

Maschinelle Spracherkennung und Übersetzung

Interview mit Heinz Schwärtzel



Prof. Dr. rer. nat.
Heinz G. Schwärtzel

ist seit 1961 bei Siemens, wo er derzeit die Hauptabteilung Systemtechnologien in der Zentralen Forschung und Entwicklung leitet. In verschiedenen nationalen und internationalen Institutionen ist er als Vorsitzender bzw. Mitglied von Aufsichtsräten und wissenschaftlichen Beiräten vertreten. Der heute 57jährige Diplommathematiker studierte Mathematik und Physik in Saarbrücken, München und Linz. Seit 1986 ist er Honorarprofessor der Universität Linz, seit 1991 auch Honorarprofessor der TU München

Ein Telefongespräch mit Japan und den USA ohne Fremdsprachenkenntnisse der Beteiligten – diese wie Science Fiction anmutende Vorstellung nähert sich der technischen Realisierung. Am 28. 1. 1993 wurde in einem internationalen Forschungsexperiment demonstriert, wie weit diese Technik bereits ist: Übersetzungs- und Spracherkennungssysteme in Pittsburgh, Kyoto und im Siemens-Forschungszentrum München waren miteinander verknüpft, so daß sich Forscher der drei Kontinente in ihren Muttersprachen am Telefon unterhalten konnten.

Redaktion: Herr Professor Schwärtzel, Siemens hat vor kurzem eindrucksvoll demonstriert, wie weit die Technik der Spracherkennung und der automatischen Übersetzung bereits ist. Was zeigte dieses Experiment im Hinblick auf die Fähigkeiten des Unternehmens in diesem Sektor?

Prof. Schwärtzel: Diese Demonstration bewies, daß die Siemens-Forschung auf dem Gebiet der Sprachverarbeitung wohl zu den weltweit führenden Institutionen zählt. Wir wollten drei Ziele erreichen: Zunächst ist festzustellen, daß es der erste echte automatisch übersetzte Telefondialog zwischen Japan, den USA und Deutschland war, der unter den Augen der Weltöffentlichkeit stattfand. Dies kann man als technische Meisterleistung ansehen. Zum zweiten haben hier Forscher aus Deutschland, Japan und den USA gemeinsam in einem globalen Verbund – wobei jeder Partner auf eigene Kosten arbeitete – in einem herausfordernden Projekt zusammengearbeitet. Dies ist, so glaube ich, ein wichtiges Zeichen für einen neue Ära technologischer Kooperationsmodelle in der Triade. Wir sehen darin auch den Beginn einer strukturellen

Veränderung in der Forschungszusammenarbeit, bei der auf Wertgewinn durch globale Kooperationen gesetzt wird. Zudem haben sich an dieser Forschungsdemonstration zwei Forschungsinstitute und ein Industrieunternehmen beteiligt. Dies beweist, daß unser Unternehmen Mut zum Risiko hat; es zeigt aber auch seine Innovationskraft. Und so kann diese Demonstration würdig in die Reihe der technischen Pionierleistungen von Siemens eingereiht werden.

Red.: Wie bedeutsam ist dieses Projekt im Vergleich mit dem, was andere Firmen machen?

Prof. Schwärtzel: Ich kenne eine ganze Reihe von Firmen, die auf diesem Gebiet arbeiten, aber Siemens ist mit seinen Partnern dabei ganz eindeutig in der Spitzengruppe. Es gibt eine Reihe von Firmen und Forschungseinrichtungen, die zu so etwas in der Lage wären, aber sie haben es bisher nicht gemacht. Ich vergleiche dieses Experiment gerne mit den ersten Fernsehübertragungen in den 30er Jahren. Es dauerte hinterher gut 20 Jahre, bis das Medium Fernsehen zu einer weltweit verbreiteten Einrichtung wurde. Und der Beginn des Fernsehzeitalters war ja durch-

aus von einer ähnlichen Atmosphäre geprägt.

Red.: Wer sind Ihre Partner bei diesem Projekt?

Prof. Schwärtzel: Die Demonstration wurde getragen von dem Konsortium C-STAR – Speech Translation Advanced Research, das vor über einem Jahr gegründet wurde. Zusammengeschlossen sind in ihm das ATR – Advanced Telephony Research Institute – in Kyoto, die School of Computer Science der Carnegie Mellon University in Pittsburgh und die Siemens-Zentralabtei-

lung Forschung und Entwicklung. Jede der drei Sprachen, Japanisch, Deutsch und Englisch, wurde jeweils in eine der anderen Sprachen übersetzt. Jeder Partner war sowohl verantwortlich für akustische Erkennung und Analyse als auch für die linguistische Analyse der eigenen Sprache in eine Zwischenform. Die Partner waren auch verantwortlich für das Übersetzen der Zwischenformergebnisse der anderen in die eigene Sprache. Grundsätzlich macht jeder alles, was seine Muttersprache betrifft. Gemeinsam werden gewisse Basisausstattungen der Systeme und die Schnittstellen definiert, an denen die Systeme zusammenarbeiten müssen.

Red.: *Haben alle Partner die gleiche Zwischenform verwendet, und gibt es besondere Schwierigkeiten bei bestimmten Sprachen oder Sprachpaaren?*

Prof. Schwärtzel: Man hat ähnliche Zwischenformen verwendet. Sie wurden unabhängig voneinander gestaltet. Aber die Zwischenform ist eine nach bestimmten Regeln definierte, formalisierte Struktur. Es ist viel einfacher, zwischen solchen formalen Strukturen hin und her zu übersetzen; man geht von der eigenen formalen Struktur in die der Kollegen in Japan oder den USA über und übersetzt von dort in die Sprache des Partners. Es gibt auch eine Annahme, daß eine einzige universale Zwischensprache ausreichen könnte. Allerdings ist das ein noch nicht beschlossenes Forschungsprojekt.

Bezüglich der Schwierigkeiten beim Übersetzen gibt es keine Sprachen, die besonders einfach oder besonders schwer zu übersetzen wären. Die Schwierigkeiten sind zwar teilweise unterschiedlich, halten sich aber insgesamt die Waage. Bei diesem Experiment wurden die Pro-

bleme auch ein wenig eingeschränkt, indem nur eine sehr spezielle Anwendungsdomäne benutzt wurde, das Anmelden zu einer Konferenz. Damit war die syntaktische und semantische Komplexität natürlich relativ gering. Dies zu erweitern, ist eine der Herausforderungen für die Zukunft. Dann muß man auch sehen, ob es schwieriger wird,

von oder in eine Sprache zu übersetzen, die nicht dem indogermanischen Sprachbereich angehört. Aber dies ist nicht zu vermuten, vielleicht mit Ausnahme bestimmter Randgebiete der Sprachforschung.

Red.: *Bei Ihrem Experiment war kurz nach dem Sprechen immer auch eine schriftliche Form der Mitteilung zu sehen...*

METAL – Fachübersetzung aus dem Computer

Vom Prototyp zum Produkt ist oft ein langer und dornenreicher Weg. Zwar gab es bereits 1979 eine erste Version des Sprachanalyse- und Sprachgenerierungssystems METAL, doch waren noch weitere zehn Jahre intensiver Forschung und Entwicklung notwendig, bis man von Marktreife sprechen konnte. Inzwischen gibt es jedoch schon über 40 kommerzielle Installationen dieses Systems. Anwender sind vor allem Dokumentationsabteilungen in der exportorientierten Industrie, da hier meist riesige Mengen technischer Texte übersetzt werden müssen.

METAL unterstützt die gesamte Prozeßkette vom Originaldokument bis zum fremdsprachigen Zieltext: Ein graphisch aufbereitetes Dokument, z.B. ein Handbuch, wird per Programm automatisch in Format- und Layout-Informationen einerseits sowie übersetzbaaren Text andererseits aufgespalten. Der Text kann dann, wiederum per Programm, auf Konsistenz der verwendeten Terminologie ge-

prüft werden; unter dem Damoklesschwert Produkthaftung ist Eindeutigkeit des Ausdrucks in technischen Unterlagen unverzichtbar.

Der deformierte Text wird mit Hilfe einer komplexen Grammatik und eines Fachwörterbuchs übersetzt. Wenn der Zieltext veröffentlicht werden soll, wird die von METAL gelieferte Rohübersetzung vom Anwender redigiert. Zum Schluß wird der übersetzte Text wieder mit den Format- und Layout-Informationen zusammengeführt, so daß das fremdsprachige Dokument wie ein Zwilling des Originals wirkt. Das vielsprachige METAL hat Schnittstellen zu einer Vielzahl von Desktop-publishing-Systemen und ein integriertes Expertensystem zur Aufbereitung neuer Wörterbuch-Einträge. Dank der modularen Struktur können METAL-Komponenten nicht nur für die Übersetzung, sondern z.B. auch für Information Retrieval, Stilprüfung und Indexierung verwendet werden.

Prof. Schwärtzel: Das Erkennen der gesprochenen Sätze und das Übertragen in eine Schriftform ist ja nichts anderes als eine hörende Schreibmaschine. Das ist der erste Zwischenschritt. Aber es ist ein sehr wichtiger Schritt, denn er beinhaltet, daß die gesprochenen Sätze linguistisch, syntaktisch und semantisch richtig verstanden worden sind. Das Übersetzen in eine andere Sprache ist dann schon etwas einfacher. Dies hat sich auch im Zeitaufwand gezeigt, denn das Erkennen hat an die zehn Sekunden gedauert, während die eigentliche Übersetzung dann wesentlich schneller erfolgte.

Wenn ein Satz gesprochen wird, dann erfolgt zunächst eine akustische Analyse – man muß die einzelnen Laute zu solchen Lautgruppen zusammensetzen, daß sie als Wörter erkannt werden. Aber die Bedeutung dieser Wörter hängt vom Kontext ab. Dieser Kontext wird über Suchbäume abgefragt, und die Suchbäume können natürlich ungeheuer groß und komplex werden. Wir haben bei der Demonstration die schriftliche Wiedergabe auch dazu benutzt, die Zeit zu überbrücken und zu zeigen, daß etwas geschieht. Denn die Prozessoren, mit denen wir gearbeitet haben, sind noch nicht so schnell, wie es für eine kommerzielle Anwendung notwendig wäre. Das ist zukünftig auch ein Kernpunkt der Entwicklung.

Red.: *Es reicht aber wahrscheinlich nicht aus, daß die Prozessoren schneller werden; man muß auch an den Algorithmen arbeiten...*

Prof. Schwärtzel: Es geht hier im Besonderen um die Verbesserung der Suchalgorithmen. Es geht auch darum, daß man all diese Phänomene einmal systematisch analysiert. Wenn ich also ein *äh* spreche, eine Sprache ganz

unruhig benutze oder einen Satz abbreche – was passiert dann? Gelingt es dann, eine Klasse oder Gruppe solcher Phänomene wieder durch eine einfache Darstellung zu beschreiben, so erleichtert das nachher die Suche ganz entschieden, d.h., die Interpretation und das Erkennen des Inhalts dieser Sätze.

Red.: Welche weiteren Probleme müssen noch überwunden werden, bevor etwas kommerziellere Versionen von Spracherkennungs- und Übersetzungssystemen zur Verfügung stehen.

Prof. Schwärtzel: Diese Konferenz war ein Forschungsexperiment und zeigte den Stand der Technik und die grundsätzliche Machbarkeit. Aber bis zu einer breiten Anwendung ist noch eine Reihe von Problemen zu lösen. Zu nennen sind Vergrößern des Wortschatzes von 700 auf 3000 Wörter, Erhöhen der syntaktischen und semantischen Komplexität durch Übergang zu umfassenderen Domänen, also nicht nur wie im Experiment, die Anmeldung zu einer Konferenz, Beherrschung der Spontaneität im Dialog – einen Dialog, wie ich ihn jetzt mit Ihnen führe, könnte das System noch nicht übersetzen. Das umfaßt z.B. Gebrauch unerwarteter Redewendungen, Abbruch und Wiederbeginn von Sätzen, Sprechen in unvollständigen Sätzen; dies alles muß noch erschlossen werden. Zudem muß die Robustheit noch wesentlich erhöht werden, die Robustheit gegen Neben- und Störgeräusche, aber auch bei einer wachsenden Vielfalt von Sprechern und Situationen. Schließlich muß, wie bereits erwähnt, das Echtzeitverhalten noch deutlich verbessert werden. Es müssen bessere oder auch neue Algorithmen entwickelt werden, und die Anforderungen an Geschwindigkeit und Speicherkapazität der Computer steigen noch enorm.

Man kann davon ausgehen, daß für die Lösung dieser Fragen noch erheblicher Aufwand erforderlich ist, aber auch eine nicht zu unterschätzende Zeit. Diese Aufgaben erfordern bienenfleißige empirische Forschung, und dazu müssen viele zusammenwirken. Unser Ziel ist daher, das Konsortium um weitere Partner zu erweitern.

Red.: Wie tief geht eigentlich das Erkennen des Inhalts, gehen wir in die Richtung von echtem Verstehen?

Prof. Schwärtzel: Die Verarbeitung geschriebener und die Verarbeitung gesprochener Sprache erfordert im linguistischen Kern das gleiche. Es geht um die gleichen Fragestellungen – wie sieht die Syntax aus, wie sieht vor dieser Syntax die Semantik aus, brauche ich Lexika, muß ich die Bedeutungsvielfalt einschränken und all diese Dinge mehr. Der Schritt vom Erkennen zur Schriftform und der Schritt vom Übersetzen aus dieser Schriftform heraus sind analog zu sehen, wie das Übersetzen von geschriebener Sprache in Systemen wie METAL. Wenn man in große Anwendungsbereiche geht, wird man wie bei METAL mit großen Lexika operieren müssen.

Zum Thema echtes Verstehen glaube ich, daß wir in diesen Forschungsgebieten Abstand nehmen müssen von Begriffen, mit denen wir unser Denken beschreiben. Das System versteht nichts. Es verhält sich nur in manchen Situations-Zusammenhängen so, daß man den Eindruck gewinnt, es verstünde etwas. Das System ist sich nicht bewußt, was passiert. All diese menschlichen Eigenschaften haben wir auch nicht vor, in Technik zu realisieren. Das widerspricht einer ganzen Reihe von Grundgesetzen der Logik. Es handelt sich hier um Systeme, die

bestimmte Eigenschaften simulieren. Aber sie sind programmiert, sie operieren in einem bestimmten formalen Rahmen auf einer wohldefinierten formalen Basis, aber sie haben nicht die Attribute von Bewußtsein und Verständnis.

Red.: Sie haben vorhin das Übersetzungssystem METAL angesprochen. Gibt es etwas für Ihr Projekt, das man von METAL übernehmen kann?

Prof. Schwärtzel: Wir haben Prinzipien übertragen, haben aber nicht speziell die Lexika übernommen. Das METAL-System ist auch viel komplexer, als das Übersetzungssystem, das wir demonstriert haben. Aber wir sind dabei, uns Gedanken zu machen, eine Art akustische Eingabe für Übersetzungssysteme wie METAL zu schaffen. Wir werden damit experimentieren.

Red.: Die Benutzer-Schnittstelle ist für solche Systeme sicher besonders wichtig...

Prof. Schwärtzel: Wir mußten die Erfahrung machen, daß Bedienoberflächen in der Informationstechnik, nicht nur bei Telefonsystemen, sehr komplex geworden sind. Deshalb haben wir vor einigen Jahren auch begonnen, der Systemergonomie einen besonderen Forschungsschwerpunkt zu widmen. Systemergonomie ist eine Kerntechnologie des Hauses geworden, und die Ziele, die wir damit verbinden, sind Bedienoberflächen, die so einfach wie möglich, aber andererseits funktional so angemessen wie notwendig sind. Dies gilt für alle Geräte und Systeme, sei es für die Kommunikations- oder die Automatisierungstechnik, für die Medizin- oder die Bürotechnik. Hierzu wird eine ganze Reihe von Untersuchungen angestellt. Wir haben ein eigenes Ergonomielabor, das diese Untersuchungen durchführt. Stichworte sind dabei Anpaßbarkeit und

Multimodalität der Bedienoberfläche, und bei dieser Multimodalität spielen Sprache und Spracheingabe sowie Wiedergabe von Information in gesprochener Sprache eine bedeutende Rolle.

Red.: Gibt es bereits Beispiele für solche neuen Bedienoberflächen?

Prof. Schwärtzel: Wir sind dabei, für den Bereich Medizinische Technik Bedienoberflächen von diagnostischen Geräten zu erproben und einzuführen. Wir haben spezielle Bedienoberflächen entwickelt für die Nutzung in den großen Entwicklungsabteilungen, in denen viel Software entwickelt wird. Und wir sind auch dabei, bestimmte Bedienoberflächen in der Automatisierungstechnik zu entwickeln, um das Bedienen und Beobachten von sehr komplexen Prozeßabläufen zu erleichtern. Aber alle diese Systeme sind noch in Erprobung. Sie erfordern noch eine lange Untersuchungszeit.

Red.: Kommen wir wieder zu Ihrem Forschungsprojekt zurück. Wie lange arbeitet Siemens bereits auf dem Gebiet Sprachverarbeitung und welche besondere Hürden waren bei der Kooperation zu überwinden?

Prof. Schwärtzel: Siemens ist seit mehr als 15 Jahren in der Forschung auf diesem Gebiet tätig. Sprachverarbeitung erfordert wie alle anspruchsvollen technischen aber auch vielversprechenden Gebiete einen langen Atem, einen langen Vorlauf. Wir haben im Verbund mit deutschen Partnern, z.B. mit Philips, schon eine Reihe von Vorläuferprojekten durchgeführt, bei denen Themen bearbeitet wurden, wie Erkennen einzelner Wörter oder sprecherabhängiges Erkennen gesprochener Sprache, im nächsten Schritt, sprecherunabhängige Erkennung von Wörtern oder, noch ein Schritt weiter, Er-

Telefongespräche mit Japan und den USA ohne Fremdsprachenkenntnisse – diese Vorstellung nähert sich der technischen Realisierung. Am 28. Januar 1993 wurden

im internationalen Forschungsexperiment C-Star Spracherkennungs- und Übersetzungssysteme in Kyoto (Japan), Pittsburgh (USA) und München (USA) und München zusammengeschaltet

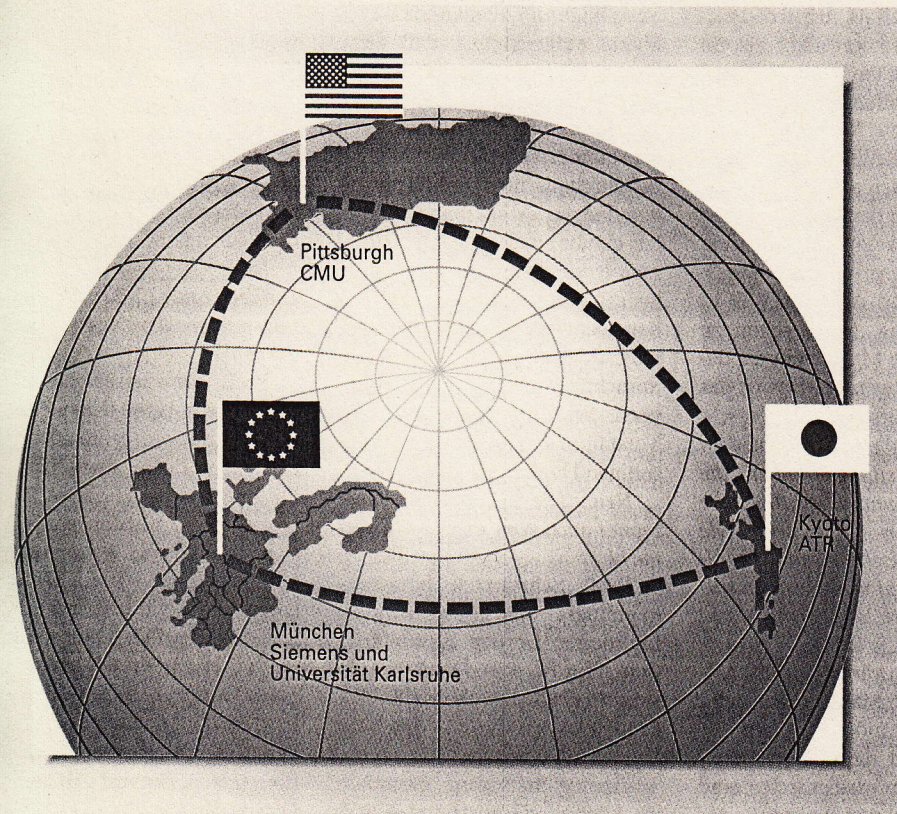


kennen fließend gesprochener Sprache. Das sind alles Schritte, mit denen man sich an dieses Ziel herantasten mußte. Bei der Vorbereitung dieser Konferenz und dem Aufbau des Konsortiums war es notwendig, und das waren nach unserem Empfinden die größten Hürden, Vertrauen zueinander zu gewinnen. Es ist ja nicht ohne weiteres feststellbar, wer gut ist, wer schlechter oder besser ist. Insbesondere in der Zusammenarbeit mit den Japanern war es so, daß wir doch viel Zeit brauchten, um administrative und rechtliche Barrieren abzubauen. Es ist dies auch das erste Mal, daß ein mit öffentlichen Mitteln finanziertes japanisches Forschungsinstitut mit einem Industrieunternehmen außerhalb Japans kooperiert. Aber es gelang!

Red.: Wie sieht überhaupt die Finanzierung dieses Projekts aus?

Prof. Schwärtzel: Das ATR in Kyoto ist im Prinzip eine private Organisation mit erheblicher staatlicher Beteiligung. Es wird jedoch auch von der Industrie gefördert. Aber der größte Teil des Budgets kommt aus öffentlichen Mitteln, im wesentlichen aus einem Fond, der durch die Teilprivatisierung der NTT entstand. Es gab damals ein Gesetz, daß der Erlös aus der NTT-Privatisierung für Forschungs- und Grundlagenarbeiten eingesetzt werden mußte. Und deshalb machte das etwas Probleme.

Die Carnegie Mellon University ist eine private Hochschule, deren Forschungsprojekte allerdings auch in erheblichem Maß von den dortigen Förderern wie der National Science Foundation oder ähnlichen Institutionen und privaten Einrichtungen finanziert werden. Ich gehe davon aus, daß auch die Vorarbeiten zu diesem Projekt über solche Mittel finanziert wurden.



Bei Siemens wurden die derzeit laufenden Arbeiten ohne öffentliche Fördermittel finanziert. Allerdings wurden Vorgängerarbeiten teilweise durch Mittel des Bundesforschungsministeriums gefördert. Insgesamt ist der Anteil öffentlicher Mittel an unserem Forschungsbudget relativ klein.

Red.: *Spracherkennungssysteme sind für verschiedene Anwendungen ja bereits seit einigen Jahren in Betrieb. Wie ist Siemens hier engagiert?*

Prof. Schwärtzel: Wir haben vor etwas mehr als zehn Jahren begonnen, solche Sprachanwendungen zu erproben. Die dabei entwickelten Spracherkennungsgeräte wurden von der Siemens-Gesellschaft CGK – Computer Gesellschaft Konstanz – vermarktet. Es ging dabei um Anwendungen, wo Daten von Mitarbeitern einzugeben waren, die ihre Hände nicht frei hatten. Wir haben auch Experimente für unseren Bereich Medizinische Technik gemacht, z.B. für die Aufnahme und Beurteilung von Röntgenbildern, wo der Arzt die Hände braucht, um die Bilder zu verschieben. Allerdings war dies kein großer geschäftlicher Erfolg. Es war, glaube ich, noch zu früh. Aber wir stellen uns schon vor, daß solche Anwendungen in absehbarer Zeit den Weg in den Markt finden werden.

Red.: *Und wie groß wird dieser Markt sein? Wenn Sie sich so stark engagieren, haben Sie wahrscheinlich auch gewisse Vorstellungen...*

Prof. Schwärtzel: Die Forschung bei Siemens stellt sich immer die Frage, welche Technologien werden die Geschäftsfelder in Zukunft benötigen. Dies ist eine bedeutende, schwierige Frage, denn sie muß beantwortet werden, bevor ein Markt überhaupt existiert. Wir müssen immer wieder Technologien mit

Blick auf einen möglichen Markt bewerten, müssen beantworten, welches Innovations- und Synergiepotential solche Technologien haben. Dies gilt insbesondere, wenn es darum geht, zu entscheiden, ob wir in diese Technik einsteigen sollen oder nicht. Innovationspotential bedeutet, welche Chancen für Neues bieten diese Technologien; es bedeutet aber auch, ob hier wirkliche Durchbrüche enthalten sind. Es gibt eine Reihe von Technologien, denen man dieses zutrauen darf, und bei denen wir uns entsprechend engagieren. An der Spitze steht derzeit natürlich die Mikroelektronik, aber auch die optische Nachrichtentechnik, die wir weiter forcieren, die Ultraschalltechnik oder auch bestimmte Software-Techniken wie Bildverarbeitung. Und beim Synergiepotential fragen wir uns, wie groß ist die Nutzungsbreite bezogen auf die Vielfalt unserer Geschäftsfelder, kann mehr als ein Geschäftsfeld davon profitieren? Auch hier heißen die wichtigsten Technologien Mikroelektronik und Software. Wir glauben allerdings, daß auch Sprachtechnologien in Zukunft ein erhebliches Innovationspotential haben, und deshalb fördern und entwickeln wir sie weiter. Allerdings ist es sehr schwierig, heute schon zu sagen, welche Märkte einmal dahinterstehen werden. Das ist das Risiko der Forschung.

Red.: *Haben Sie trotzdem schon Vorstellungen über die Marktgröße und über mögliche Anwendungen für Systeme wie den elektronischen Simultandolmetscher? Und wann werden solche Produkte auf dem Markt sein?*

Prof. Schwärtzel: Abschätzungen über Marktgrößen sind zur Zeit nicht möglich. Ich möchte auch nicht darüber spekulieren. Trotzdem wird es sicher eine ganze Reihe von Spin-offs ge-

ben, einige davon habe ich bereits erwähnt, z.B. die hörende Schreibmaschine, ein lauschender PC oder auch eine Informationsabruftechnik, mit der man in natürlicher Sprache, z.B. über Telefon, Informationen von Datenbanken abrufen kann. Auch die Steuerung von Maschinen durch gesprochene Kommandos wird immer wichtiger werden. Wir sehen vor allem eine Reihe von Unterstützungsfunktionen als nächste Anwendungen. Denkbar sind auch schriftliche Anzeigen von Rohübersetzungen als Untertitel in Videokonferenzen, so daß man sich trotz Sprachproblemen vielleicht etwas besser unterhalten kann. Aber die Wege dorthin sind zum Teil noch lang. Das Jahr 1996 oder 1997 dürfte der Zeitpunkt sein, zu dem solche Techniken im Markt sichtbar werden. Wir sind auch überzeugt – ich habe ja nur einige der technischen Probleme aufgelistet –, daß diese Technologien zum Übersetzen natürlicher Sprache Technologien für den Beginn des nächsten Jahrhunderts sind. Wir müssen da durchaus noch weiterhin einen langen Atem haben. Aber ich glaube, daß Sprachtechnologien trotz allem ein ganz erhebliches Innovationspotential haben. Sie sind besonders komplex, denn Sprache ist Denken. Sprachübersetzung heißt demnach, Wege zu finden, um das Verständnis in unserem wechselseitigen Denken auszutauschen.

Red.: *Die Marktgröße hängt immer auch vom Preis der Systeme ab...*

Prof. Schwärtzel: Spracherkennungs- und Übersetzungsfunktionen werden Zusatzfunktionen für PC sein. Dann werden sie nur einen Bruchteil des PC-Preises kosten dürfen. Wenn die Kostenentwicklung für diese Technologien soweit gesunken ist, können sie auch den Markt

erschließen. Wenn sie dagegen als Übersetzungs-Server in ein Netz eingebunden werden, kann eine solche Funktion entsprechend teurer sein. Sie wird dann andererseits auch komplexer und aufwendiger sein, weil dann auch das Problem des Vielfachzugriffs gelöst werden muß. Wir müssen jedoch davon ausgehen, daß Sprachübersetzungssysteme eher dem Bereich der Konsumelektronik zuzurechnen sind. Deshalb müssen sie billig sein, denn sonst wird auch kein Markt für sie entstehen.

Red.: *Bei der Vorführung des elektronischen Dolmetschers fiel auf, daß die Sprachausgabe in Deutsch nicht so gut war wie in den anderen Sprachen...*

Prof. Schwärtzel: Wir hatten kein besseres Augenmerk auf den Speech Synthesizer oder Vocoder gelegt, denn das ist im Grund eine wohlbekannte Technik. Wir haben einfach den besten verfügbaren Vocoder verwendet. In den USA, dort wurde ja der beste Synthesizer im Experiment eingesetzt, hat man wesentlich mehr Aufwendungen hineingesteckt. Aber beim nächsten Mal wird der deutsche Vocoder auch besser sein.

Red.: *Gibt es bereits Untersuchungen oder Studien zusammen mit potentiellen Nutzern dieser Systeme?*

Prof. Schwärtzel: Nein, wir haben noch keine Anwendungsuntersuchungen gemacht. Wir brauchen dazu noch bessere Grundlagen und bessere Prototypen. Dann werden wir versuchen, an solche Anwendungen heranzugehen. Die Untersuchungen, die wir bisher machten, waren eigentlich eher sporadischer Art, indem wir einige Siemens-Bereiche angestoßen haben, darüber nachzudenken, z.B. den Bereich Private Netze, der für seine Hicom®-Anlage eine Sprachunterstützung braucht,

die Siemens Nixdorf Informationssysteme AG bezüglich einer PC-Steuerung mit Hilfe gesprochener Kommandowörter, oder auch den Bereich Öffentliche Netze zum Thema Sprachspeichersysteme.

Red.: Neue Techniken haben stets soziale Konsequenzen. Gibt

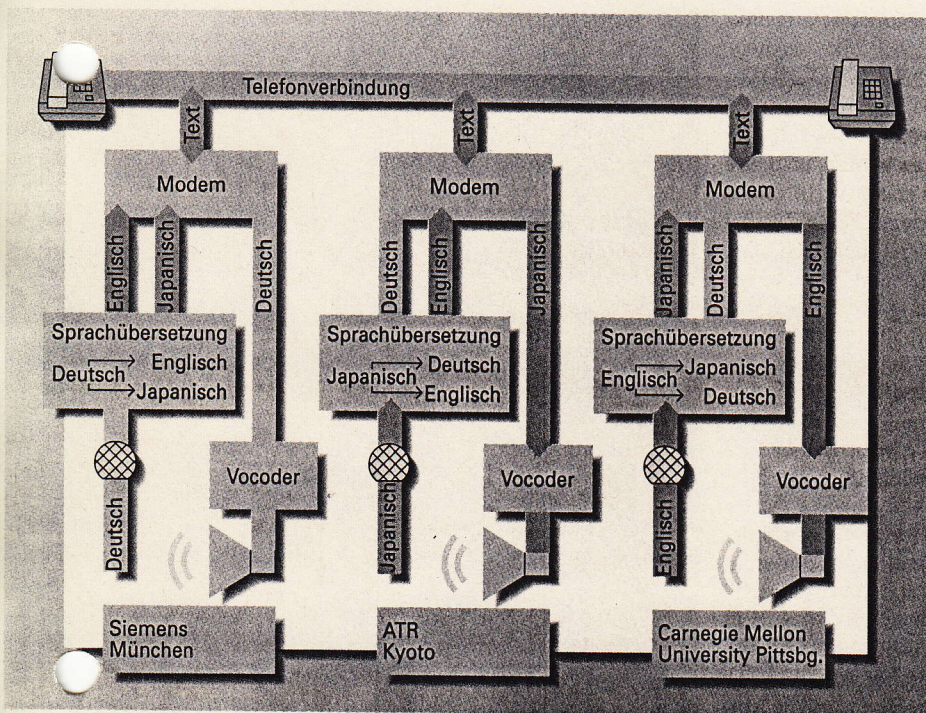
kultureller Verarmung oder gar Vereinsamung, jedenfalls nicht durch Informationstechnik gefördert werden. Wir sind etwas skeptisch gegenüber vielen Untersuchungen zur Technikfolgenabschätzung. Sie haben selten die Antwort gebracht, die man nachher erfahren mußte. Persönlich

ken. Daher sehe ich hier keine negativen sozialen Implikationen. Wir haben allerdings keine systematischen Studien gemacht und sind auch sehr zurückhaltend, über nicht abgesicherte Mutmaßungen Zukunftsszenarien aufzubauen, die sich mit großer Wahrscheinlichkeit als

die wir als System- und Netzsicherheit bezeichnen. In dieser Technologie werden Fragen des Schutzes wie auch der Zuverlässigkeit und Sicherheit von Informationssystemen studiert. Es geht dabei auch um Fragestellungen der Autorisierung, also der elektronischen Unterschrift. Diese Techniken sind grundsätzlich anwendbar und verfügbar. Es ist also nicht die Frage, wie man hier Schuld zuweist, sondern ob man sich selbst mit verfügbaren Techniken ausreichend geschützt hat, wenn man solche Gefahren vermutet. Allerdings könnte es sein, daß sich, wenn solche Probleme auftreten, auch die Rechtsprechung damit auseinandersetzen und entsprechende Lösungen anbieten muß. Ich bin jedenfalls überzeugt, daß man von der persönlichen Verantwortung für Information, die man erzeugt, nicht abgehen soll und kann. Man muß also sicher sein, daß das System nicht manipuliert. Das kann man immer erreichen.

Red.: Lassen sich Spracherkennungssysteme auch zur Identifizierung, z.B. bei Sicherheitssystemen, benutzen?

Prof. Schwärtzel: Wir haben Versuche unternommen, Sprache zur Personenidentifizierung zu benutzen. Allerdings müssen wir feststellen, daß die Robustheit, also die Erkennungssicherheit, noch nicht ausreichend ist. Es gibt jedoch Annahmen, daß die Sprache ein sehr starkes Charakteristikum und persönliches Merkmal ist, mit dem dies möglich sein muß. Die Identifizierungssicherheit der Systeme ist allerdings noch nicht ausreichend – aber die Experimente gehen weiter. Es gibt eine Reihe von Techniken, dort voranzukommen, aber sie sind noch nicht ausgereift. Noch handelt es sich um Forschung, aber der Weg ist bereits vorgezeichnet. •



Technische Struktur von C-Star - Internationales Forschungsexperiment für elektronische Telefon-Übersetzung

es schon Untersuchungen, wie solche Technologien die Gesellschaft beeinflussen?

Prof. Schwärtzel: Wir hätten diese Konferenz auch unter das Motto stellen können: Computer überwinden Sprachbarrieren. Das wäre sicher eine soziale Konsequenz, auf die wir dann hingewiesen hätten. Andererseits sind wir überzeugt, daß die Informationstechnik Menschen tatsächlich einander näherbringt und negative Konsequenzen, man spricht gelegentlich von

bin ich davon überzeugt, daß Sprache so etwas eigenes und persönliches für den Menschen ist, daß auch durch solche Techniken keine Gefahr für sie besteht. Sie entwickelt sich natürlich weiter, auch unter dem Einfluß der Technik, aber sie entwickelt sich so weiter, wie sich auch unser Denken verändert und weiterentwickelt. Im übrigen bin ich der Überzeugung, daß zutrifft, was schon Frege, Wittgenstein und Chomsky formuliert haben: Sprache ist Den-

nicht zutreffend erweisen werden.

Red.: Was ist mit juristischen Konsequenzen, wenn so ein System z.B. für Verträge genutzt wird und das System schiefe oder gar falsche Informationen liefert?

Prof. Schwärtzel: Wir haben eine Kerntechnologie definiert,