

# Tagungen

## Information Knowledge Evolution

**Helsinki 24.8. - 1.9.1988**

Die 44. Konferenz der Federation Internationale d'Information et de Documentation (FID) fand vom 24. August bis 1. September in Helsinki in der von Alvar Aalto gebauten Finlandia-Halle statt, in einem Rahmen, der im besten Sinn der finnischen Architektur edel ist. Das Organisationskomitee unter Vorsitz von Ritva Launo hatte vorbildliche Arbeit geleistet, unterstützt von Komitees für einzelne Aufgaben, in denen finnische Kolleginnen die Verantwortung trugen. Das Ergebnis war eine große Konferenz mit fast 700 Teilnehmern, die bis ins Detail des Arbeits- und des Begleitprogramms von selten erlebter Freundlichkeit und Umsicht geprägt war. Die wahrhaftig nicht verwöhnte Mitteleuropäerin erlebte zudem einmal eine Konferenz, bei der - wie der Moderator des Abschlußpanels formulierte - Männer und Frauen gleich waren.

Der folgende Bericht versucht nur begrenzt, ein Gesamtbild von der FID-Konferenz zu geben. Im wesentlichen greift er den Ausschnitt heraus, der computerlinguistisch/informationwissenschaftlich interessant ist. Im Vordergrund stehen die sprachbezogenen Themen. Hinzu kommen allgemeinere Diskussionspunkte, die für das informationwissenschaftliche und computerlinguistische Arbeitsumfeld von Bedeutung sind.

### 1 Eröffnung

Dem anspruchsvollen Gesamtthema der Konferenz entsprach die Besetzung der Eröffnungsveranstaltung. Nach dem finnischen Ministerpräsidenten Harri Holkeri, der Vorsitzenden des Organisationskomitees und dem FID-Präsidenten Michael Hill hatte Alexander King das Wort, der Vorsitzende des Club of Rome. Sein Thema war der Übergang zu einer Informationsgesellschaft, die von der heutigen Gesellschaft so verschieden sein würde wie die industrielle Gesellschaft von der

agraren. Er zählte neben der Bevölkerungsexplosion und dem Treibhauseffekt, der zu einer Erhöhung der Temperaturen auf der Erde führt, auch die sozialen Folgen der Informationstechnik zu den Triebfedern der Veränderung. Den letzten Vortrag hielt ein hochbetagter finnischer Philosoph, Sven Krohn. Er charakterisierte menschliches Wissen in einer Weise, die auch KI-Forscher nicht befremdet hätte.

### 2 Inhaltserschließung und Dokumentbeschreibung

Die Sitzung 3 "Inhaltserschließung und Dokumentdeskription" war dreifach untergliedert. In einer ersten Teilsitzung wurden linguistische Aspekte thematisiert, in einer zweiten Probleme und Trends in Klassifikation und Indexing. Hinzu kam eine Sondersitzung "Dokumentation von Industrieprodukten - sprachliche Aspekte". Da nicht alle Sprecher ihr Papier auch vortragen konnten, wurden für den Bericht die Proceedings mit herangezogen.

#### 2.1 Linguistische Aspekte in der Inhaltserschließung

Amy Warner aus Madison, WI schlug in ihrer Keynote eine gezielte Forschungsanstrengung im Information Retrieval vor, die linguistische Forschungsergebnisse integriert. Damit kann sie sich des Beifalls europäischer Kolleg(inn)en sicher sein, die sich in derselben Richtung weitere Perspektiven eröffnet haben: Sie beziehen sich über das Information Retrieval hinaus auf die ganze Informationswissenschaft. Sie sehen in der Linguistik über die Kern- und Systemlinguistik hinaus mehr auf so aktive Spezialgebiete wie Computerlinguistik, Textlinguistik und kognitive Linguistik, die informationswissenschaftlichen Interessen besonders entgegenkommen. Im konkreten Fall ist mehr Meinungsaustausch quer über den Atlantik sicher hilfreich, um einen wissenschaftlichen Neuanfang voranzutreiben.

J. Niemistö und H. Jäppinen berichteten über die Schwierigkeiten, die beim automatischen Indexing in einer stark flektierenden Sprache wie Finnisch zu lösen sind. Eine morphologische Analyse ist erforderlich. Das System MORFO zerlegt zusammengesetzte Wörter in ihre Bestandteile und reduziert flektierte Formen auf eine kanonische Form. Bei Homographen werden alle Interpreta-

tionen erzeugt.

Ebenfalls im Arbeitszusammenhang einer automatischen Indexierung für die Freitextgesuche argumentierte Gerda Ruge für eine Verbesserung des Retrievals mit booleschen Operatoren und Kontextoperatoren wie ADJACENT, NEXT oder NEAR durch eine zusätzliche Suche mit Abhängigkeitsstrukturen in Nominalphrasen, wie sie das SIEMENS-System COPSY anbietet. Die gefundenen Dokumente können anhand der Übereinstimmung in den Nominalphrasen der Suchfrage und der Dokumente in Relevanzklassen eingeteilt werden. COPSY ordnete in einem Test 85% der Dokumente in die richtigen Relevanzklassen ein. Die restlichen 15% wurden in Nachbarklassen gefunden.

Hinter dem von Carmen Cazares vorgestellten Verfahren zur automatischen spanischen Indexierung russischer wissenschaftlich-technischer Texte steht das Anliegen, lateinamerikanischen Nutzern russische Datenbanken zugänglich zu machen. Eine russische morphologische Analyse liefert die Grundformen bedeutungstragender Wörter. Sie werden mithilfe eines Lexikons durch spanische Wörter ersetzt. Bei Homonymen werden bei Bedarf mehrere Alternativen angeboten. Durch dieses Indexierverfahren wird ein Retrieval und eine Einschätzung des Dokumentinhaltes auf Spanisch möglich. Das System ordnet je nach Sachgebiet zwischen 93% und 99% richtige spanische Indexterme zu.

Der Beitrag von Ahmad Ali Al-Falahi et al. aus dem Iran beschreibt ein einfaches Verfahren, das Weiterentwicklungen des Zipfschen Gesetzes zur computerunterstützten Indexierung heranzieht.

## 2.2 Probleme und Trends in Klassifikation und Indexing

In ihrem einleitenden Überblick machten Peter Ingwersen und Irene Wormell drei Typen von informationswissenschaftlichen Forschungsansätzen aus: systemorientierte, benutzerorientierte und kognitive. Die gebräuchlichste Retrievaltechnik verwendet Boolesche Operatoren und beruht auf der exakten Übereinstimmung von Suchkette und gesuchter Kette. Die Schwächen des Verfahrens sind bekannt. Man versucht unter anderem mit genaueren Indexierungsverfahren wie PRECIS oder mit intelligenten Benutzerschnittstellen Verbesserungen zu erreichen. Erweiterte Verfahren des booleschen Retrievals und Modelle des probabilistischen Retrievals gewichten die Terme von Indexierung und Frage. Sie sind sehr viel effizienter. Ebenso gute Ergebnisse lassen sich mit Clustermethoden erreichen. Die Verfahren der Textrepräsentation richten sich nach praktischen

Anforderungen. Da sich die Verwendungen von Informationen schnell ändern, entstehen neue Anforderungen an die Repräsentation von Dokumenten und den Zugang zu ihnen. Das Hauptproblem der Dokumentrepräsentation ist die unterschiedliche Interpretation von Konzepten durch verschiedene Personen. Das Systemdesign hat sich an den Informationsproblemen der Benutzer zu orientieren. Besonders unklare thematische Retrievalfragen sind schwer zu behandeln.

Im Rahmen eines booleschen Retrieval beobachtete Miranda Lee Pao, daß eine thematische Indexierung und eine Indexierung durch die zitierte Literatur ganz unterschiedliche Retrievalergebnisse bringen. Die Indexierung durch die zitierte Literatur führt zu einer hohen Precision. Die Schnittmenge beider Indexierungen empfiehlt sich, wenn eine sehr hohe Precision angestrebt wird.

In zwei Beiträgen wurden kognitive Modelle und Verfahren herangezogen, um Prozesse der Inhaltsschließung zu erklären. Brigitte Endres-Niggemeyer schlug einen Beschreibungsrahmen für die Inhaltsschließung vor, der auf zwei adaptierten allgemeineren Modellen beruht: einem Modell des Schreibens von Sachtexten und einem Modell des Textverstehens und des Zusammenfassens von Texten. Sie konnte neben einem Forschungsprogramm erste elementare Abstractingoperationen formulieren. Irmeli Hovi untersuchte bei 35 Versuchspersonen, wie sie beim Klassifizieren vorgehen. Insgesamt entspricht die Grobgliederung des Vorgangs in drei Phasen (Analyse des Dokumentes, Feststellen der zentralen Begriffe, Ausdrücken der Begriffe mithilfe der Klassifikation) nach ISO 5963 der Realität. Das Ziel-Klassifikationssystem beeinflusst die Repräsentation, die von dem Dokument gebildet wird. Vorliegende Klassifikationen erschweren den intellektuellen Prozeß oft, indem sie zu viele alternative Klassen anbieten, ohne daß deren unterschiedliche Eignung für den konkreten Fall deutlich würde.

Eine Gruppe von Vorträgen befaßte sich mit der DK. Päivikki Karhula schilderte ein finnisches Projekt, daß im Anschluß an die Arbeiten von Irmeli Hovi eine bessere Anpassung der DK an den Gebrauch in Online-Datenbanken anstrebt. Klaus Loth und Herbert Funk stellten das System ETHICS zum Literaturretrieval auf DK-Basis vor, das an der ETH Zürich benutzt wird. Eugeniusz Szibor und Ina Shcherbina-Samojlova erklären die neue Strategie der DK-Revision. Fünf koordinierende Revisionskomitees sind für breite Themenbereiche verantwortlich. Neben Revisionskomitees arbeiten auch bezahlte Experten an der DK-Revision. Revidiert wird Schritt für Schritt, um einen Bruch zwischen der bestehenden und ei

ner revidierten DK zu vermeiden. Yukio Nakamura und Tetsuya Ishikawa präsentierten unter dem verheißungsvollen Titel "Expert system for automatic UDC number assignment" ein System zur computerunterstützten DK-Benutzung.

Maasaki Tonosaki beschrieb medizinische Fachzeitschriften anhand ihrer bibliographischen Struktur. Besonders die vorkommenden Kolumnen (Originalaufsätze, Rezensionen usw.) werden ausgewertet. Die Zeitschriften werden zu Clustern zusammengefaßt, so daß sich eine Typologie der Fachzeitschriften ergibt.

Henning Spang-Hanssen plädierte für mehr Forschung und Entwicklung zur Registererstellung. Sein wichtigstes Beispiel war der Sachindex der DK. Nur mit verbesserter Methodik können mit dem Computer produzierte die Informationsqualität der von Menschen hergestellten Register erreichen.

Eindrucksvoll war, wie M.M. Moshoeshe das mühsame Entstehen des Bibliothekswesens und die Einführung einer bibliothekarischen Sacherschließung in Lesotho nachzeichnete. In einem kleinen Land, in dem in den sechziger Jahren die ersten wissenschaftlichen Bibliotheken entstanden, ist heute der Anschluß an übernationale Bibliotheksnetze erreicht.

Neva Tudor-Silovic stellte am Beispiel einer in Dubrovnik veranstalteten Konferenzserie die informatorische Wirkung dar, die wissenschaftliche Tagungen entfalten können.

Als Hintergrund für die Bemühungen um eine bessere Methodik und mehr internationale Kooperation in der Inhaltserschließung war der Beitrag von Antje Schurek über das General Information Program der UNESCO interessant.

### 2.3 Dokumentation von Industrieprodukten - linguistische Aspekte

Das FID-Komitee "Linguistics in documentation" stellt, wie Hans Karlgren erläuterte, linguistische Fragen in einen größeren Zusammenhang. Seit zwei Jahren wird das Thema "Technische Dokumentation" behandelt. Ohne begleitendes Informationsmaterial sind komplizierte Produkte wie Papierfabriken oder Softwaresysteme nicht betriebsfähig. Die Produktion der Benutzerinformation bedeutet aufgabenorientiertes Schreiben unter professionellen Bedingungen für eine praktische Verwendung. Es handelt sich um einen besonders interessanten Spezialfall der Wissensüber-

mittlung. Eine Fallstudie in einem großen Unternehmen in Stockholm zeigte zwar Problembewußtsein, aber wenig tatsächliche Verbesserungen in den benutzten Verfahren. Insbesondere führte mehr Computereinsatz nicht zu besseren Methoden.

Kolbjörn Heggstad betonte die Notwendigkeit, Softwareprodukte und ihre Dokumentation ebenso wie andere Produkte den örtlichen Gegebenheiten anzupassen, sie zu "lokalisieren". In der Muttersprache lerne man gerade unter schwierigen Bedingungen leichter und gründlicher. Man solle doch keine Angst haben, daß man nicht international genug sei. Sinnvoller sei es, die eigene Sprache und Terminologie so zu entwickeln, daß eine lokale EDV-Kultur in ihr darstellbar sei.

Öyvind Haaland trug eine gut präsentierte Fallstudie bei: Die Dokumentation zur Oseberg-Ölbohrinsel nordwestlich von Bergen umfaßt mehr als eine Million DIN-A4-Seiten. Als Arbeitssprache sollte "Norenglish" aus Gründen der Sicherheit und der kulturellen Identität vermieden werden. Eine norwegische Terminologie war zu entwickeln. Die beauftragte Firma übersetzte, führte Kurse im technischen Schreiben durch, entwickelte eine norwegische Öl-Terminologie, überprüfte und überarbeitete die Dokumentation und die Lehrgangsunterlagen. Die Investition in die Dokumentation hat sich unter dem Gesichtspunkt der Qualitätssicherung und der Vermeidung von Zusatzkosten gelohnt.

Die Podiumsdiskussion über Language Engineering wurde ganz von Skandinaviern bestritten. Henning Spang-Hansen fand nicht, daß man eine nationale EDV-Sprache brauche: das Problem sei eher, immer auf dem laufenden zu bleiben; (sprachlich) Exotisches verkaufe sich besser; Zweisprachigkeit sei ein normales Phänomen vielen Ländern; bei knappen Ressourcen habe auch die Bewahrung der sprachlichen Identität hinter wichtigeren Zielen zurückzustehen. Dem hielt der "clever localizer" Heggstad entgegen, daß eine exotische Marke sich nur dann besser verkaufe, wenn das Produkt den einheimischen Anforderungen an Funktionstüchtigkeit standhalte. Hans Karlgren trug aus der schwedischen Erfahrung bei: Ein Volk von 8 Millionen sei zu klein, um die Fachsprachen für alle Spezialgebiete zu unterhalten. Man möge darum nicht den eigenen Nationalismus in andere Länder exportieren. Einen denkbaren positiveren Verlauf der Sprachgeschichte in Mittel- und Nordeuropa ließ er mit Luthers Bibelübersetzung beginnen: Als Basis für die deutsche Schriftsprache wählt Luther nicht, wie dummerweise passiert, einen hochdeutschen Dialekt, sondern das Niederdeutsche, das als Hansespra-

che auch die Arbeitssprache in ganz Skandinavien war. Die auf dem Niederdeutschen beruhende Schriftsprache bleibt im heutigen deutschsprachigen Raum und in ganz Skandinavien die Hochsprache. Auch das verhältnismäßig künstlich belebte Finnisch kann sich nicht durchsetzen, da statt des Schwedischen Deutsch in dieser Ausprägung der -stärkere - Gegenpart ist. In den jungen USA geht die Abstimmung, ob Englisch oder Deutsch die Amtssprache werden soll, nicht knapp zugunsten des Englischen, sondern mit der Unterstützung der skandinavischstämmigen US-Bürger zugunsten des Deutschen aus. Für Skandinavien stellt sich unter diesen Umständen das Problem gar nicht, eventuell eine vom Englischen abweichende EDV-Terminologie zu entwickeln.

### 3 Menschen als Informationsprozessen

Matti Bergström machte bei Menschen eine vierstufige Steuerung der Informationsverarbeitungsprozesse im Gehirn aus: Die "Value capacity" steuert die eigentliche Informationsverarbeitung; diese kontrolliert den Energiefluß, welcher seinerseits auf die Materie einwirkt. Wird die "Value capacity" in der Erziehung nicht ausreichend eingeübt, so entsteht eine spezielle Form der Behinderung, die das Individuum unfähig macht, Situationen ganzheitlich zu sehen und aufgrund dieser Situationssicht zu entscheiden. Die "Value capacity" scheint weder Information noch Energie zu sein. Sie ist mit den holistischen Fähigkeiten der rechten Hirnhälfte verbunden. Sie steuert auch die Motivation und die emotionalen Fähigkeiten und damit das ethische Verhalten von Menschen.

Sehr instruktiv war der Beitrag von Thomas Froehlich zu dem zentralen informationswissenschaftlichen Begriff der Relevanz. Er stützte sich auf zeitgenössische Philosophen (Schulz, Merleau-Ponty, Habermas, Heidegger, Gadamer, Taylor) um zu zeigen, daß es keine absolute Fundierung des Wissens gibt. Wir werden in eine praktische - geschichtliche und soziale - Situation hineingeboren, aus der unser Wissen kommt und in die es wieder aufgenommen wird. Auf dieser Basis entwickeln wir einen allgemeinen Begriff von Relevanz. Entwickelter Relevanzurteile beruhen auf einer geglückten Einübung in den spezielleren Lebenszusammenhang der Relevanzbestimmung z.B. bei der Recherche in Informationsbanken.

### 4 Elektronisches Wissen

Linda Smith stellte unter dem Titel "Wholly new forms of encyclopedias" die Hauptcharakteristika

von Hypertextsystemen vor. Hypertexte sind als Konzept nicht neu, aber sie gewinnen neue Aktualität durch die Tatsache, daß nun Systeme zu ihrer elektronischen Darstellung kommerziell verfügbar sind. Während gedruckte Texte im Prinzip eine lineare Lektüre annehmen, liegt einem Hypertext eine vernetzte Struktur des Textes zugrunde, der ein Leser auf selbstbestimmten Pfaden folgt oder auf solchen, die der Autor für unterschiedliche Lektüren eingerichtet hat. Genaueres ist über die Produktion, die Verteilung und den Gebrauch von Hypertexten ist noch zu erforschen.

### 5 Informationssysteme mit schnellem Wachstum

In mehreren Beiträgen (Henning Albrechtsen, Leif Brisfjord, Inez Sperr Brisfjord) wurde das in Skandinavien sehr erfolgreiche Konzept "Telecottage" erklärt. Es handelt sich um lokale Computer- und Informationsnetz-Stationen, die Infrastrukturleistungen, Ausbildung, Beratung und ähnliches vor allem im ländlichen Raum bereithalten. Sie erleichtern in ihrer Gemeinde die dezentrale EDV-Nutzung besonders durch kleine Betriebe oder Privatpersonen. Damit tragen sie zur sozialen und wirtschaftlichen Belebung gerade abgelegener Gebiete bei. Die Organisation der Telecottages ist unterschiedlich, beruht jedoch auf dem Prinzip der Selbstverwaltung.

### 6 Podiumsdiskussion "Überwindung der Schranken im Informationstransfer - lokale und globale Informationsnetze"

Zur abschließenden Podiumsdiskussion zum Thema Informationstransfer und Informationsnetze hatte Peter Canisius vier Frauen und vier Männer aus Afrika, Lateinamerika, Nordamerika und Europa auf dem Podium versammelt. Zwei Beispiele müssen genügen, um den Stil der Diskussion deutlich zu machen, die in bemerkenswerter Weise die unterschiedlichen persönlichen Standpunkte der Teilnehmerinnen) aus sehr verschiedenen Lebensbereichen zur Geltung brachte. Marjatta Okko, nach eigener Aussage die dienstälteste finnische Professorin der Informationswissenschaft, hatte die Aufgabe übernommen, das Thema in Frage zu stellen. Sie wies auf die höchst nützliche Kanalisierungsfunktion von Informationsbarrieren hin. Von ihr kam später noch der beherzigenswerte Vorschlag, von der FID-Konferenz eine wissenschaftliche Konferenz abzutrennen.

M.M. Moshoeshoe, die Vorsitzende des bibliothekarischen Berufsverbandes in Lesotho, schilderte mit aller wünschenswerten Deutlichkeit, wie sich in ihrem beruflichen Alltag die Frage nach den Informationsnetzen neben die Frage stellt, ob man die Bücher mit einem Esel transportieren sollte, und ob nicht die schlimmste Informationsbarriere darin bestehe, daß 85% der Bevölkerung Analphabeten seien. Insgesamt kam ein selten facettenreiches Meinungsbild zustande, das die Vorzüge der Konferenz noch einmal konzentriert erfahrbar machte, bevor der Moderator das Gesamtresümee zog.

Die nächste FID-Konferenz findet 1990 in Havana (Cuba) statt.

*Prof. Dr. Brigitte Endres-Niggemeyer, FH Hannover, Fachbereich BID*

## Japanbesuch im November 1988

### Ein persönlicher Reisebericht von Karin Haenelt, IPSI Darmstadt

#### Allgemeines

Vom 23.11. bis 26.11.1988 fand in Kyoto das 5. Deutsch-Japanische Forum "Informationstechnologie" statt. Daran habe ich als eingeladenes Mitglied teilgenommen.

Im Vorprogramm gab es Firmenbesichtigungen, ich nahm an einem Besuch bei ATR teil. Darüberhinaus habe ich - z. T. mit dankenswerter Unterstützung von Dr. E. Bierdümpele (GMD Außenstelle Tokyo) - folgende Besuche organisiert: ATR (2. Besuch im NL-Labor)/Osaka, EDR/Tokyo, Prof. Nagao/Universität Kyoto, GMD/ Tokyo. Das 6. Forum ist für Mai/Juni 1990 in Berlin geplant.

## 1 Advanced Telecommunication Research Institute International (ATR), Osaka

### 1.1 Besuch

Am 21.11.1988 gab es vormittags einen organisierten Rundgang durch die Labors für eine Gruppe von Forumsteilnehmern, eine Begrüßungsrede und Institutsvorstellung von Kohei Habara (Executive

Vice President). Nachmittags hatten Prof. W. Wahlster und ich ein gesondertes Treffen mit Mitarbeitern des Natural Language Understanding Labors und führten fachliche Diskussionen.

Teilnehmer des Gesprächs waren Teruaki Aizawa (Head, Natural Language Understanding Department), Kiyoshi Kogure (Senior Researcher, NLU Dept.), Kei Yoshimoto (Senior Researcher, NLU Dept.), Martin Emele (Gastforscher, IMS Stuttgart, z.Z. ATR), Wolfgang Wahlster (DFKI, Saarbrücken).

### 1.2 Organisation von ATR

Das Advanced Telecommunication Research Institute International (ATR) in Osaka wurde am 22. März 1986 gegründet. Es hat die Aufgabe, Grundlagenforschung in allen Bereichen der Telekommunikation zu betreiben und internationale Kooperationen aufzubauen. Organisatorisch ist ATR eine "umbrella company for cooperations" und unterstützt vier Forschungslabors durch Personal, Investitionsmittel und Infrastruktur. Finanziert wird das Unternehmen zu 70 % durch das staatliche Japan Key Technology Center, zu 30 % durch ATR und andere private Unternehmen. Das Finanzvolumen beträgt 20 Mio Yen bis 1989, 70 Mio Yen bis 1995. (Man vergleiche das einmal mit der deutschen Kapazität!). Das Personalkonzept sieht vor, daß erfahrene und qualifizierte Mitarbeiter von Firmen für einige Jahre zu ATR entsandt werden und dort in Forschungsprojekten arbeiten. ATR hat zur Zeit 180 Mitarbeiter. Ziel ist es, den Personalstand auf 300 auszubauen. Derzeit gibt es etwa 40 Projekte. Internationalität wird durch Gastforscher und internationale Kooperationen begründet. Im Bereich "Natürliche Sprache" gibt es derzeit Kooperationen mit SRI/Stanford, CMU /Pittsburgh, AIAI/Edinburgh. 1989 wird das Institut in das neue Institutsgebäude in der Kansai Science City (zwischen Osaka und Kyoto) umziehen.

### 1.3 Kooperation IPSI-ATR

ATR ist in seiner Struktur, Aufgabenstellung und Vorgehensweise IPSI sehr ähnlich. An beiden Instituten tragen mehrere Gruppen zur Lösung längerfristiger komplexer Forschungsaufgaben bei. Gemeinsamkeiten und Ergänzungen gibt es in mehreren Bereichen. Prof. Nagao schlug vor, eine Kooperation aufzubauen. Auch Prof. Wahlster plant, eine Kooperation aufzubauen.

### 1.4 ATR Optical and Radio Communications Research Laboratories

Forschungsgegenstand dieser Labors sind optische Intersatelliten- Kommunikation, Materialien für

neue Kommunikationsausrüstungen (klein, leicht, geeignet für verschiedene Funktionen) und weitere Beiträge zum National Space Plan. Dabei sind z.Z. folgende Methoden von Interesse: Neue Frequenzbänder, neue Techniken paralleler Informationsverarbeitung unter Verwendung nicht-linearer Techniken, neue Meßtechniken, Hochfrequenzschaltkreis- Designtechniken .

### 1.5 ATR Auditory and Visual Perceptions Research Laboratories

Forschungsgegenstand dieser Labors sind Benutzerschnittstellen, und zwar sollen hier Erkenntnisse der Forschungsrichtungen Perzeption, Kognition, Gedächtnis, Lernen umgesetzt werden. Gezeigt wurde uns viel Konnektionismus, viel massiv parallel mit den üblichen Muster-Erkennungs-Anwendungen, während Konzepte der propagierten Forschungsrichtung noch nicht zu erkennen waren. (Zu den Vorführungen gehörte die Erkennung handschriftlicher Ziffern (NEOCOGNITRON), die Restauration gestörter Bilddaten, implementiert auf NCUBE, die Erkennung von Gesichtern durch Mustervergleich mit partiellen und gestörten Eingabedaten sowie die Kompression von Bilddaten).

### 1.6 ATR Communication System Research Laboratories

Ziel dieser Labors ist es, Mensch-orientierte Kommunikationssysteme zu entwickeln. Zu diesem Zweck werden derzeit folgende Gebiete bearbeitet: Benutzerfreundliche Telekommunikationssysteme, 3D-Bilder, Stereo-Bilder, Verfolgung von Augenbewegungen und Gestik. Zur Erforschung der Gestik wird ein "data glove" für Forschungszwecke benutzt, soll aber nicht für den späteren Benutzer bestimmt sein. Vielmehr ist es das Ziel, mit der Hand direkt zeigen zu können. Der Arm wird methodisch in Zylinder eingeteilt, denen Primitive zugeordnet werden. Die Interpretation erfolgt mit einem "truth maintenance system". Auch hier liegt der Schwerpunkt der Arbeiten eher im Bereich der externen Medien, eine Beschäftigung unter semantischen und pragmatischen Gesichtspunkten war mir nicht erkennbar.

### 1.7 ATR Interpreting Telephony Research Laboratories

Für mich waren am interessantesten die Vorführungen und Gespräche in diesen Labors, über die ich daher ausführlich berichte. Ziel dieser Labors ist das inzwischen weltweit bekannte Ziel der automatischen (Simultan-) Übersetzung von Telefongesprächen, durch die Sprachbarrieren abgebaut werden sollen. Auch die anderen Labors arbeiten teilweise an diesem Ziel mit. Es ist,

wie es hieß, allerdings kein Ziel, das man in 15 Jahren für erreichbar hält. Vielmehr ist es ein Leitgedanke, eine Anforderung, die der Definition von Forschungsaktivitäten und der Identifikation von Forschungslücken dient. Zu diesem Bereich gehören drei Labors, nämlich das Natural Language Understanding Department, das Knowledge and Database Department und das Speech Processing Department. Davon hat das NLU Dept. zur Zeit 20 Mitarbeiter, darunter 18 Informatiker und nur 2 Linguisten! Auch in Japan ist es sehr schwer, Linguisten zu finden, die für solche Aufgaben qualifiziert sind, zumal auch dort entsprechende Studiengänge fehlen.

#### 1.7.1 Übersicht der Forschungsvorhaben

Das gesamte Forschungsfeld konstituiert sich durch die Gebiete

- . Spracherkennung
- . Spracherzeugung
- . Modellierung des Hörens .
- Maschinelle Übersetzung

#### 1.7.2 Maschinelle Übersetzung

Längerfristig halte ich den japanischen Ansatz für erfolgversprechend, charakteristisch sind folgende Grundgedanken:

- corpusbasiert
- kontextuelle Eigenschaften der Konversation

Zur Zeit hält man in Japan Dialoge für etwas prinzipiell Anderes als monologische Texte. Das widerspricht texttheoretischen Erkenntnissen und hängt damit zusammen, daß man auch in Japan noch keine semantisch/pragmatisch orientierten Textmodelle entwickelt. Die Bestrebungen, die Satzgrenze zu überschreiben, richten sich zunächst einmal auf Sprechakte. Der bisherige Ansatz führte zur Planung folgender Systemkomponenten:

#### . Erkennung

des Oberflächensprechaktes, Plan-Erkennung (Grundlage: Diskursplan, Interaktionsplan, Domänen-Plan, Plan-Skripts und Schemata), Abbildung im Sinn der "Situation Semantics". Für die Auflösung von Ellipsen und Mehrdeutigkeiten werden die Annahmen des "mutual belief" und "goal sharing" herangezogen.

- . An Diskurseigenschaften sollen Topik-Struktur und Diskursfokus berücksichtigt werden. Allerdings ist die Methode etwas simplifizierend: Sie beruht auf der Erkennung von Schlüsselwörtern und Schlüsselphrasen sowie Bestätigungsphrasen, die auf

der Grundlage des Corpus bestimmt werden. Darüberhinaus wird Plan-basiert gearbeitet, wobei Pläne im engeren Sinne der KI domänenbezogen definiert werden (d.h. man wird nur vordefinierte Domänen behandeln können). Gesehen wird allerdings in Japan, daß es Weltzugriffe gibt, die spezifisch japanisch oder spezifisch englisch sind (eine Erkenntnis, von der viele europäische einzelsatzorientierte Ansätze noch weit entfernt sind). Verwendete Techniken sind paralleles Parsing (unifikationsbasiert), parallele Suche von Referenz-Ausdrücken. Verfahren der inkrementellen Analyse/Synthese werden nicht erarbeitet (weil das mit Unifikationsgrammatiken nicht wirklich möglich sei).

. Transfer-Ansatz auf Sprechaktebene.

. Wissensbasis: Integration lexikalischen, semantischen und pragmatischen Weltwissens.

### 1.7.3 Gesprochene Sprache

In mehreren Projekten werden verschiedene Aspekte gesprochener Sprache untersucht, um Grundtechnologien zur Realisierung eines prototypischen Systems bereitzustellen, das kontinuierliche Sprache jeden Sprechers erkennt. Folgende Beiträge sollen geliefert werden:

. Erkennung kontinuierlicher Sprache:

- Phonem-Erkennung
- Erkennung eines großen Vokabulars - Sprecher-Anpassung
- Reduktion des "Rausehens"
- syntaktische und semantische Verarbeitung

. Sprachsynthese

- klarer, hochwertiger und natürlicher Klang
- Reproduktion der syntaktischen und semantischen Struktur, der beabsichtigten Botschaft des Original-Sprechers möglichst ähnlich
- Nachahmung der Stimmcharakteristik des Original-Sprechers

. Grundlage: Corpus mit Phonem Markierung.

. Menschliche Hör-Mechanismen und ihre Anwendung auf automatische Spracherkennungstechnologien

- psycho-akustische Experimente
- SprachWahrnehmungs- Experimente - Computersimulation von Hör-Modellen

- Hörmodellierung peripherer Signalverarbeitung
- Modellierung der Koartikulation in kontinuierlicher Sprache
- Phonem-Erkennung mit dem Ansatz neuronaler Netzwerke
- akustisch-phonemische Invarianz
- Funktionen der Prosodie (Betonung, Intonation)
- Interaktion von Wissensquellen in Sprachwahrnehmung und -erkennung

### 1.7.4 Vorführung

Vorgeführt wurde das Demonstrationssystem NADINE (Natural Dialogue Interpreting Expert), eine ingenieurartige Kopplung kommerziell verfügbarer (amerikanischer) Systeme mit folgenden Komponenten: Worterkennung (japan.), Parsing mit einem Unifikations-Formalismus, Transfer satzweise, Worterzeugung (englisch). Ein solches System kann natürlich nur Einarbeitungswert oder Demo-Effekt für ungeduldige Gutachter haben, es produziert schon witzige Ergebnisse für den Beispielsatz: "I'll be in Osaka next week".

### 1.7.5 Gesamteindruck

Die Methoden des Gesamtvorhabens erscheinen mir erfolgversprechend. Der wichtige Vorsprung gegenüber Europa entsteht vor allem durch folgende Vorgehensweise:

. Massiver Einsatz von Mitteln

. interdisziplinäre und integrierte Bearbeitung der komplexen Aufgabe

. Bearbeitung höherer Strukturebenen (Dialog, Text)

. Verbindung konnektionistischer und massiv paralleler Ansätze für die unteren Ebenen mit symbolischen Ansätzen für die höheren Ebenen. Allerdings erstreben die japanischen Ansätze noch keine modellbasierte semantische und pragmatische Verarbeitung von Dialogen und Texten oder auch multimodaler Interaktionen, wie sie derzeit am IPSI erforscht wird (Projekte KONTEXT, KOMET, COGITO).

## 2 Electronic Dictionary Research Institute, Ltd., Tokyo

### 2.1 Besuch

Vereinbart war ursprünglich ein Gespräch und eine Laborführung zum Informationsaustausch. Im Eifer der Diskussion entwickelte sich ein dreistündiges Gespräch.

Nach der Laborführung gesellten sich zu meinen Gesprächspartnern Hiroshi Uchida (Research Manager) und Toshio Yokoi (General Manager) auch noch Shiu-ya Amano (Senior Researcher, Toshiba) sowie Jun-ichi Tsujii (früher Kyoto, jetzt UMIST Manchester U.K.) und John McNaught (UMIST, Manchester U.K.). Eine weitere Diskussion und ein gemeinsames Abendessen schlossen sich an. Es war sehr spannend und unterhaltsam in äußerst freundlicher Atmosphäre.

## 3 Organisation von EDR

- . Sponsoren: Japan Key Technology Center und acht private Unternehmen: Fujitsu, NEC, Hitachi, Sharp, Toshiba, Oki Electric Industry, Mitsubishi, Matsushita
- . gegründet: 28. April 1986
- . Organisation: In der Zentrale in Tokyo arbeiten nur wenige Mitarbeiter, hauptsächlich das Management und ein kleines zentrales Labor. Die EDR Forschung wird einerseits neutral von Tokyo koordiniert, andererseits in den beteiligten Firmen dezentral betrieben. Dabei entwickeln die Firmen auch eigene (geheime) Nutzungen der gemeinsamen öffentlichen Forschungsergebnisse
- . Finanzvolumen: bis 1985 5 Mrd Yen, bis 1994 14 Mrd Yen (wie will da die deutsche Forschung konkurrenzfähig sein?)

### 3.1 Ziele des EDR

Ziel des großen Projekts ist es, Forschung und Entwicklung im Bereich sehr großer komplexer elektronischer Wörterbücher zu betreiben, wie sie für die zukünftige Technologie-Generation der Natursprachenverarbeitung sowie der Wissens- und Informationsverarbeitung benötigt werden. Als Anwendungen werden dabei genannt:

Intelligente

Textverarbeitung, Intelligente Bürosysteme, Maschinelle Übersetzung, Gesprochene Sprache, Wissensbasierte Inferenzen, schließlich sogar Robotik. Diese Arten sind bisher noch längst nicht durch elektronische Wörterbücher versorgt, da man ja z. T. noch gar nicht weiß, welche Arten

von Informationen dafür benötigt werden. Diese Frage ist Gegenstand aktuellster Grundlagenforschungsprojekte. Das Motto der gemeinsamen Aktivität von Verlagen, Computerfirmen und Forschung lautet: "Wörterbücher aus Computern, Wörterbücher mit Computern, Wörterbücher für Computer". Gemeint ist damit ein gemeinsames Vorgehen der Computer-Unterstützung jeder Art von Lexikonerstellung. Auch diese Initiative wird als Entwicklung einer Schlüsseltechnologie gewertet.

### 3.1.1 Kooperation EDR - Bundesrepublik Deutschland

Die Japaner äußern den Wunsch einer engen Zusammenarbeit mit einer deutschen Initiative. Besonderes Interesse besteht dabei im wissenschaftlichen Bereich an der Abstimmung von Schnittstellen, an zweisprachigen Lexika, an der theoretischen lexikalischen Fundierung im Bereich der Zusammenhänge Lexikon und Wissensbasis, an Methoden der Datenmodellierung sowie an Anwendungen. Im kommerziellen Bereich könnte eine solche Initiative deutsche Verlage und Computerfirmen unterstützen (internationaler Markt). Insgesamt ließe sich durch Zusammenarbeit zur Überwindung von Sprachbarrieren die Beziehung Deutschland-Japan entscheidend fundieren.

### 3.1.2 Einschätzung der Ansätze

Zur Zeit gibt es vier Typen von EDR Wörterbüchern:

- . Wortlexika (mit morphologischen, syntaktischen und semantischen Attributen sowie Angaben des Gebrauchs, der Häufigkeit und der Aussprache),
- . Konzept-Wörterbücher, die aus einem Konzept- Definitionsteil und der Beschreibung einzelner Konzepte bestehen,
- . Konkurrenz- Wörterbücher
- . Zweisprachige Wörterbücher (z. Z. Japanisch und Englisch).

Diese vier Typen von Informationen sollen miteinander in Relation gesetzt werden. Dabei werden Ansätze der Computerlinguistik, der Lexikographie und der künstlichen Intelligenz integriert. Darüberhinaus wird natürlich auch eine integrierte lexikographische Entwicklungsumgebung aufgebaut. Wir haben die Ansätze lebhaft diskutiert. Die wesentlichen Schlagworte sind: wiederverwertbar und zukunftsorientiert. Ich halte die japanischen Ansätze für fundiert und erfolgversprechend. Die Merkmale, die mir am wichtigsten erscheinen für eine entsprechende Diskussion in Deutschland sind folgende:



### 3.1.3 Gemeinsames Vorgehen von Verlagen, Industrie und Forschung

Von Anfang an werden beim Aufbau der japanischen Lexika/Wissensbasen Kenntnisse und Anforderungen verschiedener Erfahrungen berücksichtigt. Das ermöglicht den Japanern

- . einen sehr schnellen Aufbau der Datenbasis
- . einen weitsichtigen Aufbau der Datenbasis

### 3.1.4 Wahrung von Firmengeheimnissen

Das Problem ist in Japan so gelöst, daß an der expliziten und formalen Beschreibung des Sprachschatzes alle Partner öffentlich beteiligt sind, während die Anwendungen firmenintern bleiben. Eine unliebsame Erscheinung ist dabei allerdings, daß firmenintern gelegentlich aufwendige Umwandlungen der Datenmodellierung vorgenommen werden müssen (vgl. dazu aber auch 3.1.8).

### 3.1.5 Copyright

Dieses auch in Deutschland so schwerwiegende Problem soll in Japan dadurch gelöst werden, daß spezielle Einzelverträge über die Entnahme für kommerzielle Nutzung abgeschlossen werden sollen.

### 3.1.6 Vorgehensweise

Die Vorgehensweise der Japaner ist sehr pragmatisch: in einer ersten Phase, die jetzt abgeschlossen ist, wurde zunächst einmal eine riesige Datenbasis des Japanischen aufgebaut. Dies geschah auf der Grundlage von Kodierformularen, die von Leuten entworfen wurden, die alle langjährige Erfahrung in der maschinellen Übersetzung hatten. In der 2. Phase (ab 1989) beginnt die Erforschung der theoretischen Fundierung. Die Japaner sind dabei bereit, die Ergebnisse der 1. Phase wegzuzwerfen(!), neu zu konzipieren und aufzubauen. (Nur ein solcher Mut nimmt zukunftsorientierten Entwicklungen die Fesseln!).

### 3.1.7 Grammatikformalismen

Formalismen wie GPSG, LFG etc. werden von den Japanern milde als "US and EURO-formalisms" belächelt, die weder eingearbeitet wurden noch eingearbeitet werden sollen. Man beruft sich auf eine "long Japanese Experience in MT" und hält m.E. zu Recht die Anforderungen der Anwendungen und das Was der Beschreibung in diesem Zusammenhang für wesentlich wichtiger. Die Beschreibung der Phänomene der natürlichen Sprache steht im Vordergrund. In diesem Zusammenhang wurde das amerikanische Sinnieren über formale Eigenschaften von Formalismen für reine Zeitverschwendung angesehen, da es die

entsprechenden Probleme nicht löst, sondern vielmehr den vielzitierten vermißten Schlüssel unter der vielzitierten Laterne sucht, unter der er nicht liegt.

### 3.1.8 Datenmodellierung

Zur Zeit verwenden die Japaner eine Art hierarchisches/relationales Datenmodell, das keinerlei flexiblen Zugriffe gewährleistet. Auch sind die Zugriffe nicht effizient. Für jede Anwendung ist daher eine neue Datenmodellierung erforderlich. Das wird von den Japanern für eine echte Schwachstelle gehalten.

### 3.1.9 Semantik, Pragmatik und "future generation"

Besonders neugierig war ich darauf, zu erkunden, wie ernst die Japaner es meinen mit "future generation of natural language processing technology and knowledge information processing". Und mein Eindruck ist: *Sie meinen es sehr ernst!!!* Ich habe folgende futuristische tendencies bemerkt:

es wird **Grundlagenforschung** betrieben in den Bereichen:

- semantic information
- Zusammenhang von Lexika und Wissensbasen (ein Thema, dem auch ATR sich widmet)
- Überwindung von Diskursbereichsgrenzungen in Wissensbasen
- Anforderungen fortgeschrittener Sprachverarbeitung auf satzübergreifender bis textueller Ebene unter semantischen Aspekten.

Es geht also wirklich darum, auch im Bereich der Wissensrepräsentation entscheidende Fortschritte zu erzielen. Als Ansatz wurde die kontextuelle Definition von Konzepten gewählt. Dieser Ansatz war für mich natürlich auch deshalb von besonderem Interesse, weil wir ihn für die Lexika und Wissensbasen unserer Textprojekte KONTEXT und KOMET ebenfalls verfolgen. U. a. wird dazu z.B. das Longman Dictionary of Current English herangezogen, das europäische lexikologische Forschungsergebnisse am konsequentesten umsetzt. In Europa und speziell sogar in Deutschland gibt es einen großen Vorsprung im Bereich der theoretischen und praktischen Lexikographie und Lexikologie. Dieser Vorsprung sollte in der deutschen Grundlagenforschung und zum Aufbau Elektronischer Wörterbücher und schnellstens genutzt werden.

In der 1989 beginnenden Theoriephase werden zwei Anwendungen exemplarisch getestet: Maschinelle Übersetzung und gesprochene Sprache,

beide mit innovativen Techniken und Methoden (oberhalb der Satzebene, semantisch). Die erweiterten Anforderungen von Anwendungen, die zugleich Gegenstand der Grundlagenforschung sind, sollten also auch in Deutschland von Anfang an erforscht, analysiert und definiert werden.

## 4 Besuch bei Prof. M. Nagao, Kyoto University

### 4.1 Besuch

Am 22.11.1988 besuchten Prof. Barth (DFKI), Prof. v. Seelen (Uni Mainz), Prof. Wahlster und ich Prof. Nagao. Wir besichtigten kurz das Labor. Prof. Wahlster und Prof. Barth stellten Prof. Nagao und seinen Mitarbeitern das DFKI vor. Am 29.11.1988 hatte ich ein zweites Treffen mit Prof. Nagao. Bei diesem Besuch hatte ich ausführliche Gespräche mit Prof. Nagao und seinen Mitarbeitern, die mich durch das Labor führten. Ich habe dort IPSI und unsere Ansätze der Verarbeitung von Texten vorgestellt.

### 4.2 Laborausstattung

Grundausrüstung Symbolics, außerdem PSI-Maschine, SUN 4/260, Image-Processing Hardware.

### 4.3 Vorführungen und Diskussionen

#### 4.3.1 Sentence Generator

Satzgenerator auf der Basis systemischer Grammatiken. Grammatik für das Englische (entwickelt von J. Bateman), Grammatik für das Japanische.

#### 4.3.2 Textgenerierung

Dissertationsvorhaben, Behandlung von Kohärenz über truth-maintenance system.

#### 4.3.3 MD / GRADE

Übersetzungssystem des MU Projekts und Grammatik-Sprache. Übersetzung: Transfer auf der Basis von Baum- Transformationsregeln. Das System ist in der Literatur beschrieben. Vorgeführt von Jun-ichi Nakamura.

#### 4.3.4 Extraction of Semantic Information from the Longman Dictionary

Ziel ist es, sprachliche Relationen, die in der Verweis- und Eintragsstruktur eines Wörterbuches zum Ausdruck kommen, zum Aufbau einer Wissensbasis zu nutzen. (Vgl. ATR, vgl. EDR;

ein Ansatz, der jetzt an vielen Stellen in Japan verfolgt wird.) Bisher 80.000 Einträge, in Prolog. Bisher vor allem hierarchische Strukturen einzelner Wortfelder, Überlegungen gelten auch Netzen und Zirkeln.

### 4.3.5 Plan' der "Kansai Science City"

1989 beginnt zwischen Osaka, Kyoto und Nara der Betrieb eines japanischen "Silicon Valley". In diesem Zentrum werden die Kyoto-Fakultäten "Computer Science" und "Electrical Engineering" zu einer neuen Fakultät "information science" zusammengeführt. Die Sprachverarbeitung wird hier eine größere Rolle spielen als bisher. Als Ausstattung sind 50 Professoren (full professor) vorgesehen, neue Rechner und neue Software. Die Fakultät soll innerhalb von 5 Jahren aufgebaut werden.

### 4.3.6 Fachliche Diskussion

Die Diskussion im Anschluß an meinen Vortrag entwickelte sich durch lebhaftige Beteiligung der Doktoranden zu einem Kolloquium, in dem die Modellierung von Kohärenz und besonders unsere Idee der modellbasierten semantischen und pragmatischen Verarbeitung von Texten sowie die Methode der Konstruktionsprinzipien diskutiert wurden. Bemerkte wurde, daß unsere Forschungen im Gebiet natürlichsprachiger Texte über derzeitige japanische Ansätze hinausgehen.

### 4.3.7 Kooperation IPSI - Prof. Nagao

Prof. Nagao zeigte sich sehr interessiert an dem Vortrag und hat seinen Besuch für 1989 angekündigt. Interesse besteht auch an einem Austausch von Mitarbeitern.

## 5 Deutsch-Japanisches Forum Informationstechnologie

Das Forum wird von BMFT (D) und MITI (J) organisiert und dient der Förderung deutsch-japanischer Kooperationen in Forschung und Entwicklung. Es war von japanischer Seite hervorragend organisiert. Von besonderem Interesse waren für mich

- . Besuch bei ATR
- . Kennenlernen des zur Diskussion gestellten Themenspektrums
- . Vergleich mit den in meinem Arbeitsgebiet (Verarbeitung natürlicher Sprache) in Japan derzeit brisanten Themen, für die es bislang keine Kooperation mit Deutschland gibt und

die nicht alle auf dem Forum vertreten waren. Vertreten war Parallelismus, es fehlten: gesprochene Sprache, Wörterbücher und Wissensbasen, höhere KI-Gebiete (Text, Bild). (In diesen Bereichen sind die Japaner uns m.E. vor allem im massiven Einsatz von Mitteln voraus.)

- . Kennenlernen möglicher Kooperationsgebiete
- . Kennenlernen japanischer Strukturen (interne Politik, Organisationsformen)
- . Kennenlernen deutscher Japan-Partner und Interessenten

Das Forum war in drei Workshops eingeteilt:

- . Halbleiter
- . Neue Medien
- . Computer

Ich war Mitglied des Computer-Workshops. Von den 28 Teilnehmern waren 11 aus Japan, 17 aus Deutschland. Das Themenspektrum der 13 Vorträge reichte von Betriebssystemen bis zu AI-Anwendungen. Nicht für jedes Thema waren Vertreter beider Nationen anwesend. Inwiefern das Spektrum repräsentativ für Aktivitäten und Kooperationsinteressen auf bei den Seiten war, kann ich nicht beurteilen.

Die drei thematischen Blöcke dieses Workshops waren:

- . Betriebssysteme und Echtzeitsysteme
- . Anwendungen fortgeschrittener Informatik
- . Hochparallele Systeme

Aus der Perspektive meiner Interessen sei auf folgende Vorträge hingewiesen (in der Reihenfolge der Vorträge):

- . T. Ishii (Japanese Patent Office), *Electronic Patent Application System Based on the OSI-Concept*. Vorgestellt wurde die Konfiguration eines Systems zur elektronischen Verwaltung japanischer Patente, das bis 1993 verfügbar sein soll. Schwerpunkt des Projekts ist die Vernetzung einfacher verfügbarer Systemkomponenten (Eingabe, Speicherung, Ausgabe) und die entsprechende Versorgung mit Protokollen. Unsere (IPSI)-Konzeption von integrierter Publikation und Information, nämlich elektronische Unterstützung von der unteren Ebene bis zu Semantik, Pragmatik und entsprechend strukturierter Verwaltung, Zugriffe und Verarbeitung liegt offenbar jenseits dieses Projekts. Nicht einmal einfachere Methoden der Dokumenterschließung des deutschen Standards spielen hier eine Rolle.

- . W. Wahlster (Uni und DFKI/SB), *Intelligent Interfaces für Knowledge Based Application Systems: the Combination of Language Vision and Graphics in Multimodal Systems* gab eine Übersicht der deutschen Forschung im Bereich multimodaler Systeme. U .a. stellte er auch die Ansätze unserer Projekte COGITO und KOMET vor.

- . W. Howein (Siemens), *Benefits of Knowledge Based Techniques in Industrial Automation* stellte Probleme industrieller Anwendungen von KI- Techniken an einem Beispiel dar, das ich beachtenswert fand: Ein Diagnoseexpertensystem, dessen Wissensbasis eine hypertextartige Verbindung zu entsprechend aufbereiteten Dokumenten hat und so dem Benutzer offenes Hintergrundwissen unterschiedlicher Expertisen und Interessen von der Entwicklungsabteilung bis zu den Praktikern zur Verfügung stellt. Er erläuterte u. a. die Probleme des Wissenerwerbs und des Wissenstransfers.

- . T. Higuchi (ETL), *A Semantic Network Machine: IXM* stellte die Entwicklung einer massiv parallelen Hardware vor, die mit großen Assoziativspeichern den zur Verarbeitung semantischer Netze benötigten hohen Zeitaufwand vermindern soll.

- . W.v.Seelen (Universität Mainz), *Computation in Neural Net* zeigte, wie die Beobachtung netzartiger Phänomene technischer, biologischer und neuronaler Systeme integriert und mathematisch modelliert werden kann.

Er zeigte, daß er solche Modelle als Grundlage im Bereich der Bildverarbeitung und im Bereich der Steuerung eines autonomen Fahrzeugs (Verb und projekt der Firma Schenck) benutzt hat.

*Karin Haene/t, IPSI, Institut für Integrierte Publikations- und Informationssysteme der Gesellschaft für Mathematik und Datenverarbeitung (GMD), Dolivostraße 15, D-6100 Darmstadt*

## International Conference "New Directions In Machine Translation"

### Budapest 18./19.8.1988

Veranstaltet von dem niederländischen Softwarehaus BSO und der ungarischen "John von Neumann Society for Computing Sciences" (NJSZT) fand, vor dem Beginn von COLING 88, diese zweitägige Konferenz in Budapest statt. Die Proceedings lagen bei Tagungsbeginn vor.

Den knapp 120 Teilnehmern wurden in vier Sessions jeweils vier Vorträge geboten, sowie, außer einer Eröffnungsadresse am Anfang und dem Versuch einer Zusammenfassung am Ende der Tagung, ein Rahmenprogramm.

Für die einzelnen Vorträge standen jeweils 30 und für die anschließenden Diskussionen jeweils 5 Minuten zur Verfügung.

Die Konferenz diente einerseits wissenschaftlichem Austausch und andererseits der Firma BSO, ihr DLT-Projekt bekanntzumachen, in dem Esperanto als Interlingua eingesetzt wird.

In einigen Vorträgen war Esperanto Thema oder Unterpunkt oder wurde für Beispiele herangezogen. Abweichend von ihren Manuskripten gingen einige Referenten durch kritische Stellungnahmen und Beispiele auf diese Tendenz ein.

### Session I - Overviews

Mr. W.J. Hutchins, University of East Anglia, UK: *Recent Developments in MT - a review of the last five years*

Prof. Dr. Tibor Vamos, Hungarian Academy of Sciences: *Language and the Computer Society*

Dr. I.I. Oubine, USSR Centre for Translation of Scientific and Technical Literature and Documentation: *The State of the Art in MT in the USSR*

Mr. Dong Zhen Dong, China Sw. Technique Corp.: *Esperanto-based MT Research in China*

Nach der Begrüßungsansprache in der die Bedeutung der MT für kleinere Länder (hier insbesondere für die Niederlande und Ungarn) hervorgehoben wurde, wurde von Toon Witkam (BSO) die

1 Dabei handelt es sich um "DLT", ein, sich noch in der Entwicklung befindendes, sehr benutzerinteraktives System, das auch auf der COLEXPO'88 zu sehen war. Der quellsprachliche Text wird mit Hilfe von Fragen an den Benutzer in die Interlingua übersetzt. Kommunikation zwischen verschiedensprachigen Teilnehmern einer Konferenz sollen durch Konferenzschaltungen zwischen Terminals möglich werden

erste Session mit dem Thema "Overviews" eingeleitet, der Umschlag des Buches "Machine translation: past, present, future" wurde an die Wand projiziert.

Dann erschien der Autor dieses Werkes und Chronist der maschinellen Übersetzung W. John Hutchins und hielt sein Referat, in dem er die bisherige Geschichte der MT (ALPAC-Report etc.) zusammenfaßte. Der nächste Referent (Tibor Vamos) holte weiter kulturgeschichtlich aus.

Die beiden nächsten Referenten stellten den Stand der MT in der UdSSR und in China dar, ohne detailliert auf Methoden und Konzepte einzugehen. Statt dessen lag der Schwerpunkt auf Hinweisen auf Traditionen auf dem Gebiet MT und Absichtserklärungen für die Zukunft. Es entstand jedoch der Eindruck, daß die in Gebrauch befindlichen Verfahren mehr auf der Übertragung von Vokabeln als der Analyse von Syntax beruhen.

Bemerkenswert ist der Trend der zunächst auf westliche Sprachen ausgerichteten Arbeit der sowjetischen Projekte, hin zu asiatischen Sprachen, wobei die Akzeptanz gegenüber Rohübersetzungen aus "exotischen" Sprachen als größer eingestuft wird.

### Session II - Interlingual Architecture

Prof. Christian Boitet, GETA/CNRS, Grenoble, France: *Pros and Cons of the Pivot and Transfer Approaches in Multilingual MT*

Dr. Michiko Kosaka, Monmouth College, NJ, USA: *A Sublanguage Approach to Japanese-English MT*

Prof. Ivan Guzman de Rojas, La Paz, Bolivia: *ATAMIRI - Interlingual MT using the Aymara Language*

Dr. Klaus Schubert, BSOjResearch, Netherlands: *The Architecture of DLT - Interlingual or Double-Direct?*

In der Session zum Thema "Interlingual Architecture" stellte Christian Boitet theoretische Erwägungen zur Architektur multilingualer Übersetzungssysteme an, und stellte die Vorteile eines Pivot-Ansatzes heraus. Für ein multilinguales Übersetzungssystem mit z.B. 12 beteiligten Sprachen könnte es auch ökonomisch sein, die drei hauptsächlich genutzten Sprachen als standardisierte zentrale Sprachen zu nutzen und so die Zahl der Übersetzungsvorgänge zu reduzieren.

Michiko Kosaka erläuterte die Probleme bei Übersetzungen aus dem Japanischen ins Englische: Wortklassen des Englischen sind dem Japanischen oft nicht trennscharf zuzuordnen,

stattdessen herrscht ein hohes Maß von Kontextabhängigkeit, das schwierig zu formalisieren ist und einen text analytischen Zugang zur Disambiguierung erfordert. Der vorgestellte Zugang kombiniert Elemente des Interlingua- mit denen eines Transferansatzes.

Ivan Guzman de Rojas (La Paz) berichtete von seinem kleinen aber sich im Gebrauch befindenden System ATAMARI. Spätestens an dieser Stelle, aber auch während des weiteren Verlaufs der Tagung wurde deutlich, daß es als "gut" bezeichnete Systeme gibt, deren Systemarchitektur aber meist nur global beschrieben wird.

Von Klaus Schubert wurde das Projekt der BSO (s.o.) "Distribute Language Translation" (DLT) vorgestellt. In seinem Vortrag stellte er die Vorteile von Esperanto als bewährte, regelmäßige "künstliche natürliche Sprache" und ihrer Eignung zur Interlingua dar, was zu kontroversen Äußerungen unter den Teilnehmern führte.

## Session 111 - Discourse Analysis, Terminology

Dr. Christa Hauenschild Technische Universität, Berlin: *Discourse Structure - Some Implications für MT*

Prof. Dr. Jun-ichi Tsujii, Kyoto University, Japan: *What is cross-linguistically valid interpretation of discourse ?*

Dr. Christian Galinski, INFOTERM, Vienna: *Advanced Terminology Banks Supporting Knowledge-Based MT*

Mrs. Wera Blanke, Terminologia Esperanto-Centro, Berlin: *TEC - Efforts für Terminological Standardization in the Planned Language*

Christa Hauenschild vertritt einen explizit mehrschichtigen Zugang für die Analyse und Synthese, in dem syntaktischer Transfer, Semantischer Transfer und konzeptueller Transfer unterschieden werden. Darüber hinaus illustrierte sie mit Beispielen für unklare Textstrukturen die Relevanz der Diskursstruktur.

Jun-ichi Tsujii problematisierte, auch mit Blick auf verschiedene Systemarchitekturen, den Begriff der Interlingua und setzt ihn zur Diskursrepräsentation in Beziehung. Dabei wurde zunächst auf deren Leistungsfähigkeit und Definition eingegangen und mit Beispielen für Übersetzungsprobleme (englisch - japanisch), die die Referentialität betreffen, illustriert.

Probleme bei Aufbau und Unterhaltung einer Terminologiedatenbank (TDB) und ihre Kosten stellte Christian Galinski dar und wies auf

die damit verbundenen Probleme hin: Das Wissen wächst schneller als die sprachlichen Ausdrucksmöglichkeiten, Galinski nimmt einen Bestand von derzeit 30 Millionen Begriffen an. Der Aufbau einer TDB ist sehr kostenintensiv, aber er ist höchst sinnvoll und sollte von vornherein sorgfältigst gehandhabt werden.

Einen Eindruck von Umfang und Stand der Esperantobewegung und ihrer Terminologiearbeit vermittelte Wera Blanke.

## Session IV - Translation And Linguistics

Mr. Dietrich Weidmann, Schaffhausen, Switzerland: *Universal Applicability of Dependency Grammar*

Prof. Bengt Sigurd, Lunds universitet, Sweden: *Translation to and from Swedish by SWETRA - a Multilanguage Translation System*

Mr. Gabor Proszeky, National Education Library, Budapest: *Hungarian - a Special Challenge to MT?*

Prof. Claude Piron, Universite de Geneve, Switzerland: *What we can Learn from Mistakes made by Professional Translators*

Dietrich Weidmann beschrieb sein dependenzgrammatisches System, das auf den Kategorien "pilot" und "guest". beruht. Danach stellte Bengt Sigurd ein kleines Übersetzungsprojekt vor, das auf 'Referent Grammar', einer Art GPSG beruht, vor. Dieses System soll geeignet sein, Regeln für den syntaktischen Transfer zu formulieren.

Einen kurzen Einblick in den Aufbau der Sprache des Gastlandes, insbesondere in deren Morphologie, gab Gabor Proszeky, der an einer zu den DLT-Spezifikationen passenden Klassifikation des Ungarischen mitgearbeitet hat.

Von Claude Piron, der aus der Praxis des Übersetzers argumentierte, wurde angezweifelt, daß die Arbeit des Übersetzers durch Datenverarbeitung ersetzt werden kann. Seine Beispiele aus der Praxis für schwierige Fälle bezogen sich auf oneto-many-Probleme, Wortstellungsfehlern bei der Übersetzung von Namen von Organisationen, und Disambiguierungsprobleme, die über den sprachlichen Kontext hinaus Hintergrundwissen erfordern. Er wies darauf hin, daß dies die Probleme sind, die den Übersetzer viel Zeit kosten, nicht jedoch die Übersetzung der sonstigen, weitaus größeren Textmengen.

*Michael Schmidtke-Nikella*

## Literatur

NEW DIRECTIONS IN MACHINE TRANSLATION, Conference Proceedings; Budapest 18-19 August 1988, edited by Dan Maxwell, Klaus Schubert, A. P. M. Witkam; Foris Publishers, Dordrecht/ Providence 1988

## Summer School in Computational Lexicology and Lexicography

**Pisa 25. Juli bis 13. August 1988**

Der Tagung vorangestellt war ein einwöchiger Vorkurs mit einführenden Vorlesungen und Übungen z.B. in die EDV (Bandinelli/Laforenza, Pisa), LISP (Goerz, Erlangen), Lexico-Grammatik (Gross, Paris), Grammatik und Lexikon (Hajicova, Prag).

Der zweiwöchige Hauptkurs sollte primär Lexikologen und Lexikographen aber auch andere Linguisten mit Interessen an Textkorpora und linguistischer Datenverarbeitung aus der ganzen EG zusammenführen. Es fanden von Montag bis Samstag täglich acht zweistündige Vorlesungen statt, von denen jeweils zwei parallel gegeben wurden, was die Teilnahme an der Hälfte des Angebotes ermöglichte. Die Veranstalter rieten davon ab, mehr als drei Vorlesungen pro Tag wahrzunehmen, um Zeit für praktische Arbeit zu haben. Ursprünglich sollten 80 Teilnehmern zugelassen werden, es wurden jedoch schließlich über 150, was sowohl die Infrastruktur der Scuola Superiore in Pisa überforderte, in der das Gros der Teilnehmer untergebracht war, als auch zu sehr großen Kursen führte, in denen nicht soviel diskutiert werden konnte wie vorgesehen. Der hier anschließende Bericht kann nur etwas über die von mir selber besuchten Kurse sagen, während die anderen zur Information nur aufgelistet werden. Die wohl best besuchte Veranstaltung waren die zehn "Pisa Lectures on Lexical Semantics" von Charles Fillmore, Berkeley. Es handelte sich um meist in sich geschlossene Vorlesungen über Einzelthemen wie Grammatische Valenz, Semantische Valenz, Kollokationen, Idiome etc., die souverän und locker dargeboten wurden, etwa wie ein Sammler, der durch seine umfangreiche Sammlung streift

und dem Besucher einige Schätze genauer vorstellt und ihre Besonderheiten beschreibt. Ganz anders war die ebenso populäre Vorlesungsreihe von Beryl (Sue) Atkins, Redakteur im Collins Verlag, London, und aktive Lexikographin. Wie schon ihre jedem Teilnehmer zugeschickte vierseitige Ankündigung (mit zu erledigenden Vorbereitungsaufgaben) zeigte, war ihr Vorgehen äußerst systematisch. Atkins behandelte Themen wie den Aufbau eines Lexikoneintrages (der auch professionellen Lexikon-Benutzern noch viel Neues bot), das Erstellen einer Datenbank für Lexika auf der Basis eines Textcorpus (am Beispiel des Cobuild Corpus), die Einrichtung eines Pflichtenheftes für Compiler bis hin zur Aufstellung eines Finanzrahmens für ein zu erstellendes Lexikon. Daneben führte sie praktisch die Anwendung von Computern (PC mit dBase III plus) bei der Bearbeitung von Textbelegen für die Festlegung eines Lexikoneintrages vor. Besondere Höhepunkte dieser an sich schon exzellenten Reihe waren die Vorlesungen mit Alain Duval (Universität Paris und Verlag Le Robert), in denen im Wechselgespräch vorgeführt wurde, wie sich zwei Meister ihres Faches auf Definitionen für ein englisch-französisches Lexikon einigen. Wohl einmalig war in diesem Rahmen, als Charles Fillmore, der die Vorlesung im Publikum verfolgte, von Sue Atkins dazu überredet wurde, seine Ansätze an einem vorgegebenen Korpus zum Wort "risk" vor allen Teilnehmern zu demonstrieren, was er mit verblüffenden Resultaten tat. Es war für alle Anwesenden hoch interessant zu sehen, wie ein "reiner" Linguist sich mit einem realen Korpus, der nicht nur zur Illustration seiner Theorien geschaffen wurde, auseinandersetzt. Solche Veranstaltungen sind wohl einzigartig und aus diesem Grunde beeindruckend, zeigen sie doch, wie unterschiedliche Ansätze und Temperamente bei demselben Problem zu verschiedenartigen Ergebnissen führen. Sehr interessant, wenn auch wesentlich stärker mit der Computer-Seite befaßt, waren die Vorlesungen von Roy Byrd vom IBM-Research Centre in New York. Schwerpunkte waren Speicherungsprobleme und Speicherungsstrategien von großen Textmengen und Lexika. Die etwas überraschten Lexikologen und Geisteswissenschaftler hörten z.T. zum ersten Mal von binären Bäumen und deren Ableitungen, von der Pflege solcher Strukturen, von Patricia Trees, hashing und Datenkompression. Es wurde über bestehende Programme (etwa PAT, welches an der Universität Waterloo in Ontario, Kanada, erhältlich ist) und deren jeweilige Vor- und Nachteile referiert. Roy Byrd führte WordSmith, den Prototyp eines lexikalischen Datenbanksystems vor. Es handelt sich dabei um ein on-line Lexikon, das mehrere Lexika vereinigt: Webster, Longman DCE, ein Synonym-Lexikon, ein taxonomisches Lexikon, ein englisch

italienisches sowie ein englisch-französisches Lexikon. Man kann von einem Lexikon zum anderen wandern, innerhalb der Einträge weitere Begriffe markieren und diese in einem neuen Fenster anzeigen lassen, etc. Auch das von Byrd vorgeführte Programm DAM (Dictionary Access Method) verspricht große Möglichkeiten, Lexika einzurichten, zu erweitern und schnell darauf zuzugreifen. Byrd wollte zu beiden IBM-Produkten keine Angaben machen, ob und wann sie allgemein verfügbar sein würden. (Für Interessenten seine Bitnet Nummer: BYRD at YKTVMH). Bran Boguraev (Cambridge) und Ted Briscoe (Lancaster) stellten in ihrer Reihe "Computational Lexicography and Natural Language Processing" in anschaulicher Form weitgehend den Inhalt ihres zu jenem Zeitpunkt gerade bei Longman erschienen Buches gleichen Titels vor, daß sich mit den Möglichkeiten des Longman Dictionary of Contemporary English befaßt, gingen aber selten über die dort dargestellten Ergebnisse hinaus. Die Vorlesungen von Donald Hindle (AT&T Bell Laboratories) zu "Computational Analysis of Large Corpora" war zu Beginn gut besucht, verlor aber dann, möglicherweise wegen der undeutlichen Vortragsweise, der unpräzisen Gliederung und Formulierung der Vorlesungen sowie mangelnder Praxis-Relevanz schnell Zuhörer. Ebenfalls auf großes Interesse stieß Elisabet Engdahls Reihe über "Current Syntactic Theories and the Polytheoretical Lexicon", die sehr abstrakt dargestellt wurde. Zur Vervollständigung des Eindrucks vom reichhaltigen Angebot der Tagung können die übrigen länger als drei Tage dauernden Veranstaltungen nur genannt werden:

Karttunen (Stanford) und Koskenniemi (Helsinki): "Computational Morphology"; Martin Kay (Xerox und Stanford): "Natural Language and Symbolic Computing"; S. Nirenburg (CMU-Pittsburgh): "Lexicographic Needs of Machine Translation"; J. Pustejovsky (Brandeis University): "Knowledge Representation and the Computational Lexicon"; H. Uszkoreit (IBM Deutschland und Universität Saarbrücken): "Lexicon and Unification". Die Round Tables boten trotz prominenter Besetzung wenig, was nicht an anderer Stelle bereits mehrfach - und meist klarer und kürzer - gesagt worden ist. In der zweiten Woche boten eine Reihe von Teilnehmern abends nach den Vorlesungen spontan eigene Veranstaltungen an. Besonders interessant, nicht zuletzt weil das Programmpaket auf Anfrage erhältlich ist, war der Bericht über die Linguistic Data Base des TOSCA-Projektes in Nijmegen. Es handelt sich um eine Datenbank von syntaktischen Baumstrukturen von mehreren englischen Texten. Die Strukturen können für jeden Satz dargestellt werden; zugleich sind Abfragen etwa von der Art "Zeige alle Subjekte von nicht-finiten Sätzen"

und noch weit komplexere Abfragen möglich. (TOSCA, Instituut Engels-Amerikaans, Katholische Universität, Erasmusplein 1, 6500 Nijmegen).

Hervorzuheben ist die vorzügliche Organisation unter Leitung von Antonio Zampolli; das Tagungsbüro war superb besetzt und stand den Teilnehmern 12 Stunden am Tag mit Rat und Tat zur Seite. Wichtig war die eigens eingerichtete Präsenzbibliothek, in der man eine Reihe von Büchern und Zeitschriften, die nicht an allen Universitäten vorhanden sind, durchblättern konnte, und auf diese Art auf Aufsätze aufmerksam wurde, die man sonst nicht gefunden hätte. Neben dem Kenntniserwerb für alle Beteiligten werden sicherlich die gemeinsamen Mahlzeiten in dem mit Kreuzrippen überwölbten Refektorium in Erinnerung bleiben, weil man dort in entspannter Atmosphäre mit Kollegen aus der gesamten EG (und einigen anderen Ländern) Ideen und Gedanken austauschen konnte.

Insgesamt muß die Tagung als sehr gelungen bezeichnet werden, sowohl was die Themenvielfalt als auch was die Auswahl der Referenten angeht. Solche Tagungen, während derer man mit anderen Wissenschaftlern zusammen lebt und arbeitet, stimulieren und motivieren mehr als das Studium von Büchern und sind durch nichts zu ersetzen. Man muß der European Science Foundation für die Durchführung dieser Summer Schools danken und kann nur jedem wünschen, daran teilnehmen zu können. Im Januar 1989 war noch nicht bekannt, ob eine nächste Summer School durchgeführt wird. Es wäre zu wünschen, daß die verantwortlichen Gremien eine Weiterführung dieses Projektes befürworten.

*Norbert Bolz2, Köln*

## **COLING 88 - 12th International Conference on Computational Linguistics**

**Budapest 22.-27. August 1988**

Aufgrund des Umfangs der Veranstaltung und der thematisch und qualitativ sehr unterschiedlichen Beiträge ist eine globale inhaltliche Einschätzung nur schwer möglich, darum liegt der Schwerpunkt dieses Berichts auf statistischen und organisatorischen Aspekten.

Zu COLING 88 wurden 520 Teilnehmer aus 29 Ländern erwartet. Die Veranstaltung setzte sich aus der Eröffnungsveranstaltung, 137 Vorträgen, vier Panels, einem Plenum und einer kurzen Schlußsitzung zusammen.

<sup>2</sup> Anmerkung: Der Autor dankt der Deutschen Forschungsgemeinschaft, die ihm die Teilnahme an dieser Tagung ermöglicht hat.

Es waren 461 Abstracts eingereicht worden, von denen die meisten aus den USA (94), der Bundesrepublik (62) und Japan (45) kamen.

Jedes Abstract wurde von drei Gutachtern gelesen. Die ausgewählten Vorträge, sollten von den Autoren selbst gehalten werden. Darüber hinaus wurden Vorträge zur Reserve ausgewählt, von denen einige vorgetragen wurden, als die mit erster Präferenz vorgesehen Referate - meist wegen der Abwesenheit der Autoren - nicht gehalten werden konnten.

Auch in dem Tagungsband, in den beide Klassen von Vorträgen aufgenommen worden waren, waren die USA mit 40, die Bundesrepublik (und West-Berlin) mit 22 und Japan mit 18 Papieren am stärksten vertreten.

[Es folgten Frankreich mit 10, Großbritannien mit 8, die Tschechoslowakei mit 7, Italien und die Niederlande mit je 6, die Sowjetunion mit 5, Finnland, Belgien und Schweden mit je 4, Dänemark, Spanien und Schweiz mit je 2, sowie Österreich, Bulgarien, China, Irland und Israel mit je einem Vortrag.]

Zusätzlich finden sich in den Proceedings noch einige, aber nicht alle, der im Rahmen der Panels gehaltenen Vorträge.

Die Vorträge waren zu 54 Sessions zusammengefaßt worden, die sich auf 12 Themengebiete verteilt haben.

*Themen:*

mer, so daß es häufig vorkam, daß Teilnehmer die Räume zwischen den Vorträgen wechselten. Weil dafür keine Zeit vorgesehen war, kam es dabei zu Störungen der Referate.

Die Panels beschäftigten sich mit den Themen:

\* Language Engineering: The Real Bottleneck of Natural Language Processing

\* Parallel Processing in Computational Linguistics

\* Controlled Languages and Language Control

\* The Relation of Lexicon and Grammar in Machine Translation

Die Plenarsitzung fand am Morgen des letzten Tages statt und hatte das Thema "TRENDS AND PERSPECTIVES".

Proceedings: COLING BUDAPEST 1988 Proceedings edited by Denes Vargha, ca. 800 Seiten in zwei Bänden, lieferbar bei: Association for Computational Linguistics (ACL) Dr Donald E. Walker; Bell Communications Research 445 south Street, MRE 2A379, Morristown, NJ 07960, USA.

COLING 90 findet von 20. - 25. 8. 1990 an der Universität von Helsinki statt. Organisator ist die "Research Unit for Computational Linguistics, University of Helsinki, Hallituskatu 11, SF -00100 Helsinki 10, Finland".

*Michael Schmidtke-Nikella*

Die Themengebiete und die Anzahl der Vorträge im Einzelnen:

SEMANTICS	(10 Vorträge)
MACHINE TRANSLATION	(22 Vorträge)
SYNTAX AND MORPHEMICS	(10 Vorträge)
SOFTWARE TOOLS FORMAL	(5 Vorträge)
MODELS DISCOURSE	(17 Vorträge)
LANGUAGE GENERATION	(15 Vorträge)
LEXICAL ISSUES	(7 Vorträge)
UNDERSTANDING AND KNOWLEDGE	(9 Vorträge)
REPRESENTATION	(10 Vorträge)
PARSING	(16 Vorträge)
SPEECH ANALYSIS AND SYNTHESIS	(8 Vorträge)
COMPUTER- ASSISTED LEARNING	(8 Vorträge)

Es war nicht möglich, alle Vorträge zu hören: in der Regel fanden vier Sessions zu verschiedenen Themengebieten gleichzeitig statt. Die einzelnen Vorträge dauerten 30 Minuten, für eine anschließende Diskussion standen 5 Minuten zur Verfügung.

Die Zusammenfassung der Vorträge zu Sessions entsprach nicht immer den Interessen der Teilneh-