanksystemen

von heute Wissensbanksysteme entwickelt sind, ist nach Einschätzung der ESPRIT-Experten noch viel FuE-Arbeit zu leisten. Die externen Schnittstellen sollen auch gesprochen-sprachliche und mixed-mode-Schnittstellen sein. Einzelforschungsthemen sind hier z.B.:

- Erkennung von fließend gesprochenen Wörtern aus großen Wortschätzen (kurzfristig)
- Ideale Sprachsynthese auf der Basis von Text (mittelfristig)
- Sprachliche Beschreibung visuell beobachteter Szenen (langfristig).

Bei der Entwicklung von Expertensystemen für den industriellen Einsatz sind vorrangig:

- die Entwicklung geeigneter Repräsentationssprachen
- die Verbesserung der Methoden zur Datengewinnung, auch aus gesprochener Sprache, mit Ableitung von Wissen aus Daten und geeigneten Schnittstellen für die Zusammenarbeit von Mensch und Maschine
- die Entwicklung experimenteller Expertensysteme.

Auch hier rechnet man mit großem Entwicklungsaufwand.

Als ein Hauptanwendungsgebiet der fortgeschrittenen Informationsverarbeitung nennt ESPRIT die Büroautomation, die auch ein eigener Forschungsschwerpunkt des Programmms ist. Hier sind die Teilthemen:

- Büro-Workstations
- Bürokommunikationssysteme für mixed-mode-Kommunikation
- Büroordnungs- und Retrievalsysteme für Wissen
- Interaktion zwischen Mensch und System, Humanfaktoren
- Analyse von Bürotätigkeiten mit Ermessensspielraum.

Soweit ESPRIT. In seinen LDV-bezogenen Teilen dürften die Arbeitsziele von ESPRIT eigentlich keinen LDV-Forscher kalt lassen, besonders nicht, wenn er ein gewisses Interesse an Künstlicher Intelligenz hat.

EUROTRA

Über EUROTRA berichtet ausführlich S. Perschke. Darum beschränke ich mich hier auf die fördertechnischen Aspekte des Vorhabens. EUROTRA umfaßt drei Phasen. Die erste wird vom EUROTRA-Team bei der EG abgewickelt. Sie dauert zwei Jahre.

In der zweiten Phase beginnen die linguistischen Arbeiten an den sieben EG-Amtssprachen jeweils in den zuständigen EG-Ländern. Sie werden in nationalen Zentren durchgeführt, die von EG und Sitzland im Verhältnis 40:60 finanziert werden. Ein solches nationales Zentrum wird aller Voraussicht nach etwa ab 1985 auch in der Bundesrepublik existieren. Deutscherseits wird das BMFT die Finanzierung übernehmen, die GID wird Projektträger sein.
Auch auf die Computerlinguisten, die sich mit automatischer Übersetzung befassen, kommen also erhebliche Leistungsanforderungen zu.

BMFT-Förderung der LDV nach dem Leistungsplan Fachinformation

Im relativ bescheidenen Rahmen der BMFT-Förderung nach dem Leistungsplan Fachinformation gilt das gleiche für Beiträge zur Verbesserung oder Er chung von japanisch-deutscher Übersetzung.
Daneben bestehen natürli uch Fördermöglichkeiten zu allen Themen des Leistungsplan f individuellen Antrag. Aber hier muß die Initiative von de nigen kommen, der gefördert werden will.

Manfred Thiel

# Maschinelle Übersetzung und Perspektiven ihrer Anwendung:

## 1. Maschinelle Übersetzung

### 1.1. Einordnung der maschinelle[illegible]bersetzung

Bevor ich über die möglichen [illegible]rspektiven vor allem des Saarbrücker Übersetzungssystem[illegible]ichte, möchte ich einige grundlegende Bemerkungen zur m[illegible]inellen Übersetzung überhaupt und insbesondere über d[illegible]tifunktionalen Charakter einzelner Systemkomponenten ma[illegible]

Was ist 'Übersetzung' eigentlic[illegible]nn man definiert: " Übersetzung ist die Übertragung eine[illegible]tes aus der Sprache L(q) in die Sprache L(z) unter Erhalt[illegible]er semantischen Information", so ist dies eine recht [illegible]fache Sicht und berücksichtigt bei weitem nicht alle [illegible]tablen Fakten. Aber sie mag für die folgenden Überlegung[illegible]s Basis dienen. Über die Zielsprache L(z) wird [illegible]ielsweise nichts weiter ausgesagt. Es ist in erster L[illegible] an natürliche Sprache gedacht. Gleichwohl kommt aber [illegible] auch formale [illegible]rache in Betracht, beispielsweise w[illegible]m natürlichsp[illegible]achlichen Programmieren aus der natürli[illegible]Sprache in eine computerverständliche, formale Sprach[illegible]rsetzt wird. D[illegible]ben kommt als Zielsprache ebenso eine In[illegible]ungssprache in [illegible]rage oder eine evtl. baumartige Darstell[illegible]nes Komprimie[illegible] Textes, wie sie im Projekt TOPIC von Kuhlen in de[illegible]iskussion ist. Zwischen diesen Bereiche[illegible]te ich kei[illegible]nzipielle Trennung aufbauen. Es sollte [illegible]ch für ma[illegible]lle Übersetzung keine [illegible]ierende St[illegible] gegenüber [illegible]ren Disziplinen aufgebaut we[illegible]. Worau[illegible]r ankommt, ist: Wenn in diesen Disziplinen schon Parallelen festgestellt werden können, sollten sie so weitgehend wie möglich ausgenutzt werden.

Ein Verweis auf ein LDV-FITTINGS INFO in 'Das Inforum' Nr. 12, 1982, macht diese Vorstellung ebenfalls deutlich: Auf der Ebene von Subsystemen kommen Moduln vor, die in verschiedener Kombination zu anwendungsorientierten Gesamtsystemen unter Beachtung des Schnittstellenproblems zusammengeschlossen werden können. In dem erwähnten INFO "Computerlinguistische In-

haltsanalyse sucht nach linguistischen Lösungen" ist zu lesen: "Von daher würden wir Hinweise, bzw. eine Bestandsaufnahme portabler Programme, die als Moduln in die bestehenden Inhaltsanalysesysteme übernommen werden können, begrüßen." Hierbei ist an Lemmatisierung und an die Erkennung von einfachen syntak-tischen Informationen wie Kasus und Passiv- bzw. Aktiv-VP gedacht.
Gerade die Lemmatisierung ist beispielsweise ein Analysebaustein, der multifunktional eingesetzt werden kann, da er auch notwendiger Baustein einer maschinellen Übersetzung ist. Es soll allerdings auch auf Probleme hingewiesen werden:
1. In welchem Maße düerfen Moduln miniaturisiert werden? Die Erkennung morphosyntaktischer Kasus allein ist als eigenständiger Modul wohl nicht gerechtfertigt. Daher soll im folgenden von größeren Subsystemen die Rede sein.
2. Sind die Ergebnisse eines Subsystems, das für eine Anwendung A konzipiert ist, auch für eine Anwendung B akzeptabel? Mit anderen Worten: Inwieweit geht der Verwendungszweck in die Konzeption z.B. einer morphologischen Analyse ein?

## 2. Perspektiven der maschinellen Übersetzung

Mitte 1982 wurde von Prof. Kuhlen eine Delphi-Befragung begonnen, deren vorläufigen Ergebnisse ich für den Bereich maschinelle Übersetzung kurz vorstellen möchte. Herr Kuhlen hat über die ersten Resultate der Umfrage auf dem SFB-Kolloquium im Oktober 1982 in Saarbrücken berichtet.

Bzgl. der Wünschbarkeit nimmt die maschinelle Übersetzung hinter Retrieval und Question Answering und vor Indexing und Abstracting einen mittleren Platz ein. Innerhalb des Themas maschinelle Übersetzung ergibt sich für die Zusammenfassung der Bewertungen 'sehr wichtig' und 'wichtig' folgende Rangfolge von Teilbereichen:

1. Automatic Translation of restricted input (2.)
2. Interactive Translation Systems (1.)
   Development of sophisticated Multilingual lexicons (6.)
4. Integration of information and translation systems (3.)

5. Multilingual translation systems (4.)
6. Fully automatic translation systems (5.)

Die Werte links gelten für die Forschung, für die Anwendung gelten die Werte rechts in Klammern. Abgesehen von dem Ausreißer 'Multilingual Lexicons' sind sich beide Reihenfolgen ähnlich. Es fällt auf, daß auch bei der Bewertung der Forschung Aspekte im Vordergrund stehen, die auf eine relativ rasche Anwendbarkeit der maschinellen Übersetzung hinauslaufen. Dies kann dadurch begründet werden, daß spezielle softwaretechnologische oder linguistische Probleme, die in der maschinellen Übersetzung eine Rolle spielen, an anderer Stelle im Fragebogen behandelt werden.

Bei der Bewertung der 'Machbarkeit' liegen die Angaben für die Anwendung durchweg etwas niedriger als für die Forschung. Ähnliches ergibt sich bei den Angaben zur Realisation, wo die Anwendbarkeit gegenüber der Forschung um jeweils 1 bis 2 Zeitabschnitte in die Zukunft hinein verschoben ist. Die benutzbaren Zeitabschnitte sind: 1984, 1989 1996, 2010 und unrealistisch.

Weiteren, bislang nicht veröffentlichten Ergebnissen der 1. Runde soll nicht vorgegriffen werden. Aber schon auf der Grundlage der vorliegenden Daten wird die Erwartung bestätigt, daß in näherer Zukunft nicht mit einer high quality - Übersetzung zu rechnen ist, sondern mit interaktiven Systemen oder mit Übersetzungssystemen für eingeschränkten Input. Bzgl. interaktiver Systeme bin ich etwas skeptisch, da meiner Meinung nach solche Systeme u.a. zu langsam und vielleicht zu umständlich handhabbar sind. Die Vorstellung von interaktiven Systemen ist eher auf 'maschinengestützte Humanübersetzung' zu reduzieren.

Für eine Weiterentwicklung der maschinellen Übersetzung sind vor allem zwei Wege zu beachten:

1. Weiterentwicklung der linguistischen Theorien und der SoftwareKomponenten.
2. Aufbau neuer Systeme auf bestehenden.

Die beiden Ansätze werden im folgenden am Beispiel des Saarbrücker Übersetzungssystems skizziert. Problematik und Zukunftsperspektiven der angesprochenen Punkte dürften dabei unabhängig von diesem konkreten System sein.

## 2.1. Das Saarbrücker Übersetzungssystem

Das Teilprojekt A2 im SFB 100 entwickelt nun schon im 12. Jahr das Saarbrücker Analysesystem SATAN und das Übersetzungssystem SUSY. Für derart umfangreiche Forschungsvorhaben sind entsprechend große und noch größere Zeiträume notwendig, um einen Ansatz zu einem gewissen Abschluß zu bringen und nicht immer wieder in modellhaften Entwürfen stecken zu bleiben. Auf der anderen Seite werden aber derart langfristige Vorhaben oft von neueren Entwicklungen überholt. Diesem Dilemma kann durch einen modularen Systemaufbau und durch eine möglichst weitgehende Offenheit für die Integration neuer Methoden begegnet werden. Hier liegt auch die Begründung dafür, daß zu Beginn von Subsystemen gesprochen wurde.

Die Analyse in SUSY benutzt kein durchgehendes Analyseverfahren wie z. B. einen ATN. Statt dessen wird versucht, verschiedene Parsingstrategien flexibel an linguistische Problemstellungen anzupassen. Der Analyse liegt das Konzept zugrunde: Zuerst die Möglichkeiten der morphosyntaktischen Analyse soweit wie möglich ausschöpfen und Semantik nur dann zur Hilfe zu nehmen, wenn dies von Vorteil ist. In der syntaktischen Analyse sind Grammatik und Programmm größtenteils nicht in der üblichen Weise voneinander getrennt. Die semantische Analyse dagegen benutzt einen Interpreter. Die Regeln, die der Algorithmus interpretiert, sind prozeduraler Natur, mit allen Vor- und Nachteilen. Es können in dieser Sprache selbst wieder Prozeduren formuliert werden und mit Parameterübergabe aufgerufen werden.

Das System kann im Ansatz als multilingual bezeichnet werden, da mehrere Sprachen integriert sind, die alle vom gleichen Analyseprogramm behandelt werden. Die weitaus meisten Regeln der Grammatik (bzw. Teilgrammatiken) gelten für alle Sprachen. Dort, wo einzelsprachenspezifische Besonderheiten vorliegen, werden die betreffenden Regeln sprachabhängig ausgewählt. Die Grammatik des Analysesystems umfaßt also so etwas wie die Vereinigungsmenge der Regeln derjenigen Grammatiken, die im System enthalten sind.

Da es nicht möglich ist, das gesamte System zu beschreiben, soll auf einige Punkte eingegangen werden, die von allgemeinem Interesse zu sein scheinen. Darüber hinaus können natür-

lich auch nicht alle Problemfelder, die in der zukünftigen Entwicklung der maschinellen Übersetzung eine Rolle spielen können, angesprochen werden. Offen bleiben beispielsweise Fragen nach deterministischem/ nichtdeterministischem Parsing oder die Stellung des Transfers.

### 2.1.1. Lexika

Da das System SUSY stark lexikonorientiert ist, spielen deren Umfang und ihre Qualität eine besondere Rolle. Die Stellung der Lexika in der maschinellen Übersetzung wird in Zukunft verstärkt diskutiert werden müssen, da umfangreichere Wörterbücher nicht nur quanitative Veränderungen mit sich bringen, sondern auch eine qualitative Verschiebung von Problemen. So bringt ein umfangreiches Lexikon durch eine erhöhte Lesartenproduktion in der morphologischen Analyse eine Verkomplizierung der Homographenreduktion bzw. der gesamten Analyse mit sich.

Folgende Lexika mit Größenangaben liegen im System SUSY vor:

Morphosyntaktische Analyselexika:

| | | |
|---|---|---|
| Deutsch | 137.000 | Einträge (Stämme) |
| Englisch | 8.200 | " |
| Französisch | 9.400 | " |
| Russisch | 15.000 | " |

Semantische Wörterbücher:

| | | |
|---|---|---|
| Deutsch | 76.000 | Regel- und Merkmaleinträge |
| Englisch | 3.100 | " |
| Russisch | 2.600 | " |
| Französisch | 700 | " |

Tansferlexika:

| | | |
|---|---|---|
| Deutsch - Englisch | 2.200 | Einträge |
| Deutsch - Franz. | 5.500 | " |
| Englisch - Deutsch | 7.800 | " |
| Englisch - Franz. | 4.200 | " |
| Französisch - Deutsch | 9.000 | " |
| Russisch - Deutsch | 9.000 | " |

Morphosyntaktische Synthesewörterbücher:

| | | |
|---|---|---|
| Deutsch z.Zt. | 11.000 | Stämme (demnächst ca. 80.000) |
| Französisch | 8.500 | " |

Englisch 2.000 "

## 2.1.2. Texttyp

Bei der Beschäftigung mit 'real life' - Texten stellte sich eine starke Abhängigkeit der Analyse (genauer: der darin enthaltenen Grammatik) von einzelnen Texttypen heraus, die sich während eines Textes auch ändern können. So tauchen in Arbeitsanleitungen etwa folgende Sätze auf:

Abschneiden Helfer.

Schrägen schneiden Helfer.

Die Notwendigkeit der besonderen Behandlung solcher Texttypen ist offensichtlich: "Schrägen schneiden Helfer" ist für eine Grammatik für "normale" Texte nicht akzeptabel. Mit anderen Worten: Würde man auf texttypabhängige Grammatiken verzichten, ist nicht auszuschließen, daß sich Regeln einander widersprechen. Das Konzept der Texttypen lehnt sich an die Vorstellung von 'sublanguage' an, vgl. KITTREDGE/LEHRBERGER (1982), dort vor allem SAGER und LEHRBERGER. Es spielt hier teilweise die Problematik der Multilingualität hinein, da kein Grund dafür vorzuliegen scheint, zu unterscheiden zwischen dem einzelsprachenabhängigen Anstoßen einer Regel innerhalb des Konzepts der Multilingualität und dem Anstoßen einer texttypabhängigen Regel. Mehr noch: die Regeln, über die beide Konzepte arbeiten, können identisch sein.

## 2.1.3. Übersetzung von Komposita

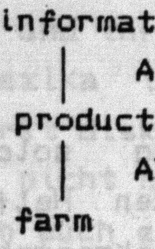
Ein Problem der maschinellen, aber auch der Humanübersetzung ist die Übersetzung von sog. 'unbekannten' Wörtern, meist Komposita, vor allem Fachtermini. Nur bei Verwendung von umfangreichen Terminologien ist sie befriedigend zu lösen. Aber auch im besten Fall bleibt für die maschinelle Übersetzung immer ein Rest von Komposita, der nicht lexikalisch übersetzt werden kann. Hier muß ein Verfahren entwickelt werden, das sowohl die Komposita morphologisch zerlegt, als

auch deren Komponenten in einen semantisch korrekten Zusammenhang bringt. Die morphologische Analyse kann als erledigt bezeichnet werden. Für die Übersetzung wird unterschieden, ob der Begriff als ein Wort oder als eine Wortfolge ausgedrückt wird.

Bei mehrwortigen Ausdrücken, beispielsweise englischen string compounds, wird ihre syntaktische Struktur ermittelt, die eine Übersetzung entweder als Kompositum oder als Nominalstruktur zuläßt. Beispiel:

farm product information

erhält die Struktur




Dies führt letztendlich zu Übersetzungen wie

Information über Agrarprodukte
information sur les produits agricoles
Agrarproduktinformation

Komposita werden einer semantischen Analyse unterworfen, die die Bestandteile und die Relationen zwischen ihnen festlegt. Diese Relationen resultieren entweder aus der Transformation der Komposita in den verbalen Rahmen, vgl.

Zuckerproduktion =====> produzieren (AKK) Zucker

oder aus der semantischen Analyse ihrer Nominalgruppenstruktur, vgl.

Information über Produkte =====>
Information (sem.Rel.) Produkte

Die semantische Analyse leistet im Zusammenhang mit semantischen Relationen auch die Erzeugung von Nominalgruppen aus Adjektiven, vgl.

golden ====> (sem.Rel.) Gold
erfolgreich ====> (sem.Rel.) Erfolg

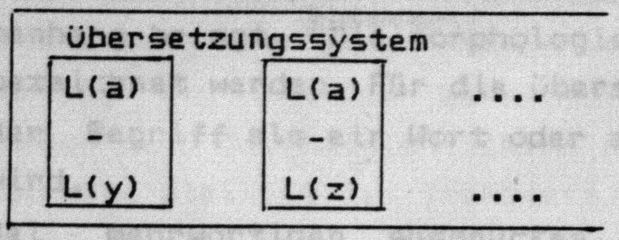
Strukturen, die auf diese Weise erzeugt werden, sind zu der ursprünglichen synonym gesetzt. Der Transfer und die Synthese wählen aus diesen dann die am besten übersetzbare aus. Einzelsprachenabhängig werden aus Nominalgruppen entweder Komposita oder wieder Nominalgruppen generiert. Für das Deutsche werden z.B. Komposita bevorzugt. Dabei werden auch Fugenmorpheme korrekt erzeugt.

### 2.1.4. Tabellen

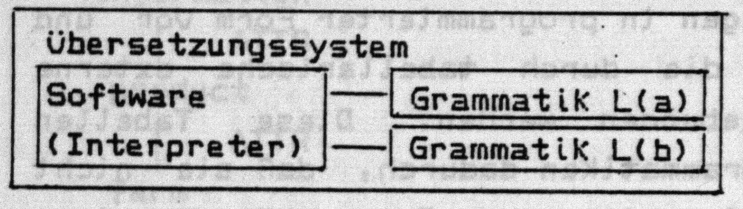
Eingangs wurde die Formulierung "In der syntaktischen Analyse sind Grammatik und Programm größtenteils nicht in der üblichen Weise voneinander getrennt" benutzt. Dies sollte näher erläutert werden.
Verschiedene Moduln benutzen zwar auch extern gelagerte linguistische Daten, sind aber keine interpretierenden Systeme. Die Regeln liegen in programmierter Form vor und verarbeiten Variablen, die durch tabellarische externe linguistische Daten gebunden werden. Diese Tabellen unterscheiden sich von Grammatiken dadurch, daß sie nicht die Struktur und die inneren Zusamenhänge eines Sprachausschnittes darstellen. Die Tabellen haben ihre eigene Syntax. Mit Inhalt werden sie durch den Linguisten gefüllt. Unter dem Kapitel "Gewichtungen" wird auf eine spezielle Tabelle, die Gewichtungstabelle, näher eingegangen werden.

### 2.1.5. Multilingualität

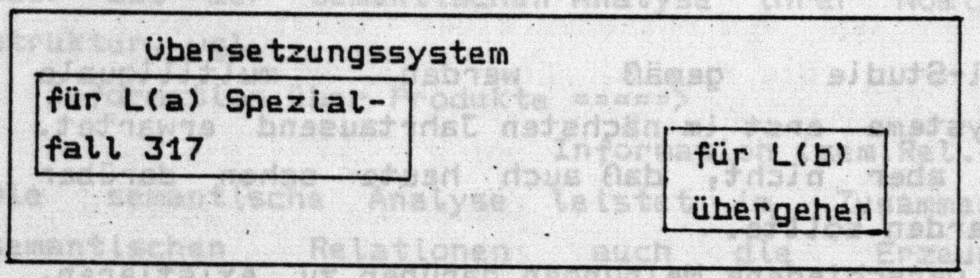
Der Delphi-Studie gemäß werden multiliguale Übersetzungssysteme erst im nächsten Jahrtausend erwartet. Das bedeutet aber nicht, daß auch heute schon darüber nachgedacht werden sollte.
Es scheinen verschiedene Meinungen darüber zu existieren, was an einem Übersetzungssystem multilingual ist. Es soll nun versucht werden, diese zu charakterisieren.
Der unterste Grad - wenn überhaupt - liegt vor, wenn Analyse und Synthese voneinander abhängig sind, d.h. wenn die Analyse der Sprache L(a) im Übersetzungsvorgang in die Sprache L(y) eine andere ist als im Übersetzungsvorgang in die Sprache L(z). Das Multiliguale reduziert sich auf eine gemeinsame Systemumgebung, wobei der interne Ablauf dem Benutzer nicht sichtbar ist. Die Analysen sind dabei nicht nur ein Duplikat voneinander, sondern echt verschieden.



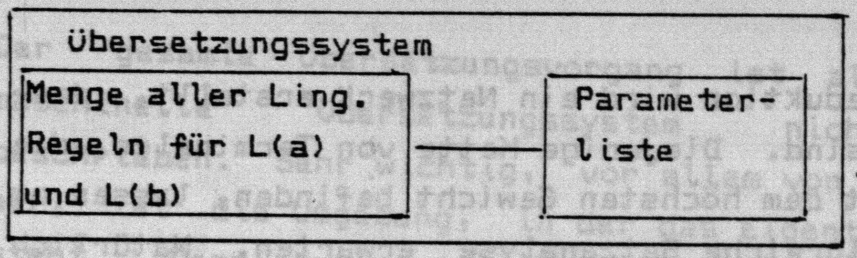

Ich möchte nun trennen zwischen Systemen, die mit einem Interpreter und einer Grammatik arbeiten und solchen, bei denen die Grammatik größtenteils programmiert vorliegt. Im ersten Fall liegen z.B. der Interpreter und die Grammatiken für L(a) und L(b) für die Analyse vor:



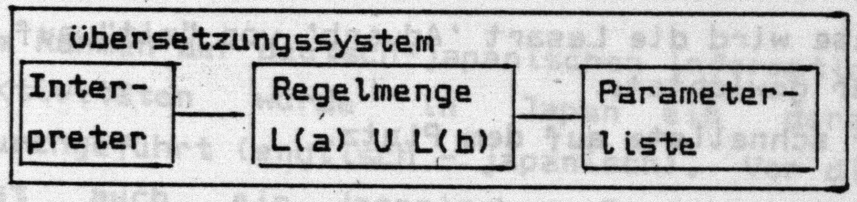

Der multilinguale Charakter ist in diesem Fall hauptsächlich festgemacht am Interpreter, an einer gemeinsamen Datenstruktur und einer gemeinsamen Sprache zur Formulierung der Gramatik. Falls die Grammatiken compiliert werden, liegen zwei getrennte Systeme vor.
Falls die Grammatik in Programmform vorliegt, kann Multilingualität durch das sehr unübersichtliche Anbauen von "Balkönchen" erreicht werden, d.h. durch Abfragen der Art: "falls Spache L(a), dann Spezialfall Nummer 317".




Die gravierenden Probleme dieser Unsystematik liegen auf der Hand. Nach kurzer Zeit wird niemand mehr wissen, wie z.B. die Grammatik der Sprache L(b) aussieht und wie sie sich von der Grammatik von L(a) unterscheidet. Das Problem tritt nicht auf, wenn es neben der Grammatik in Programmform, die die Menge aller linguistischen Regeln der Sprachen L(a) und L(b) enthält, auch eine Parameterliste gibt, die die zwischen L(a) und L(b) unterschiedlichen Regeln einzelsprachspezifisch angibt.




Diese Problemlösung soll nun auch auf interpretierende Systeme angewandt werden. Es soll deshalb geschehen, um an einem Extremfall die Problematik der Multilingualität in den Brennpunkt zu setzen.
Neben dem Interpreter besteht das System aus einer Grammatik, die die Regeln der Sprachen L(a) und L(b) enthält. Der Terminus 'Grammatik' ist in diesem Zusammenhang - und damit beginnt schon die Problematisierung - gewiß nicht ganz korrekt, da diese Regelmenge beispielsweise auch sich einander ausschließende Regeln enthalten kann.
Dazu existiert eine Parameterliste, die die zur Sprache L(i) gehörenden Regeln auswählt.




Dieses Konzept kann natürlich nur angerissen werden. Es stellt sich eine ganze Reihe von Fragen, die diskutiert werden müssen. Beispielsweise "Wie wirkt sich die Regelordnung in diesem Fall aus?" oder "Welche Auswirkungen auf das Konzept der Systemkontrolle sind zu bedenken?" Neben Fragen nach theoretischen Problemen ist zu untersuchen, wie sich der Aufwand für die Organisation der Multilingualität verhält.

### 2.1.6. Gewichtung

Unabhängig von der Konzeption des Parsers oder eines Übersetzungssystems muß man davon ausgehen, daß Verfahren zur Gewichtung in der Zukunft immer mehr an Bedeutung gewinnen werden. Ich möchte nicht näher darauf eingehen, wie Gewichtungen berechnet werden und wo sie zum Tragen kommen können, sondern ich stelle ein konkretes Beispiel für die Benutzung von Gewichtung vor, um diese Idee zu

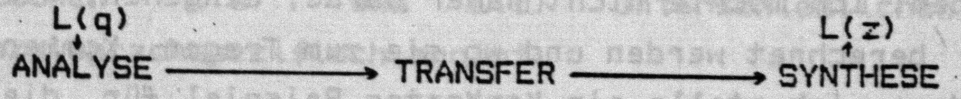
verdeutlichen.
In der Homographenreduktion wird ein Netzwerk erstellt, dessen Wege gewichtet sind. Diejenige Kette von Terminalen, die sich auf dem Weg mit dem höchsten Gewicht befinden, lassen am ehesten eine erfolgreiche Satzanalyse erwarten. Natürlich werden neben der Gewichtung auch andere Verfahren zur Homographenreduktion herangezogen. Die Gewichtung der Wege berechnet sich aus der Gewichtung von Paaren von Wortklassen. Somit repräsentiert dieses Paargewicht quasi den 'Normalfall' der Wortklassen des Paares. Nun können aber nicht alle Wörter z.B. der Wortklasse 'Adverb' die gleiche Gewichtung haben. Die Modifikation der Standardwerte ist natürlich auch abhängig von den übrigen Lesarten des Wortes. Mit anderen Worten: die Klasseneinteilung kann zu grob sein. Eine feinere würde aber ihre Handhabung und die Differenzierung der zugehörigen Gewichte erschweren. Daher wird im Wörterbuch ein Herab- oder Heraufsetzen der in der Gewichtungstabelle bei Paaren vorgegebenen Standardwerte erlaubt - einzelwortabhängig. Beispielsweise wird die Lesart 'Adverb' von "mit" auf diese Weise niedriger gewichtet:

Er ist mit der schnellste auf dem Platz.

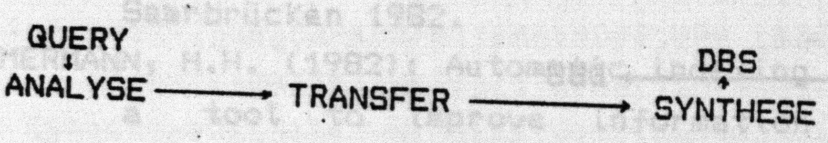
## 2.2. Multifunktionale Subsysteme

Im folgenden werden Vorschläge für Systeme entwickelt, deren Subsysteme als vorhanden oder als in Entwicklung bezeichnet werden können. Für die vorgestellten Konzeptionen versuche ich jeweils ein entsprechendes Projekt vorzustellen. Dabei beschränke ich mich, da ich dort das Umfeld am besten kenne, auf die Aktivitäten in Saarbrücken.

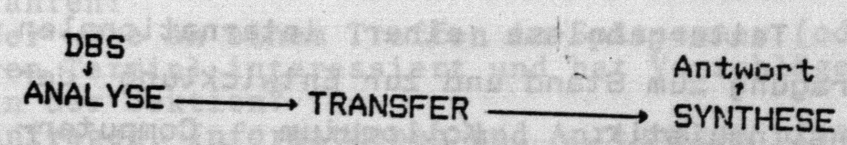
Unterteilen wir eine maschinelle Übersetzung in die Komponenten Analyse, Transfer und Synthese. Der Übersetzungsvorgang wird dann durch die Eingabe eines quellsprachlichen Textes (Satzes) und die Ausgabe eines zielsprachlichen darstellbar:



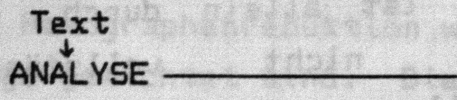
Der gesamte Übersetzungsvorgang ist allein durch das maschinelle Übersetzungssystem nicht vollständig beschrieben. Sehr wichtig, vor allem von der Benutzerseite her, ist die Umgebung, in der das eigentliche Kernsystem steht. Denn damit ist in hohem Maße seine Akzeptanz verbunden. Der Frage, wie ein bestehendes System in eine funktionierende (Human-) Übersetzertätigkeit integriert werden kann, wird z.Zt. im Projekt SUSY-BSA nachgegangen. Bleiben wir bei dieser unveränderten Architektur des Übersetzungssystems und dem Aufbau der einzelnen Komponenten und betrachten einige Beispiele: als Erstes den Einsatz eines Systems zur maschinellen Übersetzung bei der natürlichsprachlichen Anfrage an eine fremdsprachliche Datenbank:



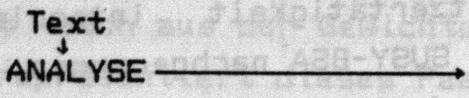

Im Rahmen der deutsch-japanischen informationslinguistischen Aktivitäten wurde in Japan ein derartiger Versuch durchgeführt (englisch - japanisch). Vor diesem Hintergrund ist auch ein japanisches Projekt zur automatischen Übersetzung zu sehen: In einem 'National Project' werden im Zeitraum 1982 - 1984 ca. 8 Mio DM investiert, um eine japanisch - englische Übersetzung von Abstracts zu ermöglichen. Ziel ist es, eine englische Datenbank mit wissenschaftlichen Titeln aus dem Japanischen aufzubauen. Umgekehrt könnte die Antwort aus der Datenbank rückübersetzt werden, vorausgesetzt natürlich, diese Antwort liegt natürlichsprachlich vor. Eine derartige Konstellation ist z. Zt. im Gespräch zwischen dem Projekt A2 des Sonderforschungsbereichs 'Elektronische Sprachforschung' und der European Space Agency (ESA).



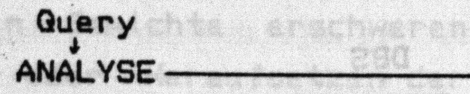

Der Einsatz der Analysekomponente läßt sich in einer Reihe von Projekten nachweisen, wie z. B. in dem 1980 abgeschlossenen Projekt "Automatische Lemmatisierung" des Sonderforschungsbereichs 100:




oder in dem System CTX des Projektes Transit: Hier dient die Analyse zum Automatischen Indexieren:




Die Analyse erlaubt aber auch eine natürlichsprachliche Formulierung der Anfrage in Form einer Problembeschreibung. Diese deskriptive Retrievalsprache erweist sich in CTX als sehr benutzerfreundlich:




LITERATUR:

FREIGANG, K.-H. (1982): Erfahrungen bei der Anpassung eines MÜ-Systems an eine neue Sprache bzw. einen neuen Texttyp. Kolloquium Computerlinguistik zwischen Forschung und Anwendung, Saarbrücken 1982.

HAHN,U./KUHLEN,R./REIMER,K. (1982): Konzeption und Aufbau des autmatischen Textkondensierungssystems TOPIC: Bericht TOPIC-1/82, Sept. 1982

KITTREDGE,R./LEHRBERGER,J. (eds) (1982): Sublanguage. Studies of Language in Restricted Semantic Domain. Berlin/New York.

KUHLEN,R. (1982): Themen der Automatischen Übersetzung: Bedeutung, Machbarkeit, Realisierungszeitraum. Vorläufige Teilergebnisse einer internationalen Delphi-Befragung zum Stand und zur Entwicklung der Informationslinguistik. Kolloquium Computerlinguistik zwischen Forschung und Anwendung. Saarbrücken 1982.

LEHRBERGER, J. (1982): Automatic Translation and the concept of sublanguage, in: KITTREDGE/LEHRBERGER

(1982), S. 81-106.

LUCKHARDT,H.-D. (1982): SUSY: Capabilities and range of application, in: Multilingua 1-4, 1982.

MAAS, H.- D. (1979): Das Saarbrücker Übersetzungssytem SUSY, in: Sprache und Datenverabreitung 1, S. 43-64

SAGER, N. (1982): Syntactic Formatting of Science Information, in: KITTREDGE/LEHRBERGER (1982), S. 9-27.

SCHMITZ, K.-D. (1982): Bermerkungen zum Umfeld eines MÜ-Systems. Kolloquium Computerlinguistik zwischen Forschung und Anwendung, Saarbrücken 1982.

THIEL, M. (1982): Systemarchitektur von SUSY unter benutzerspezifischem Aspekt. Kolloquium Computerlinguistik zwischen Forschung und Anwendung, Saarbrücken 1982.

ZIMMERMANN, H.H. (1982): Automatic indexing and retrieval as a tool to improve information and technology transfer. Bled Conference on Computational Linguistics, Bled 1982.

# Arbeitskreis maschinelle Übersetzung

Betrifft: Mitarbeit in der Sektion 'Maschinelle Übersetzung'

Auf der letzten Vollversammlung in Trier wurde im Anschluß an die Vorträge in der Sektion 'Maschinelle Übersetzung' eine Fortsetzung der Aktivitäten in dieser Sektion vereinbart.
Die Koordination wird von H.J.Weber vorgenommen.
Wer arbeitet an Maschineller Übersetzung oder bereitet ein MÜ-Projekt vor?
Wer wünscht Informationen über maschinelle Übersetzungsverfahren?
Wer wäre an einem Treffen im Spätsommer (oder zu einem anderen Termin) interessiert und hat Vorschläge für die Arbeit in der Sektion?
Anfragen, Informationen und Anregungen bitte an

H.J.Weber
SFB 1oo (A3)
Bau 4, Universität
66oo Saarbrücken
Tel. o681/811168

# Tagungsberichte

## Informationswissenschaftliches Kolloquium an der Universität Konstanz

G. Knorz, TH Darmstadt

Am 1. und 2. Juli 1983, an einem Freitag und Samstag (!), waren alle informationswissenschaftlichen Gruppen der Bundesrepublik eingeladen, sich zu einem Kolloquium in Konstanz zu treffen. In Übersichtsvorträgen sollte die Situation der drei informationswissenschaftlichen Lehrstühle (Wersig, Berlin; Zimmermann, Saarbrücken; Kuhlen, Konstanz) dargestellt werden: Personal, Ausbildungsstand, institutionelle Aufhängung, Projektförderung, aber auch informationswissenschaftlicher Standpunkt, Methoden, Perspektiven (Einladungsschreiben von Prof. Kuhlen). Ein halber Tag war der Darstellung von Entwicklungstendenzen in anderen Institutionen gewidmet (Berlin (TU), Bremen, Darmstadt, Düsseldorf, Frankfurt, Hannover, Köln, Stuttgart, Regensburg).

Je zwei Spezialvorträge der zuerst genannten Forschungsgruppen machten den Charakter des Kolloquiums als eine Fachveranstaltung aus, die im übrigen eine (wissenschafts-) politische Veranstaltung war. Für die breite Diskussion einer generellen Informationspolitik war am Freitag Abend ein großzügiger zeitlicher Rahmen vorgesehen (und am Samstag Abend auf Kosten einer ausbildungsbezogenen Veranstaltung zusätzlich eingeräumt) - doch auch ansonsten waren viele Diskussionsbeiträge von dem politischen Aspekt beherrscht. Hintergrund (und 'drohende' Perspektive) waren unter anderem das im April erschienene Gutachten des Bundesrechnungshofes zur Lage der Fachinformation und ein für den Herbst beim BMFT geplantes Hearing über Informationswissenschaft. In diesem Zusammenhang war es das Ziel, zu einem Konsens zu kommen und diesen als 'Konstanzer Thesen' zu formulieren.

Die Ausgestaltung des inhaltlichen Schwerpunktes, nämlich die Beiträge der genannten inf. wiss. Lehrstühle, kann sicher ohne Einschränkung als gelungen bezeichnet werden. Deutlich

mehr Mühe hatten die anderen Beitragenden in einem zunehmend drastisch beschränkten Zeitrahmen und bei dann doch nachlassender Konzentration des Teilnehmerkreises, eine effektive Plattform zu finden. Ein so harmloser Programmpunkt, wie 'Mittagessen im Restaurant' kann die zeitlichen Vorstellungen eines Veranstalters eben doch in unvermutetem Ausmaß sabotieren.
Die eigentlichen Probleme jedoch zeigten sich bei der Behandlung der wissenschaftspolitischen Fragen. Um es vorwegzunehmen: die angestrebten Konstanzer Thesen wird es auf Initiative des Veranstalters wohl geben, jedoch erst im nachhinein und (überspitzt und provokativ formuliert) eigentlich fast trotz und nicht etwa als Ergebnis der Diskussionen in Konstanz. Bei der Auseinandersetzung ging es - verkürzt und vergröbert dargestellt - im Kern darum, entweder den 'technokratischen' Standpunkt Kuhlens oder aber Wersigs Vorstellung einer 'Metawissenschaft' in den Mittelpunkt des Selbstverständnisses von Informationswissenschaft zu stellen. Das insgesamt Erfreuliche war, daß die Diskussion um Ausrichtung und letztlich (?) auch um (Förderungs-) Mittel nicht verdeckt (wie z. T. in anderen Kreisen), sondern sehr offen geführt wurde. Das Bedauerliche war, daß nur selten konstruktiv und zielgerichtet diskutiert wurde, sondern daß eine faire und inhaltliche Auseinandersetzung häufig durch die Verwendung selbstgezimmerter bequemer Feindbilder umgangen wurde. Während die Wünschbarkeit und Notwendigkeit eines ganzheitlichen Ausgangspunktes oder Hintergrundes für die Informationswissenschaft nicht in Frage gestellt wurde, kam aus der Ecke der 'Technokraten' (zu denen ich in diesem Zusammenhang wohl gerechnet werde) der Einwand, daß eine ausschließliche Beschränkung auf den metawissenschaftlichen Standpunkt (bei dem alle weiteren Wissenschaften, wie z.B. Informatik, KI, Soziologie, Kommunikationstechnologien, ... den Status von Hilfswissenschaften haben) zu einer - letztlich nicht als seriös anzusehenden - Wissenschaft ohne harten Kern führt - zu einer Wissenschaft, die stets soweit ausholt, daß sie der Lösung aktueller Probleme nur hinterherlaufen kann bzw. sie anderen Wissenschaften überlassen muß. In diesem Zusammenhang drängt sich mir die Frage auf, ob die Philosophie

ihren heutigen Stand hätte erreichen können, wenn die Physik auf Dauer Philosophie geblieben wäre. Bei solchen Betrachtungen weiterzufahren würde sicher bedeuten, ein eigenes Positionspapier zu erstellen, was jedoch nicht Sinn dieses kurzen Berichtes sein kann. Wie bereits angedeutet, lassen sich bei gutem Willen (und unter gewissen Notwendigkeiten) durchaus Ansatzpunkte für ein konsensfähiges Thesenpapier finden. Diese Arbeit hat sich der Veranstalter vorgenommen, bei dem man sich im übrigen für die Mühe und letztlich auch den Erfolg eines Kolloquiums mit insgesamt ca. 40 - 50 Teilnehmern bedanken muß. In jedem Fall waren die Teilnehmer am Ende wohl einhellig der Ansicht, daß weitere Treffen dieser Art und Zusammensetzung sicher eine wünschenswerte Sache sind. Wenn ich zum Schluß meine Erinnerungen noch nach besonderen Eindrücken durchforsten will, so fällt mir neben den in der Tat nicht zu vernachlässigenden sozialen und fachlichen Kontakten spontan Folgendes ein : die interessante Gelegenheit, die Ausstattung in Konstanz mit DV- und Kommunikationstechnologie zu besichtigen, die breite Bereitschaft, in den späten Abendstunden selbst noch am Samstag ein ernsthaftes Kolloquiumsprogramm mitzutragen, eine außergewöhnliche und bunte Hochschularchitektur, ein Wetter, das ganz zum Urlaubsgebiet 'Bodensee' paßt, und dann so manches Interessante und Gemunkelte aus der 'Gerüchteküche'.

Informationswissenschaftliches Kolloquium an der Universität Konstanz

1. - 2. Juli 1983

Raum C 203

---

1. Juli

11.00 - 12.30 Universität Konstanz

a) Oberblick
b) TOPIC/TOPOGRAPHIC

- Mittagessen in der Mensa -

14.00 - 15.30 Universität Berlin (FU)

a) Oberblick
b) Spezialproblem I

- 15.30 - 16.00 Kaffeepause -

16.00 - 17.30 Universität Saarbrücken

a) Oberblick
b) Spezialproblem I

(ab 17.30 Möglichkeit zur Begehung des DV-Labors)

Abendessen in benachbarten Restaurants; z.B. Mainaublick, Arche

20.00 - ... Informationspolitik
Perspektiven eines neuen IuD-Programms

2. Juli

10.00 - 11.00 Universität Konstanz

Kommunikationstechnologie und Organisationsstruktur

- Kaffeepause -

11.30 - 12.30 Universität Saarbrücken

Spezialproblem II

- Mittagessen im Restaurant -

14.00 - 15.00 Universität Berlin

Spezialproblem II

- Kaffeepause -

15.30 - 19.00 Entwicklungstendenzen in anderen Institutionen

Berlin (TU), Bremen, Darmstadt, Düsseldorf, Frankfurt, Köln, Stuttgart, Regensburg

20.00 - ... Ausbildungsperspektiven

Forschungsvorhaben

**Bei den Nachbarn im KI-Rundbrief erschienen:**

## Bericht vom LDV-Kolloquium 1983, Trier (Jahresversammlung der LDV-Fittings e.V.)

Vom 2.3. bis 4.3.1983 fand das diesjaehrige LDV-Kolloquium der LDV-Fittings an der Universitaet Trier (Tarforst) statt. Schwerpunkte des Treffens bildeten die Themen:

- Einsatz von Mikro-Computern/Textverarbeitungssystemen in der LDV
- Computergestuetzte Inhaltsanalyse (und deren Nutzung in den Sozialwissenschaften)
- Maschinelle Uebersetzung (neue Ansaetze zur automatisierten Sprachuebersetzung).

### Sektion Mikro-Computer/Textverarbeitungssysteme

Zu dem Themenbereich "Mikro-Computer und Textverarbeitung in der LDV" wurden zunaechst einfuehrende Vortraege ueber die Bereiche Hardware (H. Kuke, Gesellschaft fuer wissenschaftliche Datenverarbeitung Goettingen), Software (Th. Krug, BSP Regensburg) sowie Textverarbeitungssysteme (TVS) (S. Mueller-Zantop, IBM Duesseldorf) gehalten.
Die Tendenz des Einsatzes von Mikro-Computern (MC) geht von den Systemen mit 8-BIT-Prozessoren zu Geraeten mit 16/32-BIT-Prozessoren, die aufgrund des groesseren Adressraumes mehr RAM-Speicher besitzen (mit entsprechenden Hintergrundspeichern z.B. in Form von Winchester-Platten). Basisfunktionen, die bisher durch Software realisiert sind, werden zunehmend durch entsprechende Hardware realisiert, wobei die Verarbeitung durch Co-Prozessorsysteme verbessert wird (Parallel-Processing).
Im BS-Bereich ist die CP/M-"Familie" am weitesten verbreitet, die auf 16/32-BIT-Prozessorsysteme ausgedehnt werden soll (hier hat sie allerdings Konkurrenz in den UNIX-Betriebssystemen). Vorherrschend sind als Programmiersprachen BASIC und PASCAL, fuer den realen Einsatz von Datenbanksystemen ist ein entsprechend grosser Hintergrundspeicher (> 5 MB) notwendig.
Im Textverarbeitungsbereich duerften fuer Forschungseinrichtungen wohl eher Mikro-Computer mit Textverarbeitungssoftware interessant sein (im Gegensatz zu reinen TVS).
Die vom Arbeitskreis "Mikro-Computer" durchgefuehrte Umfrage zum Einsatz von MC / TVS brachte (mit nur 11 Ruecklaeufen) kaum verwertbare Ergebnisse; die MC werden wohl weniger als TVS eingesetzt (da dafuer meist auf den Haus(gross)rechnern mehr oder weniger komfortable TVS-Software zur Verfuegung steht), sondern in erster Linie fuer spezielle Forschungsaufgaben genutzt (wie auch wir in Berlin unseren Mikro "Osborne 1" primaer fuer Implementierungen mit PROLOG einsetzen).

Der zweite Teil des Schwerpunktes MC beschaeftigte sich mit dem Einsatz in Schule und universitaerer Ausbildung (U. Klenk, Universitaet Goettingen), und mit neuen Software(-konzepten) fuer MC (D. Metzing / F. di Primio, Uni Trier / Bielefeld) sowie interaktiver Programmierung mit MC (R. Lawler, Uni Paris

/ MIT). Dieter Metzing berichtete ueber den Einsatz objektorientierter Sprachen (am Beispiel FRL) im Projekt TESSDAT (Bearbeitung von Informationen ueber Sprachkurse in England). F. di Primio referierte ueber die objektorientierte Programmierung in LISP. Im Vortrag von R. Lawler ueber interaktive Programmierung mit LOGO ging die Darstellung der Vorteile derartiger Programmiersprachen (insbesondere der Eigenschaften objektorientierter Sprachen) etwas unter bei dem Versuch, dies im Auditorium auf einem MC-Bildschirm anhand von Beispielen ueber Fremdsprachenerlernung und Aufbau von "(Mikro-)Welt-Szenarios" (graphik-unterstuetzte Darstellung von bewegten Objekten wie Autos, Flugzeugen u.ae. in einer Spielwelt) zu demonstrieren. (R. Lawler beschaeftigt sich mit der Unterstuetzung von Lernprozessen durch Computereinsatz (Lernanreize)). Unklar blieb die Programmierung in LOGO und der Aufwand fuer die Definition von graphischen Objekten.

Sektion Computergestuetzte Inhaltsanalyse

Den zweiten Schwerpunkt der Tagung bildeten die Nutzungsmoeglichkeiten der Computergestuetzten Inhaltsanalyse (CUI) in den Sozialwissenschaften (wobei sich z.Zt. die Computerunterstuetzung auf die quantitative Auswertung von Texten hauptsaechlich mittels statistischer und syntaxorientierter Verfahren bezieht und nicht mit dem "Verstehen" von Textinhalten gleichzusetzen ist). Hiermit sollte eine Zusammenarbeit zwischen LDV und Sozialwissenschaften vertieft werden, die auf der vorjaehrigen Tagung in Koblenz initiiert worden war.
Einfuehrend referierte E. Mochmann (Uni Koeln) ueber den Einsatz der Inhaltsanalyse in den Sozialwissenschaften. Die Ergebnisse der quantitativen CUI werden unter dem Blickwinkel sozialwissenschaftlicher Theorien ausgewertet. Gegenwaertige Probleme sind die Analyse grosser Einweg-Korpora (Korpora fuer einen einmaligen Gebrauch) zu vertretbaren Kosten und die Automatisierung der Disambiguisierung (P. Mohler, Zentrum fuer Umfragen, Methoden und Analysen Mannheim). Die Analyse von Therapiegespraechen erfordert aufgrund bruchstueckhafter (syntaktisch unkorrekter) Saetze semantische (!) Analysen unter Beruecksichtigung der Kommunikationsziele (!) von Patient und Therapeut (E. Mergenthaler, Uni Ulm). Ergaenzt wurde die sozialwissenschaftlich orientierte Darstellung durch eine vergleichende Darstellung verfuegbarer LDV-Analysesysteme (K. Petersen, Bonn) und einen Ueberblick ueber "Modelle und Methoden der quantitativen Linguistik" (R. Koehler, Uni Essen). Nach dieser Darstellung der Probleme und Moeglichkeiten setzte ein reger Austausch von Informationen und Testergebnissen / -ausdrucken ein; offensichtlich traf die Wahl dieses Schwerpunktes auf starkes Interesse (wobei es allerdings noch einige Verstaendigungsprobleme auf beiden Seiten gibt).

Sektion Maschinelle Uebersetzung

Im Mittelpunkt der dritten Sektion standen die Planung zur Realisierung eines mehrsprachigen maschinellen Uebersetzungssystems (Projekt EUROTRA).
S. Perschke (EG-Kommission, Luxemburg) gab einen Ueberblick ueber die Zielsetzung und Planung des Projektes, das aus drei Phasen von jeweils ca. zwei Jahren bestehen soll:

- linguistische Modellbildung / Schnittstellendefinition / Basissoftware;
- Aufbau eines Modellsystems (mit ca. 2.500 Termini);
- Entwicklung eines prototypischen Systems (mit ca. 20.000 Termini) / Bewertung in Hinblick auf eine industrielle Entwicklung.

Die Repraesentationssprache (fuer Syntax und Semantik) soll zentral bei der EG entwickelt werden, die Transformationskomponenten fuer die jeweiligen Mitgliedssprachen in den einzelnen Laendern. Weltwissen soll dabei beruecksichtigt werden ("soweit verfuegbar"), auch ueber die notwendige Tiefe der semantischen Repraesentation ist wohl noch nicht entschieden. Der Vortrag vermittelte den Eindruck, dass beim Projektfoerderer die (doch wohl nicht realistische) Meinung vorherrscht, die bisherigen KI-Forschungsergebnisse auf den entsprechenden Gebieten seien als anwendungsreif anzusehen. Anhand des Saarbruecker Uebersetzungssystems 'SUSY' versuchte M. Thiel (Uni Saarbruecken) aufzuzeigen, welche Erweiterungen (aus dortiger Sicht) notwendig seien. Insgesamt schienen mir noch einige Unklarheiten insbesondere bezueglich der Wissens- und Textrepraesentation vorhanden zu sein.

Auf der Mitgliederversammlung hielt B. Endres-Niggemeyer einen Vortrag ueber die Perspektiven der LDV-Forschung: Die von der EG gestarteten Forschungsprogramme EUROTRA und ESPRIT beeinflussen in starkem Masse auch die deutsche Forschungsfoerderung. Hierbei wird verstaerkt auf die Einbeziehung der universitaer-wissenschaftlichen Arbeiten in die industrielle Forschung hingearbeitet, was hoffentlich nicht zu einer Verringerung langfristiger Grundlagenforschung fuehren wird!

Am Rande sei noch vermerkt, dass die "LDV-Fittings" e.V. sich auf ihrer Mitgliederversammlung zu einer Namensaenderung entschlossen haben (weil wohl immer wieder Fragen auftauchten, was denn "Fittings" hier eigentlich bedeuten sollte): sie bilden jetzt die "Gesellschaft fuer Linguistische Datenverarbeitung (GLDV)".

Michael Koenig
TU Berlin, Projekt KIT

# Projekte

**Ab 1.7.1983 wird es ein neues KI-Projekt in Stuttgart geben:**

Japanisch/deutsche Titeluebersetzung:

**SEMSYN, ein System zur deutschen Synthese auf der Basis einer semantischen Repraesentation**

Das **SEMSYN-Projekt** wird vom **Bundesminister fuer Forschung und Technologie** waehrend **des Zeitraums vom 1.7.1983 bis zum 30.6.1985** finanziert (Bewilligungsschreiben vom 10.5.1983; Foerderkennzeichen 1013203 7) und an der **Abteilung Dialogsysteme des Instituts fuer Informatik** der Uni Stuttgart (Leiter: Prof.Dr.Gunzenhaeuser) in Kooperation mit der japanischen Firma **FUJITSU** durchgefuehrt werden. Diese unterstuetzt das Projekt u.a. durch kostenlose Ueberlassung von Analyseergebnissen ihres Uebersetzungssystems **ATLAS/U** (bisher fuer die Uebersetzung vom Japanischen ins Englische verwendet).

**Projektaufgabe von SEMSYN** ist, ein **Synthesesystem fuer Deutsch** zu entwickeln, das als Teil eines Japanisch/Deutsch-Uebersetzungssystems auf der von ATLAS/U aus Japanisch erzeugten semantischen Repraesentation als Schnittstelle aufsetzt. Die von ATLAS/U zur Inhaltsdarstellung verwendeten semantischen Netze sind anderen in der KI verwendeten Formalismen vergleichbar und wurden vor allem von Fillmores Kasusrahmentheorie und Schanks CD-Netzen inspiriert (vgl. Beitrag von Uchida/Sugiyama in: COLING 80, Proceedings).

**Prototypische Anwendung von SEMSYN** wird zunaechst die **automatische Uebersetzung** von ca. 3000 **Titeln japanischer Arbeiten** aus dem Gebiet "Informationstechnologie" sein. Die Forderung nach bequemer Erweiterbarkeit auf andere Gebiete durch Lexikonergaenzung soll bereits beim Systemdesign bestimmend sein. Ein spaeterer Ausbau auf die Uebersetzung von Abstracts, die waehrend des Projekts nur versuchsweise erfolgen soll, oder Volltexten wird angestrebt.

Der Projektbewilligung war 1982 u.a. eine von der GID in Auftrag gegebene **Pilotstudie** ueber die moegliche Anwendbarkeit von ATLAS/U fuer die Projektaufgabe und ein zweitaegiger **Workshop** (29./30.11.1982) ueber diese Frage mit Vertretern von BMFT, GID und potentieller Anwender aus Patentamt, Fachinformationszentren usw. vorausgegangen.

In die **Projektleitung** von SEMSYN werden sich **Dr.K.Hanakata** und **Dr.J.H.Laubsch** teilen; als **Mitarbeiter** konnten **A.Lesniewski** (frueher bei Prof. Bolc, Warschau) und **D.Roesner** gewonnen werden.

Zur **Programmierumgebung von SEMSYN** wird - neben Zugang zu einer VAX - eine **projekteigene LISP-Maschine** (voraussichtlich: SYMBOLICS 3600) gehoeren.

-.-.-.-.-.-.-.-

Als "Einstiegsveranstaltung" fuer SEMSYN ist ein

**INTERNATIONALER WORKSHOP zu Fragen der GENERIERUNG**

geplant, der vom

**15.8. bis 17.8.1983**

auf Burg Stettenfels/Untergruppenbach stattfinden soll (also unmittelbar im Anschluss an die IJCAI, um - trotz des enorm knappen Zeitraums fuer die Organisation - moeglichst viele internationale Experten versammeln zu koennen).

Als Referenten wurden Wissenschaftler aus Europa und den USA eingeladen, die an **Generierungssystemen fuer Deutsch, Englisch** oder vergleichbaren Sprachen arbeiten. Die Referenten sollen waehrend des Workshop einen Statusbericht zu ihrer Arbeit geben. Es ist beabsichtigt, ausreichend Zeit fuer die Diskussion der vorgestellten Ansaetze und Ergebnisse zur Verfuegung zu haben.

-.-.-.-.-.-.-.-

**Kontaktadresse:**

**Projekt SEMSYN**
c/o Frau Hoermann
Institut fuer Informatik
Universitaet Stuttgart
Azenbergstr. 12
D-7000 Stuttgart 1
Tel: 0711/ 2078 - 338

# AUTOMATISCHES TEXTKONDENSIERUNGSSYSTEM T O P I C

TOPIC (Text-Oriented Procedures for Information Management and Condensation of Expository Texts) ist ein wissensbasiertes Textanalysesystem zur Zusammenfassung von deutschsprachigen Fachtexten (Volltexten) im Sinne des automatischen Abstracting. Thematisch sind die Texte dem Bereich verwaltungsorientierter Informations- und Kommunikationstechnologien zuzuordnen.

Das System wird am Lehrstuhl fuer Informationswissenschaft der Universitaet Konstanz im Rahmen des BMFT-Projekts "Informationsvermittlung" (Projektleiter: Prof. Dr. R. Kuhlen, Laufzeit: 1/1982 - 12/1984) entwickelt. Neben dem Projektleiter arbeiten 2 wissenschaftliche Mitarbeiter, U. Hahn und U. Reimer, am Entwurf und der Implementierung.

Schwerpunkte der Projektarbeit sind:

(1) Die Adaption und Entwicklung von Parsing-Techniken und -Strategien, die der Verarbeitung von fachlichen Volltexten angepasst sind; hierzu ist das Konzept der Wort-Experten (RIEGER) aufgegriffen worden, fuer das entsprechende (voll)textspezifische Parsing-Strategien (zB. partielles Parsing, Skimming) entwickelt werden

(2) Die Adaption und Entwicklung von Wissensrepraesentationstechniken und deren Realisierung in Form eines 'knowledge base management system' (KBMS); hierzu ist das Konzept der Frames aufgegriffen worden, fuer das auf Datenmodell-Ebene bereits eine rigide Formalisierung seiner Syntax und Semantik erarbeitet worden ist, um den besonderen Quantitaets-, Konsistenz- und Aenderungserfordernissen der Wissensverwaltung von Textanalysesystemen im Kontext des Information Retrieval zu genuegen

(3) Die Entwicklung von Wissenstransformationstechniken zur Erzeugung von informationell variablen Stufen von Zusammenfassungen, die fuer die Komprimierung von Volltexten unter unterschiedlichen Nutzungskonstellationen adaequat sind

Im Rahmen des Schwesterprojekts TOPOGRAPHIC wird eine interaktive grafische Benutzer-Schnittstelle fuer TOPIC entwickelt mit einer zweiseitigen Zielsetzung:

- Unterstuetzung des Aufbaus und der Aenderung der Wissensbasen von TOPIC im Sinne eines 'knowledge base'-Editors und Wort-Experten-Editors
- grafische Repraesentation der von TOPIC generierten Textwissens- und -komprimatstrukturen (s. (3))

TOPIC wird auf einem ONYX C8002 unter UNIX in C implementiert.

Kontaktadresse: Prof. Dr. Kuhlen
Universitaet Konstanz
Informationswissenschaft
Postfach 5560
D-7750 Konstanz 1

# Publikationen

Veroeffentlichungen zu TOPIC / TOPOGRAPHIC

(1) Hahn, Udo / Rainer Kuhlen / Ulrich Reimer
Konzeption und Aufbau des automatischen Textkondensierungssystems TOPIC.
Konstanz: Projekt Informationsvermittlung, Informationswissenschaft, Universitaet Konstanz, Sept. 1982 (= TOPIC-1/82).

(2) Hahn, Udo / Ulrich Reimer
Informationslinguistische Konzepte der Volltextverarbeitung in TOPIC: Text-Parsing - Frame-basierte Wissensrepraesentation - Kondensierungstransformationen und grafische Visualisierungstechniken.
Konstanz: Projekt Informationsvermittlung, Informationswissenschaft, Universitaet Konstanz, Nov. 1982 (= TOPIC-2/82).
(Erscheint im Tagungsband des Deutschen Dokumentartags 1982)

(3) Reimer, Ulrich / Udo Hahn
A Formal Approach to the Semantics of a Frame Data Model.
Konstanz: Projekt Informationsvermittlung, Informationswissenschaft, Universitaet Konstanz, Juli 1983 (= TOPIC-3/83).
(Erscheint in den Proceedings der 8. IJCAI-Konferenz)

---



**Magdalena Zoeppritz**

## Syntax for German in the User Specialty Languages System

*Ca. 500 S. Kart. ca. DM 168.–. ISBN 3-484-31909-7*
*(Sprache und Information. Band 9)*

*Kurztext:* Das Buch beschreibt eine formale Syntax des Deutschen, die als syntaktischer Analyseteil eines Systems zur natürlichsprachlichen Interaktion mit einer Datenbank entwickelt wurde. Der Schwerpunkt liegt dabei auf der Beschreibung der in diesem Zusammenhang relevanten und formal darzustellenden Erscheinungen der deutschen Grammatik.

*Interessenten:* Linguisten, insbesondere Germanisten, Computerlinguisten

The book formally describes a comprehensive fragment of German syntax and empirically motivates the rules of the fragment. The fragment is used as the German syntax component of the User Specialty Languages System, a natural language interface to a relational data base management system, i. e. a data base system that allows query and update of data stored in tabular form. The interface is designed to work with user defined data bases and vocabularies. It uses a framework for the evaluation of sentences that has come into linguistics through Montague, but has predecessors in syntax directed compiling techniques: Rules analyzing syntactic constructions are associated with interpretation functions that provide the semantic interpretation of the constructions. The rules themselves are formulated declaratively, using a modified Backus-Normal Form extended to general phrase structure rules and augmented by features.
After introductory chapters giving the necessary background on the User Specialty Languages System, and the aims and scope of the grammar, the book focusses on the description of the grammatical phenomena to be accounted for by the rule. The rules are justified and alternative solutions are discussed, including compromises due to practical requirements and system dependent restrictions. The intention of the book is to offer a coherent treatment of a comprehensive fragment of German, whose empirical base and empirically motivated results can also be useful in other frameworks.

Gerhard Knorz

## Automatisches Indexieren als Erkennen abstrakter Objekte

*VII, 243 Seiten. Kart. DM 84.–. ISBN 3-484-31908-9*
*(Sprache und Information. Band 8)*

*Kurztext:* Im Rahmen einer allgemeinen Konzeption von automatischer Inhaltserschließung und basierend auf breit angelegten experimentellen Arbeiten werden Entwicklungsmethodologie, Konzept, praktischer Ansatz und Ergebnisse einer anwendungsnahen Lösung für die begriffsorientierte automatische Referate-Indexierung behandelt.

*Interessenten:* Linguisten, insbesondere Germanisten, Informatiker *)

An der TH Darmstadt wird seit 1981 intensiv an einer anwendungsorientierten Lösung für das Problem einer automatischen, begriffsorientierten Referate-Indexierung für das Information Retrieval gearbeitet (Projekte WAI, 1978–1981 und AIR, 1981–1983). Daraus entstand die vorliegende Arbeit, die sich wie folgt in drei Teile gliedert:
(1) Ein allgemeiner methodologischer Ansatz wird für die Entwicklung von Systemen dargestellt, die auf der Grundlage unvollständigen und unsicheren Wissens Entscheidungen zu treffen haben. Dazu wird eine Modellierung der Einsatzumgebung solcher Systeme verlangt, mit der die Entscheidungsaufgabe als das ‚Erkennen abstrakter Objekte' formuliert werden kann. Kennzeichnend für die Vorgehensweise ist eine klare Abgrenzung von Heuristik und Systematik, von modellbezogenem und probabilistischem Ansatz sowie von Konstruktions- und Testphase.
(2) Der programmatische Anspruch wird bei der Behandlung des Information Retrieval unterstrichen, wobei als Ziel gilt, die Inhaltserschließung auf die Beobachtung des Retrieval zu gründen. Damit wird der Beitrag illustriert, den der Ansatz zur Integration von Theorie, experimenteller Forschung und praxisorientierter Systementwicklung leistet.
(3) Unter den Rahmenbedingungen der Arbeiten der genannten Projekte WAI und AIR (das bedeutet: Coordinate Indexing, gebundenes Vokabular, englische Fachreferate aus den Fachgebieten Food Science and Technology und Physik) werden Ergebnisse auf breiter experimenteller Grundlage vorgestellt. Vom technischen Standpunkt aus handelt es sich dabei um die erfolgreiche Anwendung eines aus der Mustererkennung bekannten Verfahrens (Einsatz von Polynomklassifikatoren, die im quadratischen Mittel an eine extern vorgegebene Klasseneinteilung adaptiert sind). Ausgangspunkt sind wörterbuchabhängig generierte Entscheidungsgrundlagen, die auf Schätzungen dafür abzubilden sind, daß der jeweils betrachtete Deskriptor den Inhalt des Dokuments adäquat charakterisiert.

*) an dieser Einschätzung fühlt sich der Autor unschuldig; an die Stelle der Germanisten wären wohl die KI- und Informationswissenschaftler zu setzen.

35

# DER ERSTE STUDIEN- UND FORSCHUNGSFÜHRER KÜNSTLICHE INTELLIGENZ IST DA!

Der »Studien- und Forschungsführer Künstliche Intelligenz« (ca. 150 Seiten) gibt einen Überblick über ein Teilgebiet der Informatik, das in den USA und anderen Ländern bereits einen festen Platz in den Lehrplänen und in der Forschung gefunden hat und unter der Fachbezeichnung »Künstliche Intelligenz« (artificial intelligence) derzeitig einen enormen akademischen Aufschwung erlebt. Wegen der vielfältigen Anwendungsmöglichkeiten der wissenschaftlichen Resultate, steht dieses Gebiet heute auch weltweit im Zentrum des technologischen Interesses.

Der Studien- und Forschungsführer gibt eine vollständige (nach bestem Wissen) Übersicht über das deutsche Lehrangebot und die deutschen Forschungsaktivitäten zur künstlichen Intelligenz (KI).

Er enthält ebenso eine kurze Einführung in die KI, informiert über die wichtigsten internationalen wissenschaftlichen und organisatorischen Aktivitäten auf diesem Gebiet und gibt eine Übersicht über die amerikanische KI-Szene und über die der anderen europäischen Länder. Neben einer Literaturauswahl enthält er alle wesentlichen Kontaktadressen, Adressen der internationalen und nationalen KI-Organisationen usw.

**Erhältlich gegen einen Unkostenbeitrag von DM 7,– bei:**
**Gesellschaft für Informatik e.V.**
**Geschäftsstelle: Dr. Rampacher**
**Postfach 1669**
**5300 Bonn 1**
**Telefon 0228/376751**

**Klaus Heß / Jan Brustkern / Winfried Lenders**

## Maschinenlesbare deutsche Wörterbücher

Dokumentation, Vergleich, Integration

*VIII, 228 Seiten. Kart. DM 66.–. ISBN 3-484-31906-2*
*(Sprache und Information. Band 6)*

*Kurztext:* Dieser Band bringt den Abschlußbericht einer Untersuchung über maschinenlesbare Lexika des Deutschen. Er enthält Übersichten über die verschiedenen Lexika, synoptische Vergleiche, Begriffsbestimmungen sowie einen systematischen Teil über die optimale Struktur eines lexikalischen Eintrags.

*Interessenten:* Linguisten, insbesondere Germanisten, Lexikographen, Computerlinguisten; Dokumentare, Informationswissenschaftler, Übersetzer

Der Band enthält den Abschlußbericht über eine vom Bundesministerium für Forschung und Technologie geförderte Untersuchung maschinenlesbarer Lexika des Deutschen. Neben einer Begriffsbestimmung und einem systematischen Teil, der linguistische Literatur und Ansätze aus dem Bereich der Künstlichen Intelligenz-Forschung berücksichtigt, bietet er Übersichten über die wichtigsten maschinenlesbaren Lexika der Projekte CONDOR, JUDO, PASSAT, PLIDIS, SYSTRAN, HAM-RPM, LIMAS, BEAST-BACON, ferner über die Vorhaben von G. Wahrig (Deutsches Wörterbuch), des Sonderforschungsbereichs 100 in Saarbrücken (SUSY/SATAN) und über die maschinelle Erarbeitung des Wörterbuches von Lutz Mackensen.
Alle Übersichten orientieren sich an dem gleichen Darstellungsschema, welches eine Vergleichbarkeit erleichtert: Rolle des Lexikons im Gesamtsystem, Herkunft des Wortschatzes, Datenauswahl und -umfang sowie Mikro- und Makrostruktur des Lexikons, linguistische Beschreibungskategorien und linguistische Theorie. Darüberhinaus findet man gezielte Synopsen auf verschiedenen Ebenen: Zum Vergleich der Wortklassensysteme; zum Vergleich der morphologisch/syntaktischen Informationen; zum Vergleich der semantischen Informationen. Weiterhin wird geprüft, ob und inwieweit ausgewählte maschinenlesbare Lexika in ihrer Mikrostruktur mit vorliegenden DIN-Normen (MATER) verträglich sind.
Die vergleichende Analyse maschinenlesbarer Lexika, die in diesem Bericht vorgelegt wird, ermöglicht es, die unterschiedlich gebrauchten Terminologien, Lexikoninformationen und Grammatikmodelle in Beziehung zu setzen und so Ansätze für eine sinnvolle Anwendung dieser Lexika in den verschiedensten Bereichen der Verabeitung natürlichsprachlicher Daten zu entwickeln.

**Klaus Wothke**

## Automatisierung strukturalistischer Phonemanalysen

*X, 175 Seiten. Kart. DM 52.–. ISBN 3-484-31907-0*
*(Sprache und Information. Band 7)*

*Kurztext:* Das Buch behandelt die Möglichkeiten zur Formalisierung und Automatisierung von Phonemanalysen. Es wird gezeigt, wie mit Hilfe des Computers die Phonemsysteme von Sprachen ermittelt werden können.

*Interessenten:* Linguisten, insbesondere Phonologen, Dialektologen; Computerlinguisten

Der Band dokumentiert die Ergebnisse eines Projektes zur Automatisierung strukturalistischer Phonemanalysen. Hauptziele des Projektes waren die Algorithmisierung strukturalistischer Phonemanalysen und die Entwicklung eines leicht handhabbaren Computerprogramms, mit dessen Hilfe Linguisten Phonemanalysen durchführen können.
Die Arbeit geht aus von einer systematischen, kritischen und vergleichenden Darstellung der von Z. S. Harris und N. S. Trubetzkoy entwickelten strukturalistischen Konzeptionen der Phonemanalyse. Anschließend wird exemplarisch an einer Auswahl existierender Phonemanalysen untersucht, inwieweit diese Konzeptionen auch tatsächlich Eingang in die Praxis gefunden haben. Basierend auf diesen Betrachtungen wird eine Konzeption zur computerunterstützten Phonemanalyse entwickelt. Neuerungen dieser Konzeption gegenüber früheren Konzeptionen sind: die systematische Synthese distributioneller und funktionaler Kriterien bei der Klassifizierung von Phonen als Allophone bestimmter Phoneme, die weitgehende Formalisierung, die partielle Automatisierbarkeit.
Anschließend wird eine der Möglichkeiten zur Programmierung der automatisierbaren Ausschnitte der Konzeption behandelt. Das Programm wird dabei sowohl vom Programmiererstandpunkt als auch aus der Sicht des Benutzers, d. h. des Linguisten, dargestellt. U. a. wird die Automatisierung folgender Arbeitsschritte beschrieben: die Ermittlung der statistischen und der kontextuellen Distribution von Phonen, die Ermittlung der distributionellen Relation von Phonen, die Ermittlung von Minimalpaaren.
Die Leistungen des Programms werden eingehend am Beispiel seiner Anwendung auf ein Korpus der Landkölner Mundart veranschaulicht. Das Buch schließt mit einer Diskussion der Einschränkungen und Erweiterungsmöglichkeiten des Programmsystems und der ihm zugrundeliegenden Konzeption.

Harald H. Zimmermann, Edith Kroupa, Gerald Keil:

CTX - Ein Verfahren zur Computergestützten Texterschließung

(Zur Veröffentlichung in der BMFT-Reihe "Information und Dokumentation" - BMFT-FB-ID - eingereicht)

INHALT:
Im Rahmen der universitären Forschungsprojekte "JUDO" bzw. "JUDO-DS" wurde ein Computergestütztes TeXterschließungssystem (CTX) im Modell entwickelt und im Labor getestet.

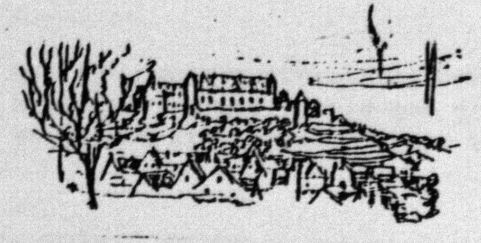
Allgemeine Zielsetzung dieses Forschungsvorhabens war die Erprobung des Einsatzes linguistischer Verfahren zur computergestützten Texterschließung für ein Information-Retrieval-System, getestet wurde am Beispiel verschiedenartiger Dokumente (Abstracts, Zeitungsartikel, Gesetzestexte und -entwürfe, Kommentare und Berichte) des Fachgebiets Datenschutzrecht.

Zum Abschluß der Forschungsprojekte wird ein Bericht vorgelegt, der die Motivation, die methodischen Grundlagen, die Verfahren und die ersten Ergebnisse darstellt. Der Überblick gliedert sich in zwei Teile: Einführend konzentriert sich die Beschreibung auf die Darstellung allgemeiner Grundlagen des Information Retrieval von textueller Information sowie eine benutzeradäquate Beschreibung der angewendeten und entwickelten Verfahrensbausteine in ihren charakteristischen Funktionen. Hierbei werden v.a. die Suchbaumtechnik und das angewendete formal-inhaltliche Stichwortverfahren gegenübergestellt.

Im zweiten Teil wird ein detaillierter Einblick in die einzelnen Verfahrensbausteine des wörterbuch-orientierten Texterschließungsverfahrens CTX gegeben, wobei der für eine konkrete Anwendung erforderliche allgemeine Aufwand anhand von Beispielen veranschaulicht wird. Ein wesentlicher Teil behandelt Struktur, Kodierung und Umfang der erforderlichen Lexika. Der Bericht schließt mit einer Bibliographie zur 'Automatischen Lemmatisierung' und zu 'Maschinellen Indexierungsverfahren'.

Gegenwärtig - nach Abschluß des Berichtszeitraums - wird in einem Folgeprojekt "Transfer informationslinguistischer Technologien" (TRANSIT) u.a. das System durch Einbeziehung erster konkreter Benutzer (z.B. in den Bereichen Patentwesen, Metallurgie und Sozialwissenschaften) optimiert und auf praktische Anwendungen ausgerichtet. Dabei werden sowohl das morphosyntaktische Verfahren (CTX-I) als auch - in der Anwendung "Sozialwissenschaftliche Literaturdokumentation" - das bedeutungsdifferenzierende Verfahren (CTX-II) getestet.

DESKRIPTOREN:
Information Retrieval, Suchbaum-Verfahren, Stichwort-Verfahren, Maschinelle Indexierung, Linguistische Analyse, Automatische Lemmatisierung, Deskriptorerstellung, Datenschutzrecht, Maschinelles Lexikon, Thesaurus, SUSY, CTX.



# Burg Stettenfels

Workshop on Language Generation

Burg Stettenfels - Untergruppenbach

Preliminary schedule

---

Mo 15.08.1983

9.00 Registration, transfers etc.

12.00 R. Gunzenhäuser
Universitaet Stuttgart
Germany

Welcome, General Introduction

13.00 Lunch and Aquaintance

14.30 K. Hanakata, J. Laubsch,
A. Lesniewski, D. Roesner

Presentation of SEMSYN Project

15.30 T.Hayashi
Fujitsu Laboratories
Kawasai, Japan

English Sentence Generation in ATLAS/U

16.30 Refreshments

17.00 G.Kempen
University Nijmegen
Netherlands

An Incremental Procedural Grammar (IPG)
for sentence formulation

18.00 H-D. Luckhardt
Universitaet des Saarlandes
Germany

Synthesis of Sentences from a Syntactico-Semantic
Deep Structure

19.00 W.Swartout
University of Southern California, ISI,
USA

Generation of NL Descriptions of Programs

Tu 16.08.1983

9.00 A.Joshi
University of Pennsylvania
USA

Grammatical Systems for Generation

10.00 K. de Smedt
University Nijmegen
Netherlands

Implementing an IPG-Generator
in an Applicative Object-Oriented
Language

11.00 Refreshments

11.30 E.Buchberger
Universitaet Wien
Austria

The Generation Component
of the system VIE-LANG

12.30 Lunch

14.30 S. Ishizaki
Electrotechnical Laboratory (ETL)
Ibaraki, Japan

Generation of Japanese Sentences from
Conceptual Representation

15.30 Refreshments

16.00 P.S.Jacobs
University of California
USA

Generation in the UNIX Consultant System

17.00 L.Danlos
University de Paris 7
France

On Intermediate Stages Needed
for Generation of Text

18.00 M.Yazdani
University of Exeter
England

Story Writing

20.00 Social Event
(eg. Vine, barbecue, ...)

---

Projekt SEMSYN
Institut fuer Informatik
Herdweg 51
D-7000 Stuttgart 1
West Germany

7101 Untergruppenbach
Telefon (07131) 70408

We 17.08.1983

9.00 D.McDonald
University of Massachusetts
Amherst, USA

Experiences with Real Time Generation -
An Overview of MUMBLE

10.00 R.P.Gabriel
Stanford University
USA

Deliberate Writing

11.00 Refreshments

11.30 D.E.Appelt
SRI International
USA

Recent Developments in Utterance Planning

12.30 Lunch

14.00 J.Kilbury
Technische Universitaet Berlin
Germany

A GPSG-Based Parser-Generator
in Prolog for German
(short)

14.30 H.Marburger & S.Busemann
Universitaet Hamburg
Germany

Generation in HAM-ANS

15.00 Final discussion

16.30 Refreshments

Transfers...

---

Please note:
To have enough time for discussions
talks should not exceed 45 minutes.

Veranstaltungen

ZENTRUM FÜR DATENVERARBEITUNG DER UNIVERSITÄT TÜBINGEN

Abteilung Literarische und
Dokumentarische Datenverarbeitung
Leiter: Dr. Wilhelm Ott

Tübingen, den 20. Juni 1983
Brunnenstraße 27
Tel. (07071) 29-2933/29-2901

# FERIENKURS

## TEXTDATEN-VERARBEITUNG FÜR ANFÄNGER

EINFÜHRUNG IN DIE SELBSTÄNDIGE BENUTZUNG DES TÜBINGER SYSTEMS VON TEXTVERARBEITUNGS-PROGRAMMEN TUSTEP:

EINGEBEN, KORRIGIEREN, VERGLEICHEN, AUSWÄHLEN, ÄNDERN VON TEXTEN;
RECHNEN MIT DEN IN DEN TEXTEN ENTHALTENEN ZAHLEN
(EINSCHL. KALENDERDATEN);
ZERLEGEN VON TEXTEN IN IHRE BESTANDTEILE;
SORTIEREN VON TEXT-EINHEITEN; INDEX-, REGISTER-ERSTELLUNG;
AUSGEBEN ÜBER ZEILENDRUCKER/TYPENRAD-DRUCKER/LICHTSATZ

(DIESE PROGRAMME STEHEN - AUSSER DEM SATZPROGRAMM - INZWISCHEN AUCH AUF DEN ANLAGEN DER UNIVERSITÄT ERLANGEN, KONSTANZ, MARBURG, REGENSBURG, TRIER, ULM UND WÜRZBURG SOWIE IM LEIBNIZ-RECHENZENTRUM MÜNCHEN ZUR VERFÜGUNG)

ORT: NEUPHILOLOGICUM, (WILHELMSTR. 50), RAUM 37
ZEIT: 10 TAGE, GANZTÄGIG, 19.-30. SEPT. 1983
BEGINN: MONTAG, 19. SEPT. 1983, 9.00 UHR

ANMELDESCHLUß FÜR TEILNEHMER, DIE NICHT DER UNIVERSITÄT TÜBINGEN ANGEHÖREN, IST DER 2.9.1983

GEZ. OTT

CALL FOR PAPERS

# RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL

July 2 - 6 1984 King's College, Cambridge

3RD JOINT BCS AND ACM SYMPOSIUM

CONFERENCE CHAIRMAN:
A Steven Pollitt
Huddersfield Polytechnic

PROGRAMME CHAIRMAN:
Keith van Rijsbergen
University College Dublin

LOCAL ARRANGEMENTS:
J Ken Moody
King's College Cambridge

TREASURER:
Peter Willett
Sheffield University

PROGRAMME COMMITTEE:

Tom Addis
Brunel University

Richard J Cichelli
American Newspaper Publishers Association

W Bruce Croft
University of Massachusetts

Douglas R McGregor
University of Strathclyde

A Patricia Miller
Marathon Oil Company Research Center

Najah Naffah
INRIA, Paris

Stephen E Robertson
City University

Hans J Schek
Technical University Darmstadt

Roger Tagg
Independent Consultant

Clement T Yu
University of Illinois at Chicago Circle

---

This is an international conference organised jointly by the specialist groups in information retrieval within the BCS and ACM. The two themes will be basic research ideas in information storage and retrieval and interesting new applications. Papers are invited on theory, methodology or application of information retrieval. Some suggested topics of interest are: Automatic Indexing, Natural Language Processing, Integrating IR with other Information Systems, IR in the context of Office Automation, System modelling and evaluation, User Interfaces, Complexity Problem, Cognitive Models, Knowledge-based IR Systems, Hardware design to enhance IR, File Organisation, Mathematical and Linguistic Models for IR.

You are invited to submit TEN copies of a detailed abstract (not a complete paper) to the programme chairman:

Professor C J Van Rijsbergen
Department of Computer Science
University College Dublin
Belfield, Dublin 4, Ireland (01) 693244

Submissions will be read by the appropriate members of the programme committee and evaluated on the basis of significance, originality and overall quality. Each abstract should:

a) contain enough information to enable the reviewers to identify the main contribution of the work;

b) explain the importance of the work - its novelty and its practical or theoretical relevance to information retrieval;

AND c) include comparisons with and references to relevant literature.

Abstracts should be no longer than ten double-spaced pages.

THE DEADLINE FOR SUBMISSION OF ABSTRACTS IS 1 DECEMBER 1983. Authors will be notified of acceptance or rejection by 15 February 1984. The accepted papers, typed on special forms, will be due at the programme chairman's address by 30 March 1984. It is intended that the Proceedings will be distributed at the conference, and will be subsequently available for general purchase.

**Termine**

29. 8.-17. 9.83: 1. Sommerschule der Deutschen Gesellschaft für Sprachwissenschaft (DGfS)
Konstanz
Rainer Baeuerle
SFB 99 - Universität Konstanz
Postfach 5560
D-7750 Konstanz

29. 8.- 1. 9.83: 6th Essen Symposium
New Trends in Electronic Publishing and Electronic Libraries
Essen
Universitätsbibliothek
Postfach 101454
D-4300 Essen

30. 8.- 2. 9.83: Meaning and the Lexicon; Second International Colloquium on the Interdisciplinary Study on the Semantics of Natural Language
Kleve, Niederrhein
"Meaning and the Lexicon"
Nijmegen Institute of Semantics
P.O. Box 1454
NL-6501 BL Nijmegen

1.September 1983 - Deadline
Seventh European Meeting on Cybernetics and System Research (EMCSR 84)
Wien
EMCSR 84-Sekretariat:
österreichische Studiengesellschaft für Kybernetik
Schottengasse 3
A-1010 Wien I
Tel.: (0222) 66 32 81 - 0, 63 61 12

1. 9.- 2. 9.83: Association for Computational Linguistics, European Chapter
Pisa
Harold Somers
Center for Computational Linguistics
U.M.I.S.T.
PO Box 88
Manchester M60 1QD
England

5. 9.- 7. 9.83: IRFIS 5 - Representation and Exchange of Knowledge as a Basis of Information Processes
Heidelberg
Dr. Hans Dietschmann
Coordinator IRFIS 5
Gesellschaft für Information und Dokumentation
Lyoner Straße 44-48
D-6000 Frankfurt/Main 71

9. 9.-12. 9.83: International Conference on Lexicography including an Ad Hoc Joint Committee Meeting of TERMINA/Inforterm/FID/AILA-Comterm
Exeter
Dr. R.R.K. Hartmann
Univ. of Exeter
The Language Center
Exeter EX4 4QH - UK

12. 9.-16. 9.83: Linguistisches Kolloquium
Bochum
H. Krenn, Romanisches Seminar
Ruhr-Universität Bochum
Postfach 102148
4630 Bochum

16. September 1983 - Deadline
Annual Technical Conference of the British Computer Society's Expert Systems Specialist Group
Cambridge
John Fox
Expert Systems 83
c/o Imperial Cancer Research Fund Laboratories
PO Box 123
Lincoln's Inn Fields
London WC2A 3PX

19. 9.-23. 9.83: GWAI 1983
Dassel/Solling
Prof. Dr. Bernd Neumann
FB Informatik der Universität Hamburg
Schlüterstr. 70
2000 Hamburg 13
Tel.: (040) 4123-4147/4152

19. 9.-30. 9.83: Ferienkurs Textdatenverarbeitung für Anfänger
Tübingen
Dr. W. Ott
Zentrum für Datenverarbeitung
Universtität Tübingen
Brunnenstr. 27
Tübingen

3.10.- 7.10.83: 13. Jahrestagung der GI
Hamburg
Prof. Dr. I. Kupka
FB Informatik der Universität Hamburg
Schlüterstr. 70
2000 Hamburg 13

4.10.- 7.10.83: Deutscher Dokumentartag
Göttingen
Deutsche Gesellschaft für Dokumentation (DGD)
Programme Chairman
Westendstr. 19
D-6000 Frannkfurt/Main

15.10.-20.10.83: Eusidic Annual Meeting
The Impact of the Electronic Office on Information Retrieval
Nizza
IMA
Mrs. H. Henderson
Uxedon Hill
Wembley, Middx. - UK

1. Dezember 1983 - Deadline
3rd Joint BCS and ACM Symposium: Research and Development in Information Retrieval
Cambridge
Prof. C.J. van Rijsbergen
Department of Computer Science
University College
Belfield, Dublin 4
IRELAND

17.11.-18.11.83: Document Preparation System - Computerunterstützte Verarbeitung von Text und Bild
**Heidelberg**
G. Rossbach
Springer-Verlag
Tiergartenstr. 17
6900 Heidelberg 1
Tel.:(06221)487-371

Themen: Interaktive Editier- und Formatiersysteme. Neue Drucktechnologien. Integration von Text und Bild. TeX/Scribe/TROFF - ein Vergleich. Integration Textverarbeitung - Textretrieval. Implementierung von Formatiersystemen. Die Schnittstelle der Druckerei zu Formatiersystemen. Entwicklungstrends im Satzmaschinenbau

14.12.-16.12.83: Annual Technical Conference of the British Computer Society's Expert Systems Specialist Group
**Cambridge**
John Lumley, Expert Systems 83
Cambridge Consultants, Science Park
Milton Road
Cambridge CB4 4DW

24. 4.-27. 4.84: Seventh European Meeting on Cybernetics and System Research (EMCSR 84)
**Wien**
EMCSR 84-Sekretariat:
österreichische Studiengesellschaft für Kybernetik
Schottengasse 3
A-1010 Wien I
Tel.: (0222) 66 32 81 - 0, 63 61 12

2. 7.- 6. 7.84: COLING-84
**Stanford**

2. 7.- 6. 7.84: 3rd Joint BCS and ACM Symposium: Research and Development in Information Retrieval
**Cambridge**
Prof. C.J. van Rijsbergen
Department of Computer Science
University College
Belfield, Dublin 4
IRELAND

24. 7.-27. 7.84: 9th Cranfield International Conference Mechanized Information Transfer
**Cranfield**
The Marketing Department
INSPEC, Station House
Nightingale Road
Hitchin, Herts SG5 1RJ - UK

8. 8.-10. 8.84: AILA World Congress 1984
(Association international de Linguistique appliquée)
**Brüssel**
University of Brüssel
AILA World Congress 84 - ITO/VUB
Pleinlaan 2
B-1050 Brüssel

# Presseschau

FÜNFTE COMPUTERGENERATION / Analyse des IW

## Japaner tüfteln emsig an künftigen Denkmaschinen

**mid DÜSSELDORF. Amerikaner und Europäer liegen im Wettlauf um eine bis zur Marktreife entwickelte „fünfte Computer-Generation" vorerst weit hinter der japanischen Elektronik-Industrie zurück. Dies stellt das Kölner Institut der deutschen Wirtschaft (IW) in einer Analyse zur Informationstechnologie fest.**

Bei der fünften Computer-Generation, den sogenannten intelligenten Systemen, handelt es sich um Anlagen, die ähnlich wie das menschliche Gehirn Daten und Informationen miteinander verknüpfen und daraus neue Erkenntnisse ableiten können. Davon erhoffen sich die Ingenieure eine Vereinfachung des Mensch-Maschine-Dialogs.

Die wichtigsten Elemente dieser in Japan bereits im Laborstadium befindlichen neuen Computer-Generation sind: sogenannte Ein- und Ausgabebausteine für Sprache, ein sich selbst organisierender Datenspeicher als „Knowledge basis", Hochgeschwindigkeitskomponenten, hohe Speicherfähigkeit und „eigene Lernfähigkeit". Das IW weist darauf hin, daß die japanische Industrie solche „Denkmaschinen" bis Mitte der 90er Jahre serienreif auf den informationstechnologischen Weltmarkt bringen will.

Zur Zeit wird in den USA, Japan und Europa intensiv an der „vierten Computer-Generation" gearbeitet. Hierbei sollen die Systeme eine zuverlässige Fähigkeit zur „Mustererkennung" und „Musterverarbeitung" erhalten, beispielsweise Handschriften, Bilder oder Sprache erkennen und verarbeiten. Die Produktentwicklung und Anwendung dieser Systeme steht in diesen 80er Jahren im Vordergrund.

### Tokio forciert die fünfte Generation

Das Institut meint, mit Prototypen dieser „denkenden" Computer-Generation sei zwar erst in den 90er Jahren zu rechnen, doch zeichne sich bereits heute ab, daß Japan auf diesem Gebiet einen erheblichen Entwicklungsvorsprung besitze. Wie keine andere Regierung habe die japanische die fünfte Computer-Generation gefördert und vorangetrieben. So habe Tokio 1982 ein eigenes Forschungsinstitut gegründet, das sich ausschließlich mit der Entwicklung der neuen Denkmaschinen befaßt — das „Institute for New Computer Technology" (ICOT). Dieser staatlich-industriellen Gemeinschaftseinrichtung, so das IW, steht allein für die Sachaufwendungen in den nächsten zehn Jahren umgerechnet eine Milliarde DM zur Verfügung. Das sei etwa der doppelte Betrag, den die US-Regierung für ähnliche Forschungszwecke bereitstelle.

Nach Feststellungen des IW gibt es in den USA und Europa noch kein Projekt, das ähnlich umfassend wie das japanische angelegt ist. Allenfalls auf Teilgebieten seien — beispielsweise bei IBM oder Bell — Forschungsanstrengungen erkennbar. Die Bundesrepublik Deutschland ist an dem Programm der Europäischen Gemeinschaft — „Grundlagen für ein europäisches strategisches Programm für Forschung und Entwicklung auf dem Gebiet der Informationstechnologie" (ESPRIT) — beteiligt, das sich mit der Entwicklung neuer Computersysteme befaßt.

Ebenso wie die EG-Kommission äußert das IW Zweifel daran, ob die wissenschaftlichen und technischen Aktivitäten der Gemeinschaft ausreichen, um den japanischen Vorsprung aufzuholen. Nach Auffassung der Kommission selbst liege das Nachhinken der Europäer vor allem darin, daß die Entwicklungsarbeiten zu sehr zersplittert seien.

Diese Ansicht teilt auch das IW und bemerkt, daß die bisherigen europäischen Anstrengungen auf diesem Gebiet jedenfalls nicht an mangelnder Ausstattung mit Forschungsgeldern gescheitert sein können. Zwar hätten die Japaner ihre staatlichen Ausgaben für Forschung und Entwicklung (FuE) in den letzten sieben Jahren auf umgerechnet 17,9 Mrd. DM im Jahre 1982 verdreifacht, doch habe die Bundesrepublik mit 18,7 Mrd. DM durchaus Schritt halten können

Auch bei der Intensität der Forschungsförderung schneidet die Bundesrepublik Deutschland nach Ansicht des Kölner Instituts nicht schlecht ab. Mit einem Anteil der staatlichen FuE-Aufwendungen am Bruttoinlandsprodukt von 1,2 % hätten die Deutschen 1981 deutlich vor den Japanern (0,7 %) gelegen.

Die entscheidende Gemeinschaftsaufgabe der nächsten Jahre besteht nach Auffassung des IW darin, über die ESPRIT-Ansätze hinaus in einer konzertierten technologischen Aktion die kapitalintensiven Forschungsprojekte im Bereich der anwendungsorientierten Informationstechnik zu bündeln. Ohne eine verstärkte europäische Zusammenarbeit werde Japan am Ende dieses Jahrhunderts wahrscheinlich Sieger im Wettlauf um die beste Position im Kreis der Informationsgesellschaften sein.

Wie sehr sich fast alle großen Industrieländer bereits merklich in Richtung Informationsgesellschaft entwickelt haben, läßt sich nach Meinung des IW vor allem an dem prognostizierten Strukturwandel im Beschäftigungssystem zeigen. Dazu führt das IW Beispiele an:

- In der Bundesrepublik Deutschland verrichteten 1980 knapp 40 % der Erwerbstätigen überwiegend sogenannte Informationstätigkeiten, 1950 waren es erst 19 %.
- In den USA lag der Anteil der Erwerbstätigen im Informationssektor 1980 sogar bei 46,8 % (1950: 30,8 %).

Handelsblatt
15./16.April 1983

Dennoch, so die IW-Analyse, zeichne sich gerade für die USA und Europa die Notwendigkeit ab, planend und forschend auf die fünfte Generation von Computern zuzugehen. Schwerpunkt müsse es dabei sein, die Handhabung vorhandenen Wissens, die Erarbeitung neuen Wissens, das Ziehen von Schlußfolgerung und die Fähigkeit, aus Erfahrungen zu lernen, maschinell zu realisieren.

# Looking for artificial intelligence

Through the 1960s and 1970s, the esoteric end of the computer industry was engaged on a quest for an elusive entity which came to be known as "artificial intelligence". It was not a wholly respectable pursuit – partly because of its associations with Frankensteinian monsters.

Another reason it was not taken too seriously was that only the researchers involved in it really believed it would ever come to anything. Unkind commentators referred to the "artificial intelligentsia". During the 1970s, in fact, a Science Research Council report (Lighthill) suggested that the whole thing was a waste of time, and work in Britain almost ceased.

But about 18 months ago the whole thing suddenly became respectable again – mainly as a result of the Japanese initiative to develop the so-called fifth generation of computers. This time, however, the researchers were careful to avoid anthropomorphic terms like artificial intelligence and spoke instead of knowledge engineering and expert systems.

Now, in the wake of yet another government report – say the Alvey Committee – the whole area of artificial intelligence is due for a revival in Britain. But this time it has pracical objectives rather than esoteric ones.

Export systems – or, as the Alvey report rather cumbersomely terms them, Intelligent Knowledge Based Systems (IKBS) – are aimed at capturing human expertise in a machine. This has profound implications on the way that computers will be used in the future and, of course, for the interface between people and machines.

Alex d'Agapeyeff, chairman of the British Computer Society's specialist group on expert systems, puts it this way: "It consists of taking the informal knowledge which people use to do their jobs and putting that into a computer as a set of rules. After all, the human brain is the best problem solving mechanism we know – why shouldn't we use it?"

He argues that computer systems have traditionally worked on the basis of mathematically oriented formulas, or algorithms in the accepted computer jargon. "But that is not necessarily the way that we do it," he says.

"Machines should do things the way humans do and explain the way they reach conclusions. It will make the computer a far better tool if humans can understand them." He suggested that more work could have been done by the specialists in artificial intelligence to explain the way such systems work. "But you don't get Phds in this country for explaining things simply."

The Government has recently sponsored a survey of the state of expert systems work in this country and d'Agapeyeff is assessing the results. He said it was too early to make any general observations.

One large commercial concern that has been looking in earnest at what expert systems can offer is ICI. Bernard Kelly of ICI has been experimenting with an expert system to help diagnose engineering problems in stress corrosion cracking. This involves assessin the properties of various metals when they come into contact with corrosive chemicals.

"We built the system to assess the technical feasibility of building this sort of knowledge into a computer and also as a demonstration to management that it would be useful.

"It has been very successful from both points of view and we have learnt some valuable lessons. For one thing we discovered that when you build a very large system the recognized ways of building your knowledge base are not always adequate. This has been a fault in the academic area where they have worked on smaller problems."

But Kelly sees the most important lessons coming from the way that an expert system fits into an organization. "ICI can see potential for expert system throughout its operations – anywhere there is a need for judgment. But we must take care to reorient the organization around the new systems – you cannot assume that you can introduce a new tool without altering the organization which uses it."

There were a number of identifiable benefits from expert systems, include a greater access to knowledge, the opportunity to combine two or more areas of expertise and a cross fertilization of knowledge between disciplines.

According to David Rodway, software technology director at Software Sciences, a major area where expert systems could have an impact is in the design of computer software itself. "We may be able to make use of our knowledge of designing systems to model new requirements. But we would have to ensure that we limit the problem. If we take the general case the number of rules we would need would become enormous."

Rodway disagrees with the view that Kelly puts forward on bulding large systems. "We have to limit ourselves to the smaller problems first to get an idea of the shape of the larger problems," he said. The other area of breakthrough that Rodway sees as important is in the interface between people and machines.

US expert Ed Feigenbaum recently commented that the first breakthrough would come in the area of consumer electronics in providing a so-called "user friendly" interface to computer systems. Rodway cited this example of the way that people were beginning to think about the practical applications of such systems.

Chris Naylor, the author of a forthcoming book on the subject, suggests that the whole area of expert systems has been surrounded by a false mystique. He has built his own expert system using an Apple computer and the popular novice programming language, Basic.

"An expert system is just another computer program," commented Naylor. "A simple program to calculate the area of a circle embodies a small amount of human expertise. What are referred to as expert systems nowadays are just more sophisticated versions of this same idea." Naylor's book, *Build Your Own Expert System* (Sigma Technical Press/John Wiley) serves as a good introduction to the subject and includes descriptions of most of the well-known expert systems in operation currently.

The idea of building human judgement into a machine raises important questions about the way we humans think about problems and can, at first, seem a little frightening. In the longer term, however, it could lead to a new style of computer system which is more forgiving than those we are used to.

If it leads to us being able to get the machines to do what we want with no fuss, then it should be worth it

**Philip Manchester**

The Times
22.März 1983

FAZ

Seite 16 / Mittwoch, 25. Mai 1983, Nr. 119

# Nachts — wenn der Übersetzer schläft . . .

## „Logos" — ein neues System zur reingetreuen Übersetzung / Von Axel Schnorbus

FRANKFURT, 24. Mai. Manche sprechen vom Zeitalter der Elektronik und meinen damit die Computer, von denen eine grundlegende Veränderung von Wirtschaft und Gesellschaft erwartet wird. Doch die Maschine allein bleibt dumm, wenn nicht Programme sie zum Laufen bringen. Das eigentlich Revolutionäre an der technisch-wissenschaftlichen Entwicklung heute ist daher die „Software", diese Abfolge logischer Schritte, entwickelt nach „Algorithmen", mathematischen Formeln oder besser gesagt: Rechenvorschriften, mit denen bestimmte Probleme gelöst werden sollen.

Seitdem die Computer dank der Mikroelektronik nun so leistungsfähig geworden sind und ständig noch leistungsfähiger werden, sind der Kreativität der Software-Entwickler kaum noch Grenzen gesetzt. Bereits 1946 war es amerikanischen Mathematikern gelungen, einen Algorithmus zu entwikkeln, um Sprache per Programm zu beschreiben. Doch alle Bemühungen in den darauffolgenden Jahren, von einer Sprache in die andere zu übersetzen, scheiterten am „Unvermögen" des Computers. Seine Speicherkapazität reichte damals für solche Aufgaben einfach nicht aus.

Heute werden für Sprachübersetzungen nicht einmal mehr große zentrale Rechner benötigt. Auf der Messe in Hannover ist jetzt zum erstenmal ein Übersetzungsprogramm vorgestellt worden, für das ein Kleincomputer vollkommen ausreicht. „Logos", eine amerikanische Entwicklung, läuft auf einem Textverarbeitungssystem der Firma Wang und übersetzt rund um die Uhr über 20 000 Wörter vom Deutschen ins Englische. Das sind über dreißig Seiten. Hans Georg Martin, Geschäftsführer der deutschen Tochtergesellschaft der Logos Computer Systems Inc., Bedford/Massachusetts, versichert, daß die Satzstellung fast immer stimme und daß nur zehn Prozent des übersetzten Textes vom „Humanübersetzer" nachredigiert werden müsse. Das Einzigartige an „Logos" ist für Martin die Fähigkeit des Programms, den menschlichen Gedankengang beim Übersetzen nachzuvollziehen und somit sinngetreu übersetzen zu können.

Das System besteht aus einem Grundwörterbuch, das einen Wortschatz von 25 000 deutschen Wörtern enthält; es ist durch Fachwörterbücher aus den verschiedensten Gebieten — insgesamt sind es 50 — erweiterbar. Der Benutzer von „Logos" kann zudem seine eigenen firmenspezifischen Begriffe speichern. Das ist wichtig. Denn erfahrungsgemäß braucht ein Übersetzer mehrere Monate, manchmal sogar Jahre, um sich in einem neuen Fachgebiet und in der „Begriffswelt" seines neuen Unternehmens zurechtzufinden. Diese Einarbeitungszeit entfällt künftig.

Das System verfügt über Sprachtabellen und Programme, die die Sprache nach den Regeln des Dudens analysieren. Alles, was „dudengerecht" geschrieben ist, wird von ihm auch richtig ins Englische übersetzt. „Logos" kann aber auch Fehler finden. Wenn es im deutschen Text beispielsweise heißt, eine Maschine soll „nach den Vereinigten Staaten" verschifft werden, korrigiert „Logos" und übersetzt richtig „in die Vereinigten Staaten". Aus schlechtem Deutsch kann freilich nie gutes Englisch werden. Die Übersetzung ist immer nur so gut wie die Vorlage. Wohl wird daran gearbeitet, das System künftig mit mehr „künstlicher Intelligenz" auszustatten, damit es sich Regeln und Gesetze merke und lernfähig werde. Aber bis dahin ist es noch ein langer Weg.

*Mit „Logos" soll sich die Übersetzungskapazität eines Unternehmens um mehr als das Dreifache erhöhen lassen. Eine einfache Rechnung wird dafür aufgemacht: Ein Übersetzer „kostet" je Zeile zwischen 3 und 7 DM. Für „Logos" werden je übersetzte Zeile (mit 55 Anschlägen) 55 Pfennig berechnet. Das Programm liegt bisher in der deutsch-englischen Version vor. Englisch/Deutsch soll noch in diesem Jahr angeboten werden. Übersetzungen aus und in das Französische und Spanische sind die nächsten Ziele. Logos wurde 1969 vom Linguisten Budd Scott gegründet. Bis 1979 war das Unternehmen ausschließlich für die amerikanische Regierung tätig.*

Schon heute merkt sich „Logos" jedoch neue Vokabeln. Das beginnt mit dem Übersetzungsvorgang: Der deutsche Text, im Computer abgespeichert, wird aufgerufen und vom Programm nach unbekannten Wörtern und nach Rechtschreibfehlern durchsucht. Am Bildschirm kann der Übersetzer sofort korrigieren und die fehlenden Wörter eingeben, wobei das System diesen — in einem Frage-und-Antwort-Spiel mit dem Übersetzer — gleich die richtigen Bedeutungen zuordnet. Denn es kommt auf die sinngetreue Übersetzung an. Jedem Wort sind daher Codes zugeordnet, die dessen semantische Bedeutung bestimmen. Insgesamt 256 verschiedene semantische Kategorien kennt „Logos". So wird das Wort „verlegen" stets richtig übersetzt, ob nun die Schlüssel „verlegt worden sind", ein Mädchen bei einer bestimmten Frage „verliegen worden ist", ein „Kabel verlegt" oder ein „Buch verlegt" worden ist.

Trotzdem ist das System weiter auf menschliche Hilfe angewiesen. Wenn „ein Verleger verlegen wird", muß es nämlich passen, weil ihm unbekannt ist, ob dieser nun „ein Buch verlegen" will oder ob ihn die knifflige Frage eines Konkurrenten „verlegen gemacht hat".

„Logos" wird sich daher wohl auch kaum für die Übersetzung von Literatur eignen. Brauchbar ist es hingegen, wenn es um die Übersetzungen von Bedienungsanleitungen, Handbüchern, Dokumentationen oder den Schriftverkehr geht, kurz; überall dort läßt es sich sinnvoll einsetzen, wo es auf die Sache, auf die Terminologie ankommt, nicht aber auf die Sprache selbst. Wo Mehrdeutigkeiten auftreten, wo die Sprache Ausdruck für Gefühle, Eindrücke, Zustände oder Reflexionen ist, wo das Einfühlungsvermögen und die Auslegung des Lesers gefordert sind, da ist „Logos" nicht geeignet.

Ob es künftig ein Programmsystem geben wird, das wie ein Mensch von einer Sprache in die andere übertragen könnte, ist zweifelhaft, zumindest heute noch. Sprachen sind hochkomplexe Gebilde. Sie sind zudem lebendig, entwickeln sich ständig weiter. In seinem Bemühen, Sprache zu analysieren, wird der Computer immer zurückbleiben, auch bei noch so viel „künstlicher Intelligenz". Die Lernfähigkeit des Computers kann durch Programme weiter verstärkt werden, aber er wird auch mit noch so viel Aufwand menschlicher Intelligenz nicht imstande sein, selbst schöpferisch tätig zu werden, selbst Regeln zu entwickeln. Das ist beruhigend.

Auch „Logos" wird daher immer auf die Hilfe des Menschen angewiesen bleiben. Seine Schöpfer verstehen dieses Programm daher auch als ein „Instrument in der Hand des Übersetzers". Während „Logos" zeitraubende Routinearbeiten ausführt, bleibt dem Übersetzer Zeit für jene Aufgaben, die seine Sprachkompetenz, seine Urteilsfähigkeit erfordern. Ohnehin wird die Maschine einen technischen, mit vielen Fachausdrücken gespickten Routinetext von 200 Seiten immer genauer, vor allem einheitlicher und damit zuverlässiger und zudem ungleich schneller übersetzen können — und das auch nachts, wenn der Übersetzer schläft. Die Rohübersetzung wird dann im System abgespeichert und morgens vom Übersetzer am Bildschirm überarbeitet.

Müssen die Übersetzer damit um ihre Jobs bangen? Wohl kaum. Solche Programme werden ihre Arbeit freilich produktiver machen. Man wird künftig auch Texte übertragen können, die aus Gründen der Wirtschaftlichkeit bisher unübersetzt geblieben sind, neue Fachaufsätze etwa, zur schnellen, aktuellen Information der Mitarbeiter. Der technische Fortschritt führt längst nicht immer zur Einsparung von Arbeitsplätzen, er dient sogar in erster Linie — wie die kurze Geschichte der Datenverarbeitung immer wieder zeigt — zur Steigerung von Produktivität und Effizienz und damit auch der Wettbewerbsfähigkeit.

# The Campus Race For 'Computer Literacy'

By James Lardner

THE BIGGEST thing in American higher education these days is the computer. On campus after campus, in department after department, the drive is on to gear up for the computer revolution.

A few trailblazing colleges — that's how they see themselves, at any rate — have decided that every student should have his own computer. Carnegie-Mellon University recently announced a joint plan with IBM to design a computer network with individual work stations for CMU's 5,500 students, each "a lot more powerful" than the IBM Personal Computer. IBM will be paying the development costs — roughly $20 million in software alone — for a system that promises to make CMU a "prototype for every major university in the country", according to president Richard M. Cyert.

All 7,600 students at Philadelphia's Drexel University will have computers as of the 1983-84 school year, and Martha B. Montgomery, one of the prime movers behind the plan, says: "I think it will be standard practice and I think it's coming very fast."

Another leader is Rensselaer Polytechnic Institute in Troy, N.Y., which recruits outstanding science and engineering students "pretty much the way Notre Dame recruits football players", according to James L. Moss, director of computing services.

For the last three years, Rensalaer has been giving personal computers to a group of 20 top freshmen. (Last year the model chosen was an Atari 800; this year it is the IBM Personal Computer.) The school has been monitoring the effects of its experiment on grades, social activities and study habits, and hopes to reach conclusions about "all of the social and academic implications of personal computer ownership", according to Moss.

So far, the schools that have gone this route are ones that emphasize technical and vocational subjects. But "computer literacy" has become the goal of the moment on many more broadly-focused campuses, and there are signs that the humanities and the social sciences may, in the end, take the new technology to their breasts just as fervently as the hard sciences.

"The computer can be used in almost every course in a broad curriculum," says Fred Siff of George Mason University, which is about to launch a two-year general-education experiment with 300 members of each incoming class taking a common group of 12 courses — all, if the planners have their way, involving computers.

At Drexel, Montgomery says the department of humanities and communications, which she heads, will use the new computers most, principally for word-processing. "The obvious first clear advantage of the word processor," she explains, "is that it makes revision so much easier. It takes some of the drudgery out so that students are more willing to edit their own work, and hardly anything exists that isn't improved by editing . . . [and] when it's easy to correct your mistakes, you can be a little more playful with language."

The academic world may associate computers with technical fields, she says, but "where an invention begins and where it ultimately leads are very different places. The real reason that the humanities have to take some leadership in this is that we can't expect the electrical engineers to know what the humanistic concerns are that computers can serve . . . We've got to get a critical mass of humanists who know about these things".

"The innovators on campus, the gatekeepers, may appear in any department," says Andrew R. Molnar, a computer education specialist with the National Science Foundation. Certain schools, he points out, have established excellent reputations for their use of computers within particular fields without being across-the-board leaders in computing.

At the University of Chicago, a vast library of French literature has been stored and filed in the university's computer system for linguistic analysis and other purposes. The University of Iowa, as an experiment, has put a terminal on the desk of every faculty member in political science — for word-processing, data and communications. And a unique "instructional design group" at the University of Minnesota creates computerized programs for faculty members in all fields.

Among the results is a line of software law-school instruction programs (to be marketed in partnership with Harvard) that enables students to have Socratic dialogues with their computer terminals. "It's the Kingsfield-in-'The-Paper-Chase' sort of dialogue except it's done through the computer program," explains Russell Burris, professor of law and instructional psychology at Minnesota.

Computer use is one area where the oldest and best-endowed schools are not invariably the leaders. In Siff's subjective opinion, George Mason's approach compares favorably with the much-vaunted "core curriculum" program unveiled four years ago by Harvard — an initiative many critics found disappointingly traditional and whose lack of daring was widely attributed to the strong attachment of Harvard's well-entrenched faculty to established courses and specialities.

George Mason's new program represents a major break from the past, according to Siff, and he attributes his school's willingness to make that break in part to the fact that it is only 10 years old, with a correspondingly small number of tenured faculty members of middle years and beyond.

Size is also one of the factors that govern a school's flexibility.

"The average public institution is lucky if it can give 10 hours [of computer time] per student per year," says Robert Gillespie, vice provost for computing at the University of Washington, whose student body numbers 33,000.

By contrast, Gillespie cites the example of Dartmouth, where the average student gets upward of 50 hours of terminal use a year. At Dartmouth, 95 percent of students get "hands on" experience and 75 per cent are regarded as "computer literate", according to a recent survey. There are 600 terminals on campus for 4,000 undergraduates and 900 graduate students.

As fast as many colleges are moving, it is not fast enough to suit all. "Computer literacy" is a goal that, in many minds, comes chained to a threat — the threat of rampant "computer illiteracy" in an age when, professionally speaking, that could be fatal.

"We may be creating an underclass of people who are not only unemployable but uneducable," says the NSF's Molnar, "because they complete the school system without having had any experience with computing". Because of computers, Molnar says, there has been an "information explosion" rendering the body of documents and data in almost every field so vast that "nobody can 'know knowledge' any more, they can only know where to look". And "knowing where to look" means knowing how to use computers.

"In Thomas Jefferson's time, 3,000 books was a library," he says. "Now we're talking about millions of documents such that it takes about 23 years to read all of the things you'd need to know in chemistry alone." In economics, astronomy, genetic sciences and other fields, schools with inadequate computer facilities, and the students who go there, "are out of big science, are precluded from doing cutting-edge research," says Molnar. And the country as a whole, he warns, runs the risk of falling behind such nations as France and Japan, which are investing in computer instruction "at a rate that only centralized systems can do".

Unfortunately, the academic world is widely perceived as slow-footed by its nature — a tortoise racing the hare of American commerce. Teachers trained and tenured in a computer-free age are unfit, it is argued, to decide what the next generation of college graduates ought to know, or, even if they perceive the need correctly, to act on that perception.

"When I started in school," recalls Robert Gillespie, "I bought a slide rule that cost me about the price of 10 books, and I have a theorem that says when the price of a computer for word-processing comes down to the price of 10 books, everyone will buy one. And that's not so far away."

© 1983 The Washington Post Co. All rights reserved.

30. Januar 1983

SPRACHÜBERSETZUNG / 25 000 Begriffe

# Logos für Deutsch/Englisch

HANDELSBLATT, Mittwoch, 13.4.1983

lp HANNOVER. Logos Computer Systems Deutschland GmbH, Frankfurt, als Tochter der amerikanischen Logos-Gesellschaft seit Anfang 1983 in der Bundesrepublik tätig, äußerte sich in Hannover recht optimistisch, im Alleingang die Wirtschaft in der Bundesrepublik für ihr Übersetzungssystem vom Deutschen ins Englische zu gewinnen. Zwei führende Computerhersteller in der Bundesrepublik gehören bereits zu den Kunden. Bei den Banken bestehe großes Interesse. Logos liefert nur die Software. Übersetzt wird mit Computern von Wang, die auf eigene Rechnung die Hardware liefert.

Das Software-System kostet 24 000 DM und umfaßt die eingehende Schulung des Kundenpersonals sowie freie Benutzung für drei Monate. Danach müssen knapp 5000 DM pro Monat Miete gezahlt werden. Das „elektronische Wörterbuch" umfaßt 25 000 Begriffe. Für die verschiedenen Wirtschaftszweige wird dazu Software für Spezialausdrücke geliefert. Die amerikanische Logos ist weitgehend für die Regierung tätig. Angefangen hat man seinerzeit mit vietnamesisch und Farsi, der Sprache für den Iran.

Handelsblatt
14. April 1983

BMFT-*Journal* · Seite 9

Nr. 2 · März 1983

# Halbleitertechnik verbessert Computer-Sprachverarbeitung

Dies ist das Ergebnis einer abschließenden Diskussion von Experten anläßlich eines Workshops zum Thema Sprachverarbeitung, der von der Gesellschaft für Mathematik und Datenverarbeitung als Projektträger für die Förderung der Informationsverarbeitung veranstaltet wurde. Vor den Fachleuten aus dem gesamten Bundesgebiet standen die vom Bundesminister für Forschung und Technologie geförderten Projekte zur Sprachverarbeitung zur kritischen Diskussion.

Bislang gehörten sprachverstehende Maschinen in die Welt der Science-fiction oder in die Umgebung aufwendig ausgestatteter Labors. Selbst bescheidene Ansätze lagen wegen der hohen Kosten fern jeglicher kommerzieller Realisierung. Mit den Fortschritten in der Halbleitertechnik und dem damit verbundenen Preisverfall sind einige Barrieren für den kommerziellen Einsatz von einfachen Sprachverarbeitungsgeräten weggefallen. So finden heute Geräte zur Ausgabe synthetischer Sprache oder solche, die Einzelwörter sprecherabhängig bei geringem Wortschatz erkennen, ihre Anwendung in den unterschiedlichsten Bereichen von der Fertigung und Lagerhaltung bis zum Konsumartikel.

In Erwartung eines sich öffnenden Marktes, der im Gefolge der heute angebotenen einfachen Geräte bald leistungsfähigere Produkte verlangt, haben viele Firmen ihre Forschung auf dem Gebiet der Sprachverarbeitung verstärkt. Auch in der Bundesrepublik sind die Grundlagenarbeiten auf diesem Gebiet intensiviert worden, nicht zuletzt dank der BMFT-Förderung, die seit etwa 3 Jahren im Rahmen der Informationsverarbeitungsförderung gewährt wird. Für die Zukunft gilt es vor allem, die Brücke zwischen den mehr signalorientierten zu den linguistischen Verfahren zu schlagen. Für beide Ansätze sind gute Vorarbeiten geleistet.

Vom Stand der Forschung und Entwicklung auf dem Gebiet der Sprachverarbeitung in der Bundesrepublik konnten sich die Experten auf dem Workshop selbst überzeugen.

Es bestand Übereinstimmung darin, daß angesichts des beschränkten Forschungspotentials auf diesem Gebiet in der Bundesrepublik die Zusammenarbeit der Forschungsgruppen verstärkt werden müsse und daß hier die BMFT-Förderung unterstützend wirken könne.

Die Zusammenarbeit der Firmen Philips und Siemens auf dem Gebiet der automatischen Sprachverarbeitung ist in diesem Sinne ein Schritt in die richtige Richtung und schafft eine günstige Konstellation für ein erfolgreiches Vorgehen.

# Der sprechende Computer lernt immer mehr hinz

## Die elektronische Sprachverarbeitung ist vielseitig einsetzbar / Wissenschaftler warnen vor Mißbrauch

Mit den elektronischen Wörterbüchern, wie sie seit kurzem angeboten werden, haben sie nur wenig gemein: die Computersysteme zur maschinellen Übersetzung natürlicher Sprachen, an denen in zahlreichen Forschungsinstituten gearbeitet wird. Auf der 5. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS) in Passau wurde vor kurzem von dem Mathematiker Hans-Dieter Maas ein schon etwas betagtes, aber noch immer hochinteressantes Projekt vorgestellt: SUSY, das „Saarbrückener Übersetzungs-System".

Dieses System, entwickelt im Rahmen des Sonderforschungsbereiches „Elektronische Sprachforschung" der Universität des Saarlandes, hat die maschinelle Übersetzung von deutschen, englischen und französischen Texten zum Ziel. Dazu sind zwar umfangreiche „Wörterbücher" nötig: das Deutsch-Lexikon von SUSY zählt 136 000 Einträge, auch für moderne Computer eine ganze Menge. Die eigentlichen Schwierigkeiten beginnen jedoch mit der genauen Bedeutungsbeschreibung dieser Wörter, die ja möglichst unabhängig von einer bestimmten Sprache sein soll.

Ein Problem ist beispielsweise die Mehrdeutigkeit der Wörter auch in ihrer grammatischen Verwendbarkeit: ein deutsches Wort ist beispielsweise im Schnitt 1,76-deutig, ein englisches sogar 2,3-deutig. Im Satzzusammenhang multiplizieren sich diese Mehrdeutigkeiten, und so bringen es selbst relativ unscheinbare Sätzchen rasch zu Millionen verschiedener Lesarten. Durch die Satzbauregeln wird diese Vieldeutigkeit zwar stark reduziert, aber oft genug bleiben doch mehrere Bedeutungen übrig.

SUSY bewältigt etwa 25 000 Wörter pro Stunde; andere Systeme, wie SYSTRAM in München, sind bereits zehnmal so schnell. Die Güte der Übersetzungen läßt allerdings noch viel zu wünschen übrig: im allgemeinen sind sie nur als Roh-Übersetzungen für einen menschlichen Bearbeiter zu gebrauchen. Immerhin sind die so übersetzten Texte bereits viermal billiger als bei der Bearbeitung durch einen menschlichen Übersetzer allein. Die Stärke der maschinellen Übersetzung liegt bei Fachtexten wie zum Beispiel Gebrauchsanweisungen; hierzu arbeitet man mit dem Bundessprachenamt in Hürth bei Hannover in einem Pilotprojekt zusammen. Ein weiterer Projektpartner ist das Patentamt in München. Die mit SUSY gewonnenen Erfahrungen sollen schließlich in EUROTRA einfließen, einem internationalen Projekt zur maschinellen Übersetzung zwischen den Sprachen der EG-Länder. Der Prototyp soll in fünf Jahren fertig sein, die Entwicklungskosten sind auf 50 Millionen Mark veranschlagt.

Computer können jedoch mehr, als menschliche Sprache nur zu übersetzen. An der Forschungsstelle für Informationswissenschaft und Künstliche Intelligenz der Universität Hamburg versucht man, ihnen ein Verständnis der natürlichen Sprache beizubringen. Wolfgang Hoeppner stellte auf der DGfS-Jahrestagung das Projekt HAM-ANS (Hamburger anwendungsorientiertes natürlichsprachliches System) vor. Der Vorteil der natürlichsprachlichen Kommunikation mit dem Computer liegt vor allem darin, daß man nicht eigens eine künstliche Programmiersprache zu lernen hat. Der Kreis der möglichen Benutzer wird damit entscheidend erweitert, die Verständigung mit dem Computer insgesamt flexibler.

Zur Entwicklung von HAM-ANS arbeitet man mit sehr alltagsnahen „Diskursbereichen", die unterschiedliche Anforderungen an das System stellen. Beispielsweise hat man umfangreiche Informationen über Fischerei-Expeditionen im Südpolargebiet gespeichert. Hier geht es vor allem darum, Methoden zu entwickeln, um möglichst unkompliziert an die gewünschten Informationen in einer großen Datenbank heranzukommen (was in der Informatik als das „Retrieval"-Problem bekannt ist). Ein anderer Diskursbereich ist die Verkehrsüberwachung mit Hilfe von Kameras an einer Kreuzung. Hier müssen die bewegten Objekte zunächst mit einem Bildfolgen-Analysesystem erkannt und anschließend in einer „Modellwelt" repräsentiert werden. Ein Problem liegt hier in der Wiedergabe von zeitlich sich ändernden Sachverhalten und von Bewegungsverben.

Ein dritter Diskursbereich ist die Hotelreservierung. Das System nimmt dabei die Rolle des Hotelportiers ein, der Auskünfte erteilt und Buchungen tätigt. Hier sollen dem Computer vor allem die Regeln der Gesprächsführung beigebracht werden. Die Hotelreservierungs-Situation ist dafür besonders geeignet, weil sie stark standardisiert ist und in ihr sowohl die Ziele des Gastes wie des Portiers festgelegt sind. So können die für Sprachverstehen und Sprachproduktion wichtigen Einschätzungen des Gesprächspartners beispielshaft untersucht werden.

Obwohl HAM-ANS' Leistungen, verglichen mit denen eines Menschen, noch bescheiden anmuten, benötigt er eine umfangreiche Software. die eingegebenen Sätze werden zunächst mit Hilfe eines Lexikons und geeigneter Wort- und Satzbauregeln in eine maschinengerechte „Repräsentationssprache" übersetzt. Diese Sprache „versteht" das System mit Hilfe eines sogenannten „semantischen Netzwerks", in dem beispielsweise gespeichert ist, ob ein Begriff als Unterbegriff eines anderen zählt. Zusätzlich benötigt man noch ein weiteres Netzwerk, mit dem der Computer sich merkt, über was bisher bereits die Rede war. Nur so kann es überhaupt zu einem sinnvollen, zusammenhängenden Dialog kommen. Darüber hinaus muß das System Zugang zur Datenbank haben, es muß seine Entscheidungen auf Anfrage begründen können, es muß lernen, selbst die Gesprächsinitiative zu ergreifen — bei Unklarheiten oder um den Gast zur Zimmerbuchung zu drängen.

Sowohl SUSY. als auch HAM-ANS sind auf die Dateneingabe mit Hilfe einer Tastatur angewiesen. Vor allem auch wirtschaftlich interessant ist es aber, dem Computer das Verstehen gesprochener Sprache beizubringen. Dies ist freilich technisch außergewöhnlich schwierig, weit schwieriger als die automatische Sprachproduktion. Zwar werden bereits seit zehn Jahren Spracherkennungsgeräte kommerziell vertrieben, doch diese meistern nur einen kleinen Wortschatz und, wenn überhaupt, nur wenige grammatische Regeln. Zudem müssen sie vorher auf die Stimme des jeweiligen Benutzers individuell getrimmt werden und fordern ihm eine sehr sorgfältige Artikulation ab. An zahlreichen Instituten und Firmen wird deshalb an einer natürlicheren Spracherkennung gearbeitet, unter anderem auch an der Universität Erlangen-Nürnberg.

Das von dem Informatiker H. Niemann geleitete „Erlanger System zur Erkennung kontinuierlicher Sprache" will Erfahrungen mit einem Modell der Intercity-Zugauskunft gewinnen. Die akustischen Eingabedaten sind dabei auf die Telefon-Bandbreite reduziert, um von Anfang an möglichst anwendungsnahe Bedingungen zu setzen. Vorerst will man sich auf 1380 Wörter und 225 grammatische Regeln beschränken.

Um ein Wort zu identifizieren, wird das Sprachsignal alle 20 Millisekunden auf bestimmte Merkmale hin analysiert — beispielsweise daraufhin, ob das Signal gerade stimmhaft wie bei den Vokalen ist oder Reibelaut-Qualitäten wie bei f, s oder ch aufweist. Diese 20-Millisekunden-Segmente werden dann einzelnen Lauttypen zugeordnet. Anschließend wird das so analysierte Wort mit einem eingespeicherten Lexikon verglichen, in dem die dudengemäße Standardaussprache der Wörter vorliegt. In der Regel bietet der Akustik-Analysator jedoch mehrere Alternativen an, da noch nicht alle Laute zuverlässig identifiziert werden können, die Lautqualität von Sprecher zu Sprecher stark schwankt und die unterschiedliche Sprechgeschwindigkeit für weitere Zuordnungsschwierigkeiten sorgt.

Nicht nur an diesem Beispiel zeigt sich, daß die elektronische Sprachverarbeitung auf der einen und die Sprachwissenschaft und Sprachpsychologie auf der anderen Seite sich gegenseitig befruchten könnten. Bislang allerdings ließ die Verständigung zwischen beiden Bereichen vieles zu wünschen übrig — wegen der verschiedenen Denkweisen, aber auch wegen der unterschiedlichen Theorieformulierungen, welche die Forschungsziele und -Ergebnisse der jeweiligen Gegenseite undurchschaubar machten. Ein wichtiger Schritt zur Verbesserung des Informationsaustausches ist möglicherweise die neue Programmiersprache COMSKEE, die in Saarbrücken entwickelt wurde und auf die Bedürfnisse von Linguisten zugeschnitten ist.

Bei der abschließenden Diskussion auf der Passauer Tagung ging es jedoch nicht nur um theoretische Verständigungsschwierigkeiten, sondern vor allem um den Mißbrauch, der mit der elektronischen Sprachverarbeitung getrieben werden könnte. Insbesondere die Erlanger Forschungen könnten, so warnten einige Wissenschaftler, den Grundstein für eine effektive Telefonüberwachung großen Stils legen: Gespräche, die sich durch bestimmte Schlüsselwörter „verdächtig" machen, könnten automatisch aufgenommen und dann von Beamten näher untersucht werden. Natürlich bestritten die Referenten, an solchen Projekten zu arbeiten; es wäre jedoch nicht zum ersten Mal, daß Grundlagenforschungen den Weg zu ungewollten Anwendungen bereiten. Es ist gut möglich, daß nach Physik, Chemie und Genetik nun auch die Sprachwissenschaft ihre Unschuld verliert — eine Aussicht, die den Sprachwissenschaftlern trotz der damit verbundenen Beschäftigungsmöglichkeiten keineswegs behagte.

MANFRED KRIFKA

Frankfurter Rundschau 28. 4. 1983

# Mitteilungen aus der GLDV

## Neues aus dem Vorstand

Vorstandssitzungen vom 3. März 1983 in Trier
und vom 6. Mai 1983 in Frankfurt

---

### o Neue Mitglieder

Neu in die Gesellschaft für Linguistische Datenverarbeitung (GLDV) wurden aufgenommen:

Maximilian Herberger, Hartmut Warlich, Dr. Peter Scherber, Karin Haenelt, Heinz Schay, Beatrice Bagola, Holger Bagola, Ute Hafner, Armin Rau, Martin Schweizer, Prof. Dr. J. P. Köster, Dr. Manfred Thiel, Hans-Joachim Dezelski, Nico Weber, Rainer Koch, Claudia Toussaint, Guido Wirtz, Gregor Büchel, Hans-Peter Wolf, Stephan Mehl, Prof. Dr. Erhard Mergenthaler.

Der Verein zählt damit zum 6. Mai 1983 105 Mitglieder.

### o Werbebroschüre

Die Vorlage der Werbebroschüre wurde mit geringfügigen Änderungen einstimmig angenommen. Die Broschüre wird in einer Auflage von 5000 Stück als Faltblatt hergestellt.

### o Tagungsbände 1983

Der Tagungsband "Microcomputer" des Kolloquiums in Trier ist bis auf drei Beiträge fertiggestellt (Stand: 6.5.1983). Für die Herstellung liegt die Zusage eines Verlages (Buske) vor. Der Band wird voraussichtlich im Herbst 1983 erscheinen.

Die Beiträge zur CUI erscheinen als Beiheft zum LDV-Forum.

o Jahrestagung 1984

Die nächste Jahrestagung ist für Ende Februar 1984 in Heidelberg in Kooperation zwischen der Universität Heidelberg und dem Wissenschaftlichen Zentrum der IBM geplant. Zur Vorbereitung der Tagung erfolgt ein Call for Papers mit dem Thema "Trends in der linguistischen Datenverarbeitung".

o Beirat

Die Herren Wahlster und Lehmann haben die Wahl in den Beirat angenommen. Der Beirat ist aufgefordert, sich zu konstituieren.

o Sonstiges

Die auf der Mitgliederversammlung in Trier beschlossene Satzungsänderung ist in die Wege geleitet worden.

Die Ausstellung von Mitgliedsausweisen wurde vom Vorstand einstimmig abgelehnt mit dem Hinweis darauf, daß als Beleg für die Steuererklärung die Quittung über den bezahlten Mitgliedsbeitrag ausreicht.

**Unbekannt verzogen sind:**

**Ellen Winkler, Gomaringen**
**H.U. Block, Aachen**
**Manfred Gehrke, Duisburg**

**Wer kennt die neue Anschrift?**

Protokoll der Mitgliederversammlung vom 03. März 1983
in Trier von 15.15 bis 19.20

Top 1: Regularia

Der erste Vorsitzende Krause eröffnete die Versammlung und stellte fest, daß die Mitgliederversammlung mit 37 anwesenden Mitgliedern beschlußfähig war. Die 6 anwesenden Gäste wurden zur Teilnahme an der Mitgliederversammlung zugelassen.

Der Versammlung lagen 4 Stimmübertragungen zu TOP 8 vor.
Als Kassenprüfer für den Kassenbericht 1982 wurden ohne Gegenstimme G. Knorz und E. Kroupa bestimmt.
H.-D. Lutz wurde ohne Gegenstimme zum Wahlleiter für die Neuwahl des Vorstands gewählt.

Die folgende Tagesordnung wurde einstimmig beschlossen:

1. Regularia
2. Bericht des Vorstands
3. Kassenbericht
4. Entlastung des Vorstands
5. Neuwahl des Vorstands
6. Bericht des Beirates
7. Neuwahl des Beirates
8. Vorstandsantrag auf Namensänderung:
   (§1: Name der Vereinigung):
   Der Verein LDV-Fittings e.V. möge fortan den Namen "Gesellschaft für Linguistische Datenverarbeitung e.V. (GLD)-Wissenschaftliche Vereinigung zur Förderung der maschinellen Verarbeitung natürlicher Sprache" führen.
9. Vortrag von Frau Endres-Niggemeyer über Fördermöglichkeiten im Bereich LDV
10. Planung 1983/84
11. Verschiedenes

TOP 2: Bericht des Vorstands

Der 1.Vorsitzende J. Krause gab den Rechenschaftsbericht über die Aktivitäten im Zeitraum 1981/82.

1981 und 1982 fanden je 4 Vorstandssitzungen statt.

Ziel des Vereins ist es, eine Standortbestimmung für die LDV zu finden und mit Nachbarorganisationen zu kooperieren; dies wurde v.a. durch Tagungen und Workshops angestrebt.

Im Berichtszeitraum fanden folgende Veranstaltungen statt:

Tagungen, Workshops...

| | |
|---|---|
| Dez. 1980 | Kolloquium "LDV-Ausbildung und Berufsperspektiven". Saarbrücken |
| März 1981 | Workshop "Datenbanken", Sektion: Datenbanken und natürliche Sprache Kooperation mit GI. Heidelberg |
| April 1981 | Expertengespräch "Automatisches Indexieren" Kooperation GID + KLD (DGD). Frankfurt |
| Nov. 1981 | Workshop "Philologische Anwendungen der LDV". Tübingen |
| Februar 1982 | LDV-Kolloquium "LDV und Nachbarn". Koblenz |

März 1983 LDV-Kolloquium 1983
-Mikrocomputer/Textverarbeitung
-Computerunterstützte Inhaltsanalyse
-Maschinelle Übersetzung
Kooperation Sozialwissenschaften.

Die Ergebnisse der Aktivitäten fanden Niederschlag in folgenden Publikationen:

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

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

IZ Sozialwissenschaften (Bearbeiter K.Petersen): Software zur linguistischen Datenverarbeitung. Sonderdruck für die Mitglieder der LDV-fittings. Bonn 1983

LDV-fittings INFO:
1981: 4x
1982: 4x

Arbeitspapiere der Arbeitskreise

An Arbeitskreisen waren mit unterschiedlichem Erfolg ins Leben gerufen:



AK Ausbildung und Berufsperspektiven
(Studiengägne mit LDV-Anteil)

AK Phonetik (Diestelmann)

AK Evaluierung natürlichsprachlicher Komponenten von Informationssystemen (beendet 1982).

AK LDV und Nachbarn

AK Mikros und Textverarbeitung

AK Sozialwissenschaften und LDV

Die beiden letztgenannten Arbeitskreise hatten die Vorbereitung für das diesjährige LDV-Fittins-Kolloquium übernommen. Der Vorsitzende wies darauf hin, daß sich die Vorbereitung der Kolloquien durch die Arbeitskreise bewährt habe.

Top 3: Kassenbericht

Der Kassenbericht schließt an den am 18.02.82 in Koblenz vorgelegten Kassenbericht an. Er endet mit dem 31.12.82.

1982 waren 15 Neuaufnahmen und ein Austritt zu verzeichnen. Der Verein zählte damit Ende 1982 84 Mitglieder. Darunter sind ein institutionelles Mitglied (IDS) und sechs studentische Mitglieder.

Der Kontostand des Vereins betrug am 31.12.82 DM 4.423,40. 1982 wurden DM 2.476,80 eingenommen, an Ausgaben fielen DM 739,40 an.

Die Einnahmen des Vereins aus Mitgliedsbeiträgen betragen zur Zeit jährlich DM 2.162.-, davon waren für 1982 noch DM 514.- an Außenständen (DM 218.- davon werden mit dem Lastschriftverfahren 1983 eingezogen). An dem Lastschriftverfahren sind von 84 Mitgliedern insgesamt 66 beteiligt.

Ohne den Druck des jährlichen Kolloquiums-Bandes wird 1983 mit Ausgaben von DM 800-1000 gerechnet.

Top 4: Entlastung des Vorstands

Der Vorstand wurde einstimmig entlastet.

Top 5: Neuwahl des Vorstands

Als Kandidaten für die Wahl zum 1.Vorsitzenden wurden J. Krause und B. Endres-Niggemeyer vorgeschlagen.
B. Endres-Niggemeyer lehnte die Kandidatur ab.
Mit 36 Stimmen (0 Gegenstimmen, 1 Enthaltung) wurde J. Krause zum 1. Vorsitzenden wiedergewählt. J. Krause nahm die Wahl an.

Für die Wahl zum 2. Vorsitzenden wurden P. Hellwig, G. Knorz, E. Kroupa und G. Willée vorgeschlagen. G. Willee lehnte die Kandidatur ab. Auf Antrag fand die Abstimmung schriftlich statt.
Bei 2 Enthaltungen und 1 ungültigen Stimme entfielen auf P.Hellwig 22 Ja-Stimmen, auf E.Kroupa und G.Knorz je 6 Ja-Stimmen. Damit war P. Hellwig zum 2. Vorsitzenden gewählt. P. Hellwig nahm die Wahl an.

Für die Wahl zum Schatzmeister wurden als Kandidaten B. Schaeder (in Abwesenheit) und E. Kroupa vorgeschlagen.
Ein Antrag auf schriftliche Abstimmung wurde abgelehnt.
B. Schaeder wurde mit 21 Stimmen wieder zum Schatzmeister gewählt. Auf E. Kroupa entfielen 9 Stimmen.
B. Schaeder nahm die Wahl an. Dem Vorstand lag eine entsprechende schriftliche Einverständniserklärung vor.

Für die Wahl zum Schriftführer wurden die Kandidaten E.Kroupa und A.Schmidt nominiert. Beide nahmen die Kanditatur an. Dem Antrag auf schriftliche Abstimmung wurde stattgegeben.

Für das Amt des Schriftführers wurde E.Kroupa mit 19 Stimmen gewählt. Auf A.Schmidt entfielen 12 Stimmen. E.Kroupa nahm die Wahl an.

Top 6: Bericht des Beirates

Im Berichtszeitraum 1982 fanden keine Beiratssitzungen statt. Jedoch hat der Beirat den Vorstand in seiner Arbeit unterstützt. Der Vorstand war in mehrfachem Kontakt mit einzelnen Beiratsmitgliedern.

Top 7: Neuwahl des Beirats

Als Kandidaten für die Wahl des Beirats wurden vorgeschlagen (ja=Annahme der Kandidatur, nein=Ablehnung, n.a.=nicht anwesend):
Wahlster (n.a.), Kuhlen (n.a.), A.Schmidt (nein), Lutz (ja), Lenders (ja), Endres-Niggemeyer (ja), Batori (ja), Ott (ja), Mochmann (n.a.), Metzing (n.a.), Rieger (n.a.), Günthner (n.a.), Niederehe (nein), Lehmann (n.a.), Schneider (n.a.), Zimmermann (nein), Krallmann (n.a.), Kirsch (ja).

Die Abstimmung fand schriftlich statt. In den Beirat gewählt wurden (bei den nicht anwesenden Mitgliedern vorbehaltlich ihrer Zustimmung):
Lutz (mit 30 Stimmen), Batori (24), Endres-Niggemeyer (24), Ott (23), Wahlster (21), Lehmann (17), Lenders (16).
Auf die übrigen Kandidaten entfielen:
Mochmann (15 Stimmen), Metzing (10), Kuhlen (9), Rieger (8), Krallmann (7), H.-J.Schneider (5), Kirsch (4) und Günthner (4).

Die anwesenden gewählten Mitglieder nahmen die Wahl an. Von den nicht anwesenden gewählten Mitgliedern wird die Zustimmung schriftlich eingeholt. Bei einer Ablehnung rücken die Mitglieder mit den nächst meisten Stimmen nach.

Top 8: Vorstandsantrag auf Satzungsänderung

Es lag ein Antrag auf Satzungsänderung vor, der lautete: "Der Verein LDV-Fittings e.V. möge fortan den Namen 'Gesellschaft für Linguistische Datenverarbeitung e.V. (GLD)-Wissenschaftliche Vereinigung zur Förderung der maschinellen Verarbeitung natürlicher Sprache' führen."

Der alte Vorstand unterstützte den Antrag auf Namensänderung. P.Hellwig leitete die Diskussion mit der Begründung für die Unterstützung des Antrags ein (die ausführliche Begründung ist im Fittins-INFO Nr. 01/1983 nachzulesen).

Die Hauptargumente, die in der lang und engagiert geführten Diskussion für die Namensänderung angeführt wurden, lauteten: Der Verein solle auf nationaler Ebene zu einem Sammelbecken werden für alle, die mit der automatischen Sprachverarbeitung zu tun haben. Auf internationaler Ebene solle er Ansprechpartner für internationale Organisationen (wie ALC-EC) sein und integrierend wirken. Dazu bedürfe es eines Namens, der deutlich die Ziele des Vereins zum Ausdruck bringe. Der Name LDV-Fittings sei mißverständlich und der veränderten Situation nicht mehr entsprechend.

Die Argumente, die für eine Beibehaltung des Namens angeführt wurden, waren im wesentlichen folgende:
Der Name sei traditionell gewachsen; uneindeutige Namen förderten die Diskussion über den Verein; der Name sei als Motto für die Zielsetzng der Kooperation und Koordination von Forschungsvorhaben auf dem Gebiet der linguistischen Datenverarbeitung gewählt worden und diese Zielsetzung habe sich nicht verändert. Neue Mitglieder machten ihren Eintritt nicht vom Namen sondern von der geleisteten Arbeit abhängig.

Bei einem Meinungsbild ergab sich schließlich eine große Mehrheit (35:2) für eine Namensänderung.

Anschließend wurde der neue Name diskutiert.
Es wurde festgelegt, daß über den neuen Namen, das zugehörige Kürzel und die Satzungsänderung in der angegebenen Reihenfolge abgestimmt werden solle. Als neue Namen waren vorgeschlagen:

(1) Gesellschaft für Computerlinguistik
(2) Gesellschaft für Linguistische Datenverarbeitung
(3) Gesellschaft für Sprache und Datenverarbeitung
(4) Gesellschaft für Sprachdatenverarbeitung

Der 2. Vorschlag wurde mit 26:7:2 angenommen. Der Antrag, das neue Kürzel möge "GLDV" lauten, wurde einstimmig angenommen. Dem Antrag, den neuen Namen ohne Untertitel zu führen, wurde mit 25:4:8 zugestimmt. Der abschließende Antrag auf Satzungsänderung lautete:

Der §1(1) möge wie folgt geändert werden:

Die Vereinigung führt den Namen "Gesellschaft für Linguisitsche Datenverarbeitung (GLDV) e.V." und soll in das Vereinsregister eingetmgen werden.

Bei der Abstimmung waren 32 stimmberechtigte Mitglieder anwesend, es lagen 4 Stimmübertragungen vor.
Der Antrag auf Satzungsänderung wurde angenommen: 31:3:2.

Top 9: Vortrag von Frau Endres-Niggemeyer

Frau Endres-Niggemeyer stellte die Förderungssituation im Bereich LDV vor. Sie berichtete vor allem über die Fördervorhaben ESPRIT und EUROTRA. Einzelheiten werden im nächsten Fittings-Info nachzulesen sein.

Top 10: Planung 1983/84

1. Es ist geplant, die Kooperation mit der "Association for Computational Linguistics (ACL)" und der "Deutschen Gesellschaft für Sprachwissenschaft auszubauen.

2. Als Thema für das nächste Kolloquium wurde vorgeschlagen, einen aktuellen Überblick über den Stand der Linguistischen Datenverarbeitung in der BRD zu geben. Dies soll durch einen Call for Papers (rechtzeitig bis ca. Juli) vorbereitet werden. Als Tagungsorte wurden Heidelberg (in Zusammenarbeit mit IBM) und München (mit SIEMENS) vorgeschlagen.

3. Die Arbeit in den Arbeitskreisen soll vorangetrieben werden. Im Arbeitskreis Mikrocomputer soll die begonnene Arbeit durch Marktbeobachtung fortgesetzt werden. Der AK Sozialwissenschaften soll auf dem Soziologentag mit dem Thema CUI auftreten. Es wurde der Vorschlag gemacht, einen AK Maschinelle Übersetzung ins Leben zu rufen.

4. Es wurde angeregt, Kontakte mit DDR-Wissenschaftlern auf dem Gebiet der LDV zu intensivieren und sie zu Kolloquien einzuladen. Dagegengehalten wurde, daß dies daran scheitern werde, daß DDR-Wissenschaftler in der Regel nur für internationale Tagungen Ausreisegenehmigungen bekämen; daß aber schriftliche Kontakte bestünden.

5. Clearingaktivitäten:
Seit Herbst 1982 werden mit der GID Verhandlungen über die Fortführung der Softwaredokumentation geführt. Es sollen die Möglichkeiten für einen Antrag für ein Projekt: "LDV-Fittings-Softwaredokumentation" durch Krallmann/Krause mit einer Laufzeit von 2 Jahren geprüft werden. Der vorgeschlagene Umfang: 1WM, 1WHK, Geldmittel für Kontakte.



Top 11: Verschiedenes

1. Der Vorstand wurde aufgefordert, eine Mitgliederliste an alle Mitglieder zu verschicken; und außerdem ein Faltblatt mit der Satzung und den Zielvorstellungen des Vereins den Mitgliedern zugänglich zu machen.

2. Es soll geprüft werden, ob für alle Mitglieder Mitgliedsausweise ausgestellt werden sollen.

25.03.1983

J. Krause
(1. Vorsitzender)

E. Kroupa
(Schriftführer)