



UNVOLLENDET blieb der Turm zu Babel, weil die Menschen sich plötzlich nicht mehr verstehen konnten. Das babylonische Sprachgewirr mit den Mitteln der Technik auflösen soll ein elektronischer Dolmetscher.

# Dolmetscher im Taschenformat

„Halloisdortdaskonferenzsekretariat“ – so oder ähnlich hört sich wohl der Satz für jemanden an, der von der deutschen Sprache keine Ahnung hat. Abgesehen davon, daß die Bedeutung des Satzes unklar bleibt: Der Unbedarfte kann daraus nicht erkennen, daß dies eine Frage ist, die aus fünf Worten besteht. Genauso unbedarft aber ist der Computer, der bei der Firma Siemens steht und Teil eines Übersetzungs-Telephon ist. „C-Star“ ist eine Entwicklung von Siemens und der Universität Karlsruhe sowie von japanischen und amerikanischen Forschungsinstituten. Anfang 1993 soll der elektronische Dolmetscher drei „Knoten“ in Deutschland, Japan und den USA verbinden und es Menschen ermöglichen, sich zu einer Tagung in einem Land anzumelden, dessen Sprache sie nicht sprechen.

Mit C-Star üben die Wissenschaftler, was in „Verbmobil“, einem vom Bundesforschungsministerium (BMFT) finanzierten Projekt, entwickelt werden soll. Über 30 Arbeitsgruppen von Hochschulen und Industrie haben es sich zum Ziel gesetzt, in den nächsten acht Jahren einen tragbaren Dolmetscher zu konstruieren, der gesprochene Sprache versteht, sie übersetzt und innerhalb kürzester Zeit den Inhalt akustisch wiedergibt; ein ehrgeiziges Ziel, wie Martin Kay von der Stanford Universität meint. Der Linguist, der ein Gutachten für das BMFT verfaßt hat, plädierte trotz der Hürden auf dem Weg zum Dolmetscher im Taschenformat dafür, das Projekt zu fördern.

## Ohne Pausen

Wenn Menschen sprechen, entsteht, technisch gesehen, ein akustisches Signal, das sich mit der Zeit verändert. „Um diese Schwingungen im Computer zu verarbeiten, wandelt man sie in ein Frequenzsignal um“, erklärt Siemens-Ingenieur Otto Schmidbauer. Das menschliche Ohr ist nicht für alle Frequenzen gleich sensibel: Zum Beispiel werden tiefe Töne besser gehört als hohe. Solche psychoakustischen Effekte ermöglichen es, die Menge an Daten, die in den Tönen steckt, etwas zu reduzieren, ohne dabei an Information einzubüßen. Aus den Frequenzen oder Sprachmerkmalen berechnet ein sogenannter akustisch-phonetischer Prozessor eine Folge von Lauten (Phoneme). Dies geschieht an Hand eines Lexikons,

das rund 45 Phoneme und deren typische Frequenzsignale gespeichert hat.

Mit der Folge von Phonemen ist es aber nicht getan. Worte müssen erkannt werden und zunächst einmal die Wortgrenzen. Sie aufzuspüren bereitet den Maschinen große Schwierigkeiten. Für einen Menschen dagegen ist das kein Problem. Er hört einen gesprochenen Satz als Folge von Wörtern, getrennt durch kurze Sprechpausen. In Wirklichkeit gibt es aber keine Pausen in einem normalen Redefluß. Aus dem akustischen Signal jedenfalls lassen sich die Wortgrenzen nicht erkennen.

„aaanaaiissskkrrriim“ hat der akustisch-phonetische Wandler gerade als Lautfolge ausgemacht. „Bei der Wortsuche läßt sich der Computer von einem Aussprachelexikon und einem Sprachenmodell leiten“, berichtet Schmidbauer. Im Lexikon gespeichert sind Worte und deren Lautsprache, welche die Forscher an Hand der Aussprache zahlreicher Redner ermittelt haben. Das Sprachenmodell spiegelt die Grammatik wieder, indem es vorgibt, welche Ausdrücke aufeinanderfolgen können. Beim Erkennen eines Satzes „tastet“ sich das Computerprogramm an der Phonemefolge entlang und berechnet an Hand von Wahrscheinlichkeiten die plausibelste Lösung für „aaanaaiissskkrrriim“.

„Der Aufbau unseres Systems ist ziemlich typisch für elektronische Spracherkenner“, beschreibt Schmidbauer. Wesentlich sei der integrierte Ansatz. Integriert heißt, daß die einzelnen Ebenen – Phoneme, Wörter, Sätze – parallel aktiv sind und daß auch Rückkopplungen zwischen ihnen erlaubt sind. „Früher hat man versucht, hierarchisch vorzugehen, von unten nach oben, von Lauten zu Sätzen. Damit sind die Sprachforscher gescheitert.“ Als wenig erfolgreich haben sich auch die sogenannten wissensbasierten Systeme erwiesen, Modelle, welche die Leistungen des Gehirns durch starre Regeln beschreiben. „Sprachsignale sind zu kompliziert, um sie mit Regeln zu beschreiben“, bemängelt der Siemens-Forscher.

Was sein Modell allerdings auch nicht kann, ist „aaanaaiissskkrrriim“ eindeutig zu interpretieren. Soll mit der Lautfolge „An ice cream“ gemeint sein, oder „and I scream“? Derartige Ambiguitäten lassen sich größtenteils nicht aus der Sprache,

sondern nur aus dem Kontext heraus lösen. Das ist mit ein Grund dafür, warum die Konstrukteure das Repertoire von Wörtern einschränken müssen, wenn sie erreichen wollen, daß ihr Erkennungssystem zuverlässig arbeitet. Rund 500 bis 1000 Wörter können heutige Programme bewältigen, einen Bruchteil der zwanzig- bis dreißigtausend Wörter, die ein Durchschnittsmensch zu seinem aktiven Wortschatz zählt. Ein belesener Zeitgenosse schöpft gar aus einem Repertoire von 50 000 bis 100 000 Wörtern.

## Tricks beim Diktiergerät

Zwar gibt es Systeme, die mehr als tausend Wörter erkennen können. So stellte die Firma IBM am vorgestrigen Dienstag die deutsche Version eines im amerikanischen Forschungslabor Yorktown-Heights (New Jersey) entwickelten Diktiergeräts „Tangora“ vor, das 20 000 Wörter mit einer Treffgenauigkeit von 95 Prozent unterscheiden kann, was unter Experten als geringe Fehlerquote gilt. Vom Prinzip her arbeitet Tangora ähnlich wie das System von Siemens. In zwei Punkten unterscheidet es sich jedoch wesentlich.

Das Gerät wird mit Sätzen eines Sprechers trainiert und kann auch nur dessen Sprache zuverlässig erkennen. Diktiert einmal nicht der Chef persönlich, sondern dessen Stellvertreter, kommt es schnell zu Verwechslungen. Ein weiterer Trick erlaubt den großen Wortschatz: Um das Problem der Wortgrenzen zu lösen, müssen die Texte mit deutlichen Pausen zwischen den Wörtern gesprochen werden. „aaanaaiissskkrrriim“ ist dann nicht mehr erlaubt, sondern nur noch „aan iisskkrrriim“ – womit die Zweideutigkeit aus der Welt geschafft ist.

Die künstlich getrennte Sprache mag fürs Diktieren noch brauchbar sein. Im Alltag redet man aber nicht so. Deshalb arbeiten die Verbmobil-Forscher auch verstärkt daran, spontan gesprochene Sprache zu erkennen und zu übersetzen. „Ähm“ und „eh“, Räuspern und Zögern muß der tragbare Dolmetscher meistern können, ebenso die syntaktisch falschen Sätze, die oft im natürlichen Redefluß vorkommen.

„Spontane Sprache ist das größte Problem bei Verbmobil“, meint Alexander Waibel. Der Wissenschaftler von der Uni-

versität Karlsruhe hat mit seiner Arbeitsgruppe ein Übersetzungssystem entwickelt, das neben den herkömmlichen statistischen Methoden auch neuronale Netze verwendet, Rechenverfahren mit vielen Elementen, die parallel arbeiten, ähnlich wie die Nervenzellen im Gehirn. „Neuronale Netze lassen sich leicht trainieren“, so Waibel. Zudem eigneten sie sich für den integrierten Ansatz, da sie mehrere Wissensquellen, etwa Aussprache und Grammatik, gleichzeitig verarbeiten könnten.

Bei der automatischen Übersetzung stoßen die Forscher, neben den Problemen mit der spontanen Sprache, auf weitere Schwierigkeiten. „Der englische oder japanische Satz muß nicht nur den gleichen Wahrheitswert wie der deutsche haben“, erläutert der Linguist Hans-Ulrich Block von Siemens. „Er sollte ihm auch stilistisch gleichen. Das heißt, in einem höflichen Dialog darf Verbmobil die Anforderung an den Gesprächspartner, sich zu setzen, nicht mit „setzen“, sondern mit „nehmen Sie Platz“ dolmetschen.

Um Fachliteratur zu übersetzen gibt es seit längerem Computerprogramme. Sie arbeiten jedoch mit zwei erheblichen Einschränkungen. Sie sind in ihrem Wortschatz auf das Fachgebiet limitiert. Zudem liefern sie entweder nur eine Rohübersetzung, die noch korrigiert werden muß, oder sie bedienen sich einer eingeschränkten Grammatik. „Für das Dolmetschen eines Gesprächs sind diese Begrenzungen inakzeptabel“, sagt Block. Andererseits sind die Forscher weit davon entfernt, die Probleme, die beim Übersetzen entstehen, lösen zu können, mehrdeutige Sätze zum Beispiel. „Karl beobachtet den Mann mit dem Fernrohr“ – das Fernrohr könnte sowohl Karl als auch dem Mann gehören.

Wenn ein Satz dann übersetzt ist, muß der geschriebene Text wieder zu Sprache werden. Wer einmal einen Computer hat sprechen hören, weiß, daß man ihn zwar versteht, daß er aber unnatürlich klingt. Der synthetischen Sprache den menschlichen Touch zu geben, gehört daher auch zum Verbmobil-Projekt. Bisher hat man sogenannte Generierungsmuster verwendet, starre Regeln, die das Spontane an der Sprache nicht genügend berücksichtigen. Klang und Melodie bemühen sich jetzt Sprachforscher zu modellieren.

JEANNE RUBNER