

Mit Humor, Höflichkeit und Zwischentönen kommen Übersetzungscomputer noch immer nicht zurecht

Unberechenbare Sprache

Von Daniel Perrin

Kayoko Miyama kichert. Wieder habe ich mich vergriffen zwischen *iu*, *ossharu* und *mooshiageru*, drei Verben, die alle „sagen“ bedeuten. Sag's doch schnell per Telephon? Nicht im Ferngespräch nach Japan! Zur Wahl stehen immerhin: eine Ehrerbietungsform, eine Empfängerhöflichkeitsform und eine Hörerhöflichkeitsform. Patzer wirken so peinlich wie ein „Hallo, altes Haus!“ im Bewerbungsgespräch.

Mit dem automatischen Telephonübersetzer soll nun alles gut werden. Im Advanced Telecommunications Research Institute International (ATR) in Kyoto forschen Physiker, Mathematiker, Informatiker, Linguisten und Kommunikationswissenschaftler an der „Überwindung der sprachlichen Barrieren“ und der „Erleichterung der Kommunikation unter den Völkern dieser Welt“, wie das Institut seine Ziele beschreibt. Wenn die Maschine fertig ist, soll sie gesprochene Sprache im Telephonnetz fließend und ohne Zeitverzug übersetzen. Deutsch, Englisch und Japanisch sind vorerst vorgesehen als Sprachen der Völker dieser Welt. Fürs Ausländische hat man potente Partner gewonnen: Die Probleme des Computers mit der englischen Sprache versucht die Carnegie Mellon University (CMU) in Pittsburgh zu lösen, Siemens in München kämpft gegen die Krämpfe der Maschine mit dem Deutschen.

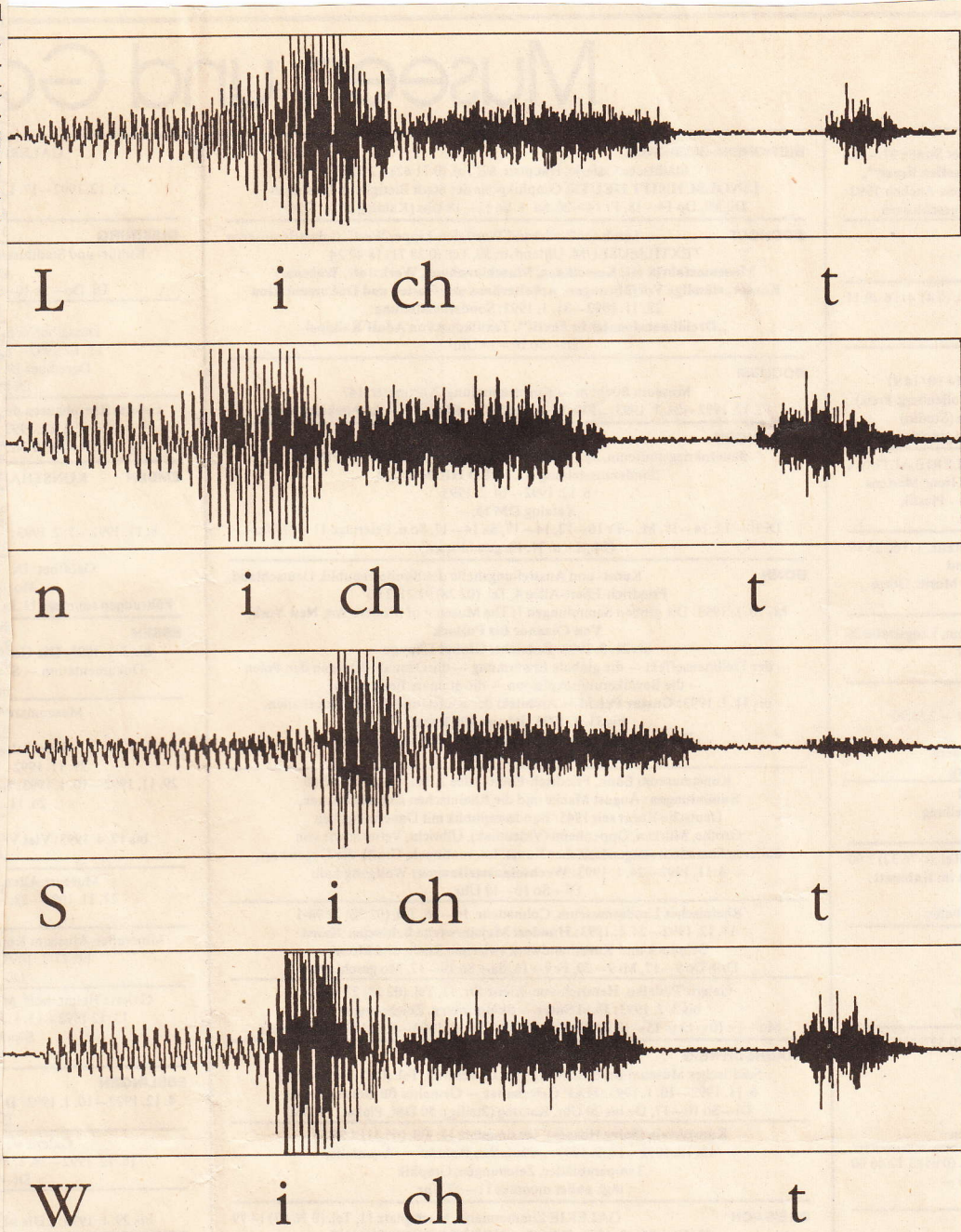
Die gesprochene Sprache narrt den Computer, der nie verstehen kann und doch erkennen soll: Watisdat? Wosischdess? Hat Susi den schönsten Mund weit und breit oder den schönsten Mund, weit und breit? Wie wacklige Seilbrücken führen Lautreihen über Abgründe der Unklarheiten und Mißverständnisse. Persönlich gefärbte Sprechweise oder versteckte Sinnzusammenhänge verwirren die lautsprachliche Verständigung in jeder Sprachgemeinschaft. Dazu kommen die Tücken der einzelnen Sprachen. Das Deutsche sperrt sich dem Zugriff der Maschine mit Wortzusammensetzungsmonstern wie eben und einem barocken Schatz an Flexionsformen. Was Englisch immer *spoke* heißt, lautet hier sprach, sprach, sprach oder sprachen. Selbstverständlich überlagern sich die Probleme. Die toten Sprachen, die Toten sprachen, die Totensprachen – bei Hades und Persephone, alles ist möglich! Aber das erste kommt wohl häufiger vor als die anderen zwei: *di:to:tən|spra:xən* meint mit größter Wahrscheinlichkeit „die toten Sprachen“. Dies nun hilft leblosen Sprach-„Erkennern“ weiter. Soll der Computer Lautketten identifizieren, vergleicht er die Eingabe mit seinem Mustervorrat, findet ähnliche Formen und greift daraus die wahrscheinlichsten. Er versteht die Sprache nicht, er mißverstehet sie nicht, er zählt und rechnet sie durch.

Das klingt so technisch und ist doch beim Menschen geklont. Beim telephonierenden Menschen zum Beispiel: Bitte, wie war Ihr Name? Sie sehen mich nicht, in Ihr Ohr schwappte etwas wie Peräng, aber wer heißt schon Peräng, sieht ja aus wie ein Crashtest. Sie bitten, doch zu buchstabieren: Paula-Emil-Richard-Richard-Ida-Nordpol – aha, Perrin, französisch! In Deutschland versteht man meinen Namen am Telephon nur mit Erklärung, in Japan sowieso nicht, in Frankreich aber, wo er geläufig ist, sofort.

Das liegt bestimmt nicht am Telefonsystem und auch nicht daran, daß wir mit fremdsprachigen Ausdrücken prinzipiell nichts anzufangen wüßten. Nein, wir erwarten Peräng einfach nicht, wissen der unscharfen Lautfolge kein Schriftbild

zuzuordnen. So hilflos würde sich ein Spracherkennungssystem an den Grenzen seines Programms fühlen, könnte es fühlen. Denn was uns Umwelterfahrung und Gespür, sind dem Computer Sprachdatenbank und Statistik. Hier findet er Lautbilder, Lautfolgen, Wörter, Wortfolgen und Sätze programmiert, vorgeschrieben. Bei jedem Element stehen Wahrscheinlichkeitswerte für das Vorkommen des Elements an sich und im Umfeld seiner Nachbarelemente. Nach Herr heißt es wohl eher Perrin, nach Bundeshauptstadt eher Berlin. Da reihen und beeinflussen sich die Wahrscheinlichkeiten in Ketten, der Weg zur gesamthaft plausibelsten Variante führt über einen Berg Mathematik. Wie man die Aufgabe löst, hat Andrej Andrejewitsch Markow vorgerechnet. Markow starb 1922; Hidden-Markow-Modelle kommen seit den achtziger Jahren für Natural Language Processing (NLP, Verarbeitung natürlicher Sprachen) zur Sprache.

Wie verläuft der Prozeß? Ich will Kayoko mitteilen, ich sei heute in Bombenstimmung zum Arbeiten, und wähle dazu unvorsichtigerweise genau dieses Bild, sage also in die Sprechmuschel: du:ʔiçbinhøøtə... Der digitale Telefonapparat übersetzt die Schallschwingungen in Ziffernreihen, je nach Übertragungsqualität in Tausende bis Millionen von Nullen und Einsen pro Sekunde. Aus diesem Ziffernstrang pickt sich nun der Übersetzungscomputer periodisch Stichproben: beispielsweise sucht er jede Hundertstelskunde nach Extremwerten. So speckt er den Datenwanst ab zum Umriß der Lautkette. Noch fließt ein Laut in den anderen über. Aber schon setzt das Programm Schablonen an. Da, am Anfang, paßt doch das d, und zwar die gespeicherte Variante 128! (Für Fachleute: ein Allophon (d) des Phonems/d/) Allerdings könnte es auch noch die 34 von t sein oder, im dritten Rang, ein f. Falls d 128 zutrifft, müßte der Laut bis hier gehen, und der neue beginnt dann mit jener Kurve dort – ja, wieder sitzt eine Schablone, u 19. Nachher kommt Pause, das bedeutet vielleicht nichts, vielleicht Wortende, vielleicht Satzende. Programmmanager an Worterkennungsschleife: Die Lauterkennungsschleife hat mit höchster Wahrscheinlichkeit die Folge d-u-Pause identifiziert:



licht, nicht, Sicht, Wicht: Wenn der Computer die Schallfrequenzen dieser Wörter untersucht, unterscheiden sie sich für ihn kaum, außer beim „t“. Welches Wort gesprochen wurde, muß der Rechner aus dem Sinn des Textes ergründen – und das fällt ihm schwer

„In wenigen Jahren schon werden uns viele Maschinen aufs Wort gehorchen“

Ist d-u ein Wort? Ja, ist eins. Vermutlich steht als erstes Wort der Aussage also „du“. Der weitere Verlauf der Analyse bekräftigt diesen Verdacht: dui, duich, duichb undsoweiter sind im Wörterbuch nirgendwo als deutsche Wörter gespeichert. Die Satzschleife müht sich dann lange vergeblich, Muster für die Wortfolge du-ich-bin . . . zu finden, erkennt aber nach „Bombenstimmung zum Arbeiten“ die Aussage als vollständigen Satz mit vorgestelltem Anredewort, genau nach einem der gespeicherten Satzbaumuster.

Als vorläufig letzten Schritt prüft die Bedeutungsschleife im Textzusammenhang, auf welche Personen die Pronomina du und ich verweisen, ob auf Angesprochene und Sprecher – oder auf Dritte, wie in direkter Rede des Typs „Peter sagte: Bin ich umständlich!“.

Bis hierher hat das System mit jedem Schritt neue Varianten in Betracht gezogen. Jetzt zieht es Bilanz: Je weniger Widersprüche innerhalb einer

Variante die Analysen auf den Ebenen des Lauts, des Worts, der Satzstruktur und der Bedeutung aufwerfen, desto wahrscheinlicher erscheint diese Variante. Die wahrscheinlichste wird als Tiefenstruktur – in einer Art internationaler Ziffernsprache – abgelegt und von hier aus in der Zielsprache neu aufgebaut, zum Beispiel in Japanisch. Auch da gibt es Modelle für Bedeutung, Satzbau, Wortfunktion und Lautgestalt. Und was näseln nach Billionen von Rechenschritten, nach Zigtausenden von verworfenen Hypothesen aus Kayoko Miyamas Hörer: „*Joshi, kio wa jaru zo.*“ Verrückt doch, wie unser Hirn, die nichtkünstliche Intelligenz, dagegen locker antritt.

Die Computerindustrie tritt nicht locker an. ~~Ganz so wie oben beschrieben, arbeitet noch kein~~ System. IBM hat dieser Tage mit den Speech Server Series eine „hörende Schreibmaschine“ auf den Markt gebracht, die 20 000 Flexionsformen und Wörter erkennt – wenn man nach jedem Wort eine Pause einlegt. ATR, CMU und Siemens wollen Anfang 1993 den Prototyp eines Simultanübersetzers für Telephonate vorführen. Der Automat übersetzt fließend gesprochenes Deutsch, Englisch und Japanisch „von sehr eingeschränktem Wortschatz“. Mit der Anzahl verschiedener, erkennbarer Wortformen steigt die Zahl der Hypothesen und damit der Wahrscheinlichkeitsrechnungen gewaltig an, der Computer wird langsam unbrauchbar für die Simultanübersetzung. Weil man in der Spracherkennung und -übersetzung mit komplexen statistischen Modellen arbeitet, wird der Rechenaufwand prinzipiell hoch. Also muß die Hardware an Leistung zulegen. Schneller werden die Rechner mit ASICs, Application Specified Integrated Circuits, maßgeschneiderten Schaltkreisen also – und mit Parallelprozessoren. Während herkömmliche, serielle Rechner einen Befehl nach dem anderen abarbeiten, schmeißen Parallelrechner mehrere Jobs zugleich. Ihnen gehört die Zukunft; die Gegenwart gehört ihnen nicht. Bis ich mit Kayoko am Draht

in meiner Muttersprache durch Gedankenwelten bummeln kann, vergehen „noch Jahre“, sagt Siemens.

Da bleibt den Juristen doch etwas Zeit, über Datenschutzprobleme nachzudenken. Ein deutscher Hersteller digitaler Telephonzentralen lobt im Werbeprospekt, sein Produkt biete „schon vom Eingabemedium her alle Voraussetzungen für die optimale Arbeitszeiterfassung“, und die Einzelverbindungen könnten „u. a. mit Zeit, Datum, Verbindungsdauer und gewählter Rufnummer überprüft“ werden. Ein Sprachübersetzungssystem verrät mehr. Der Computer muß eine Dialoggeschichte anlegen, soll er verdeckte Bedeutungen analysieren, Pronomina zuordnen, Themenbezüge herstellen können. Praktisch heißt das: Da wird gespeichert, welche Wörter wie oft und in welchem Zusammenhang auftauchen. Und ich Dussel habe Bombe gesagt! Bombenstimmung! Wer immer gegen den Terror dieser Welt antritt und dafür aufs System zugreift, hat mich automatisch ertappt. Seit Linguisten mit Informatikern zusammen an Natural Language Processing arbeiten, befürchten einzelne Vertreter der Zunft, nach Physik, Chemie und Biologie verliere auch die Sprachwissenschaft „ihre Unschuld“. So klangen Voten etwa 1983 auf der Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft. 1986 warnte die Gesellschaft für Linguistische Datenverarbeitung in ihren Thesen zur NLP-Forschung gar vor der „militärischen Anwendung“.

Kürzlich schrieb ein Wissenschaftler eines großen deutschen Unternehmens für Elektronik und Kommunikation: „In wenigen Jahren schon werden uns viele Maschinen aufs Wort gehorchen.“ Hoffentlich kommt dann wenigstens in den Übersetzungsmaschinen die Höflichkeitsform automatisch richtig. Wenn nicht, wird sich Kayoko Miyama an den rüden Ton der Technik gewöhnen wie an den Lärm in den Straßen Kyotos. Oder sie bleibt ganz Japanerin. Dann bin ich aber gespannt, wie der Computer ihr Kichern übersetzt.