

Computer übersetzt Telefongespräche

Elektronischer Dolmetscher

Linguistischer Prozessor kennt schon
700 Wörter und 400 grammatikalische
Regeln – Von Egon Schmidt

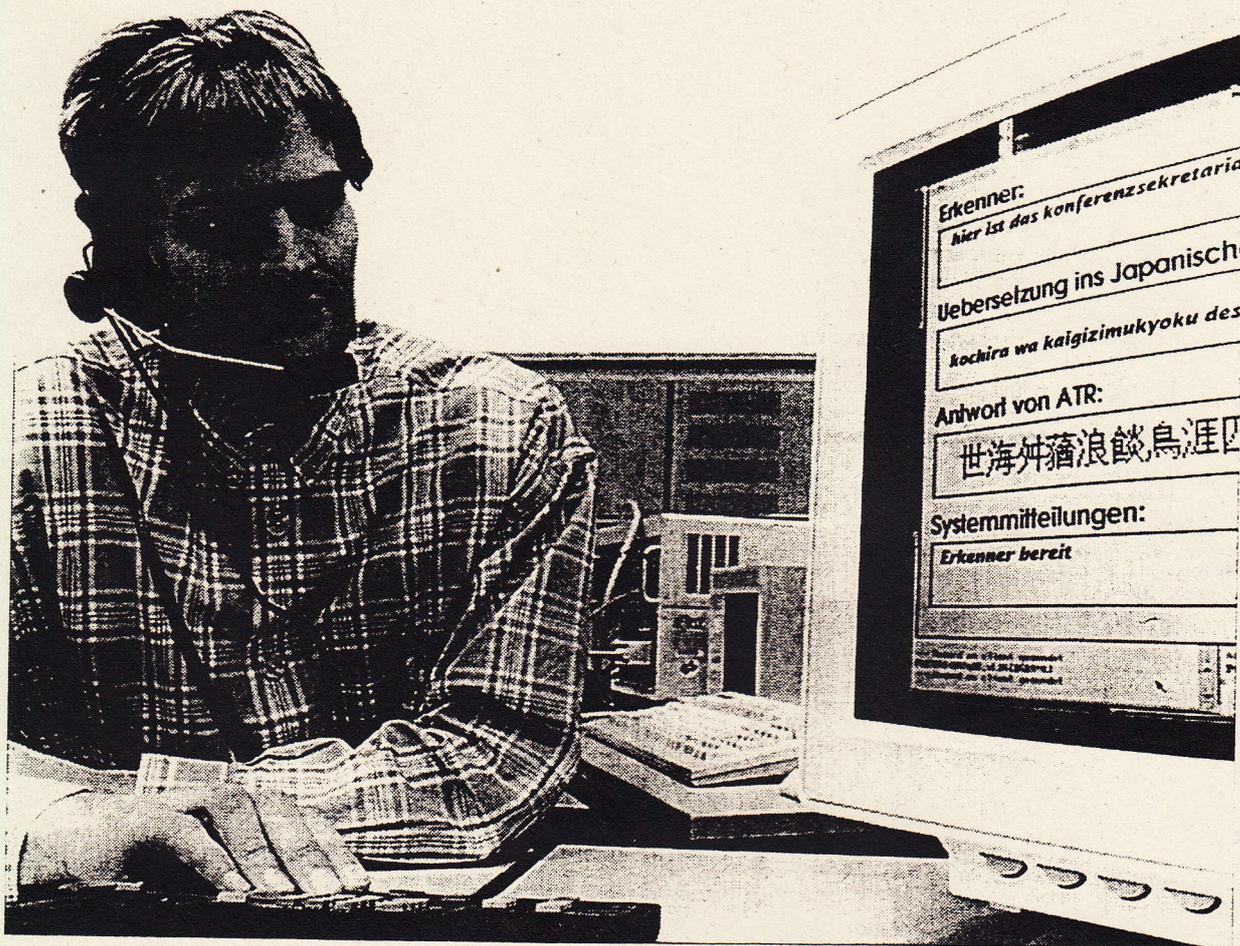
*VDI-N, München, 12. 2. 93 –
Es ist kein Traum mehr, daß je-
mand in München auf Deutsch ins
Telefon spricht und im fernen Ja-
pan nach ein paar Sekunden die
Übersetzung zu hören ist. Kürzlich
bewies eine Vorführung den Fort-
schritt bei Maschinen, die Gespro-
chenes verstehen, es übersetzen
und in einer Fremdsprache synthe-
tisch hörbar machen.*

Vermutlich wird man eines Tages als bedeutenden Meilenstein der Technik feiern, was der Presse am 28. Januar im Siemens-Forschungszentrum im Münchener Stadtteil Neuperlach vorgeführt wurde: nämlich Bildtelefon-Verbindungen über ISDN-Leitungen zwischen München, Pittsburgh und Kyoto, bei denen die jeweils flüssig und keineswegs mehr isoliert-abgehackt gesprochenen Sätze der Bayern, Amerikaner und Japaner maschinell in die Sprache ihrer Gesprächspartner übersetzt wurden.

Die eindrucksvolle Demonstration in München am 28. Januar zeigte technische

Fortschritte, die neben den Siemens-Forschern auch Kollegen von der Universität Karlsruhe, der Pittsburgher Carnegie-Mellon-Universität und des Internationalen Forschungsinstituts für Fortschrittliche Telekommunikation (ATR) in Kyoto sich an die Fahnen heften können. Sie zeigte in Ausschnitten, daß die heutige Technik immerhin schon mit zwölf kurzen Dialogen von je rund zwölf kürzeren, aber auch längeren Frage-Antwort-Sätzen zurechtkommt und dabei besonders „den natürlichen Sprachfluß verstehen kann“, wie Siemens-Manager Prof. Heinz Schwärzel hervorhob. Dies nämlich sei „bedeutend aufwendiger als das einfache Erkennen einzelner Wörter“, das nur gelinge, wenn in unnatürlich-abgehackter Weise gesprochen werde.

Schwärzels Unternehmen ist im Rahmen dieses weltumspannenden Vorhabens zur Entwicklung von Techniken, die sprecherunabhängig funktionieren und sogar ohne kurze Trainings-Phasen zum besseren Erkennen der Stimme des individuellen Benutzers auskommen sollen, in wesentlichen Bereichen engagiert. Die Münchner entwickeln nämlich das Teilsystem zum akustischen Erkennen, zum linguistischen Analysieren und schließlich



Nachdem ein deutscher Satz gesprochen wurde, erscheint wenige Sekunden später der Schriftzug auf dem Bildschirm des Übersetzungssystems. Unmittelbar danach ist auch die Übersetzung ins Japanische ablesbar, die per ISDN-Leitung übertragen und mit einem Synthesierer hörbar gemacht wird.

Foto: Siemens

auch zur akustischen Synthese speziell der deutschen Sprache.

Sieht man von gewissen Schwächen der deutschen, im übrigen nicht von Siemens selber stammenden – und technisch nicht weiter problematischen – Sprachsynthesier-Einheit ab, so verlief die Vorführung so eindrucksvoll, daß Siemens-Forscher Peter Kleinschmidt gern seine Worte abgenommen wurden, es existiere derzeit „weltweit nichts Besseres“. Denn trotz der vorführungsbedingt schwierigen Umstände arbeiteten die Systeme offenbar bestens. Sie zeigten auch, daß das erst gut ein Jahr lang arbeitende, deutsch-amerikanisch-japanische „Consortium Speech Translation Advanced Research“ (C-STAR) in erstaunlich kurzer Zeit viel auf die Beine gestellt hat.

Während Sprachsynthesierer heute in jedem besseren, „sprechenden“ Taschenwörterbuch zu finden sind und während auch das Erkennen gesprochener Einzelwörter heute schon mehr und mehr Standard wird – so etwa beim Steuern von Maschinen durch kurze Befehlsörter – ist das Erkennen fließend gesprochener Rede eine immense Aufgabe. So arbeitet Siemens bei seinem Statistischen Sprach-Erkennen (SSE) mit einem selbst entwickel-

ten Gerät, das dank hauseigener Asic-Bausteine 100 Mips (Millionen Instruktionen pro Sekunde) an Leistung aufweisen und damit jeden Personalcomputer um Längen schlagen soll. Trotzdem braucht es noch einige Sekunden, ehe ein gesprochener Satz so klar erkannt ist, daß er in Schriftform umgewandelt und dann weiter an das eigentliche Übersetzungssystem geschickt werden kann.

Dieses Übersetzungssystem heißt bei Siemens Linguistischer Kernprozessor, kurz LKP. Laut Kleinschmidt umfaßt er zur Zeit „mit 400 Grammatik-Regeln die gebräuchlichsten grammatikalischen Konstruktionen der Umgangssprache“, wodurch er eine Fülle von gesprochenen Sätzen „in eine sprecherunabhängige, logische Form“ überführen könne, aus der heraus dann wieder Gesprochenes generiert werden kann. Dabei ist anzumerken, daß diese sprecherunabhängige, logische Form in München, Pittsburgh und Kyoto vorerst ein wenig unterschiedlich aufgebaut ist, aber diesen Übersetzungsprozeß nicht zu stören scheint. Bei seiner Entwicklung stand das Übersetzungssystem Metal Pate, das für geschriebene Texte geeignet ist.

Beim heutigen Stand der Dinge kann

der Textübersetzer mit immerhin schon 700 gängigen Wörtern umgehen, wobei andere Wörter, Pausen sowie Füllwörter „äh“ vermieden werden müssen; denn sie würden das System, das nur eindeutig und grammatikalisch korrekte Sätze verarbeiten kann, irritieren. Doch im Rahmen des von Bonn geförderten Vorhabens „Verbmobil“ soll nun schrittweise eine robuste Technik entwickelt werden, die auch Spontansprache sicher erkennen und übersetzen kann.

Bei der Vorführung in Neuperlach konnte man direkt miterleben, wie ein Siemens-Forscher jeweils einen Satz in ein Mikrofon sprach, worauf dieser nach einigen Sekunden als Schrift auf einem Bildschirm erschien und bestätigt wurde, daß der Spracherkennung den Text richtig erkannt hat. Dann wurde daraus fast ohne Zeitverlust der japanische Satz erzeugt – in deutschen Schriftzeichen – gleichfalls auf dem Bildschirm erschien. Unmittelbar nachdem dieser japanische Text via Modem nach Kyoto geschickt wurde, dauerte es dort dann noch einige Sekunden, ehe der dort stationierte Synthesierer für die japanische Sprache daraus einen hörbaren Satz geformt hatte, den die anwesenden direkt verstehen konnten.