

Verschwendete Steuergelder

(Radebrecher, Der Verbomobil-Forschungsprototyp, c't 7/97)

Schön, man liest mal wieder vom Verbomobil. Immer wieder geistert dieses System durch die Medien, mit Zwischen- und Zwischenzwischenpräsentationen. Und immer wieder frage ich mich: Was machen die Leute eigentlich? Mein Eindruck ist, daß da Steuergelder verschwendet werden.

Fassen wir mal zusammen: 1992 stellt das DFKI ein Projekt bei Herrn Riesenhuber vor. Man wolle gesprochene Sprache nicht nur erkennen, sondern auch noch übersetzen, und nicht etwa deutschen Text ins Englische, sondern Japanisch sollte es werden. Das Ganze sollte in Echtzeit stattfinden, und weil es schöne Zukunftsmusik war, sollte das ganze System auch noch tragbar – mobil – sein. Man fühlt sich an Perry Rhodan erinnert, da gab es diese Maschine, die beliebigen Sprachinput aufnahm, und vorne kam dann Intergalaktisch raus.

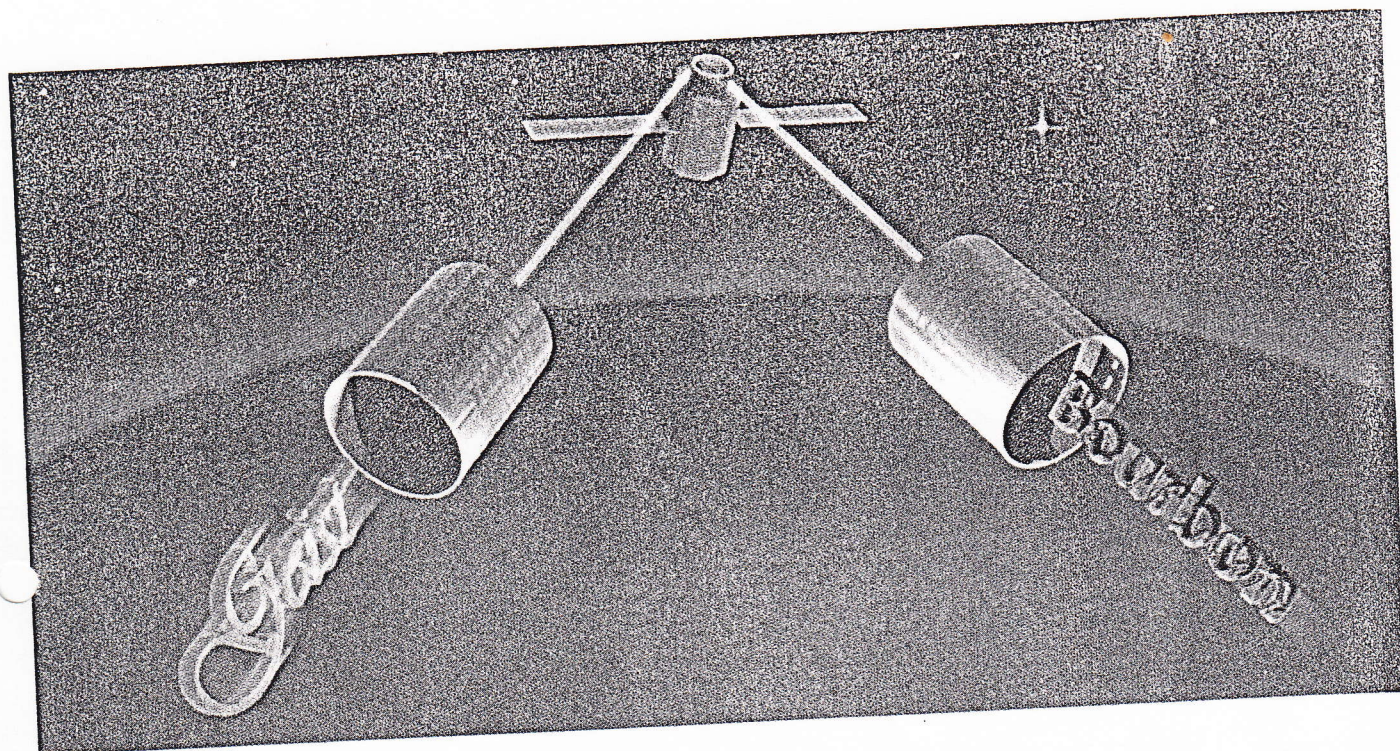
Heute sehe ich einen Bericht, in dem gezeigt wird, daß das System 2500 Wörter kennt (wie viele können eigentlich diese Langenscheidt-Wörterbücher, die wie Taschenrechner aussehen?) und sage und schreibe Laute grafisch darstellen kann

und einen Satz wie 'der Konferenzraum ...' in blanken Unsinn übersetzt. Die grafische Darstellung sieht aus wie bei einem Programm, das bei meinem Soundblaster in der Schachtel lag. Die Sprache erkennt mein Voice-Type, wenn ich auf die Wortgrenzen achte, und übersetzen, ja das kann vielleicht dieses Langenscheidt-Taschenrechnerwörterbuch besser.

Ok, vielleicht übersehe ich den konzeptionellen Kniff, der die Grundlagenforschung weiterbringt. Aber der Aufbau erscheint so langweilig wie das Mensa-Essen. Es werden Worte erkannt, geparkt, die Parsinginformation wird an eine Semantikkomponente weitergegeben und dann kommt Unsinn raus. Die einzig eventuell interessante Komponente, die nicht schon aus vorigen Projekten so bekannt wäre, finde ich in der Grafik unter der 'Prosodieanalyse'. Aber ist die wirklich 100 Millionen wert?

Alles in allem wundere ich mich, daß man dem Zukunftsminister mit diesen Ergebnissen offensichtlich immer noch Geld abringen kann. Übersetzer aber können aufatmen: Ihre Aufgabe bei Konferenzen wird noch lange nicht von einer Maschine gelöst.

Max Horneck, Freiburg



Hubert Kurth

Radebrecher

Der Verbmobil-Forschungsprototyp

An der maschinellen Sprachübersetzung hat sich bis heute noch jede Software die digitalen Zähne ausgebissen. Kein Grund zur Verzweiflung, sagt man sich kollegial beim Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) und beim Bundesministerium für Bildung, Wissenschaft und Forschung (BMBF) – und holt im gleichen Atemzug das millionenschwere Brecheisen aus der Tasche.

Bereits 1995 stellte Bundesforschungsminister Dr. Jürgen Rüttgers der Öffentlichkeit die damals noch Verbmobil-Demonstrator betitelte Software vor, deren Wortschatz einen Umfang von knapp 1300 Wörtern zu diesem Zeitpunkt besaß und ausschließlich auf das Teilgebiet der Terminverhandlung ausgerichtet war. Daran hat sich auch bei der Weiterentwicklung, dem Verbmobil-Forschungsprototyp, nichts geändert, dafür wurde der Wortschatz auf zirka 2500 Wörter erhöht. Zusätzlich zur deutschen Sprache versteht das System jetzt, wenngleich nur in sehr geringem Umfang, Japanisch. Das Grundkonzept, die Verständigung in einer quasi-

neutralen und weltweit verbreiteten Zwischensprache – auf dieser Welt Englisch – zu erreichen, wurde beibehalten.

Dies zeigt bereits, ebenso wie die Einschränkung auf ein engbegrenztes Anwendungsgebiet und einen relativ geringen Wortschatz, daß man nicht zu schnell nach den Sternen greifen will, denn durch Verwendung einer Zwischensprache wird die Anzahl der notwendigen Sprachpaare minimal gehalten. Dennoch stellt das Projekt, in das nach offiziellen Angaben in der ersten, von 1993 bis 1996 dauernden Phase insgesamt 96 Millionen DM geflossen sind, einen nicht zu vernachlässigenden Anspruch: Die Sprecherun-

abhängige Erkennung und Analyse von Spontansprache, inklusive Übersetzung und Sprachsynthese in das Englische.

Feine Gesellschaft

Die Aufzählung der Verbmobil-Partner läßt denn auch einiges erwarten, außer dem Saarbrücker DFKI tauchen 19 weitere deutsche und zwei nordamerikanische Universitäten und Forschungszentren auf – mit der Pittsburgher Carnegie Mellon University und dem Center for the Study of Language and Information in Stanford hat man namhafte Partner im Bereich der Künstlichen Intelligenz (KI) gewonnen. Wesentlich interessanter aber sind die Partner aus der Industrie. Neben Philips und IBM, die ja schon einiges Aufsehen mit ihren Sprachtechnologie-Produkten erregt haben, zählen Siemens, Daimler-Benz und einige andere bekannte Firmen zum erlauchten Kreis der Ausgewählten. Fast ein Drittel der gesamten Fördermittel stammt von diesem Industrie-Konsortium, eine Tatsache, die nahelegt, daß es um mehr als reine Grundlagenforschung geht.

Begrenzte Spontanität

Im wesentlichen besteht der Verbmobil-Prototyp aus mehreren, untereinander kommunizierenden Modulen, die teilweise

Rückmeldungen verarbeiten können und ihr Ergebnis an die nächste Stufe weitergeben. Auf die Festlegung einer einheitlichen Programmiersprache wurde verzichtet; die Module können in unterschiedlichsten Codes wie Lisp, Prolog, Fortran, C oder C++ erstellt werden, eine spezielle Gruppe am DFKI integriert sie dann zu einem Gesamtsystem.

Die Eingabe erfolgt durch explizite Aufnahme des Sprechers, das heißt, ein Klick mit der Maus löst die Aufnahme aus; wird die Maustaste losgelassen, gilt die Aufzeichnung als beendet, und das System legt los. War die Eingabe zu laut, so moniert dies umgehend eine Computerstimme und fordert zur erneuten Eingabe auf.

Nachdem die Bearbeitung der Aufnahme, welche auch für die Komprimierung des Sprachsignals sorgt, abgeschlossen ist, darf der Spracherkennung zeigen, was er kann. Im Idealfall – die Eingabe erfolgt durch einen bekannten Sprecher – liefern beide im Prototyp vorhandenen Alternativen, auf der einen Seite das Modul von Daimler-Benz, auf der anderen der Erkennung der Universität Karlsruhe, relativ gute Ergebnisse mit Wortfehlerraten zwischen 10 und 30 Prozent – vorausgesetzt, der Sprecher hält sich an den Wortschatz und artikuliert sehr deutlich. Von Sprecherunabhängig-

keit kann allerdings bis jetzt noch keine Rede sein, wir jedenfalls erzielten keine allzu großen Erfolge bei unseren Tests auf der CeBIT. Die vom Spracherkennungsmodul als möglich erkannten Wörter werden intern mit Wahrscheinlichkeiten bewertet und in einem Worthypothesengraphen repräsentiert.

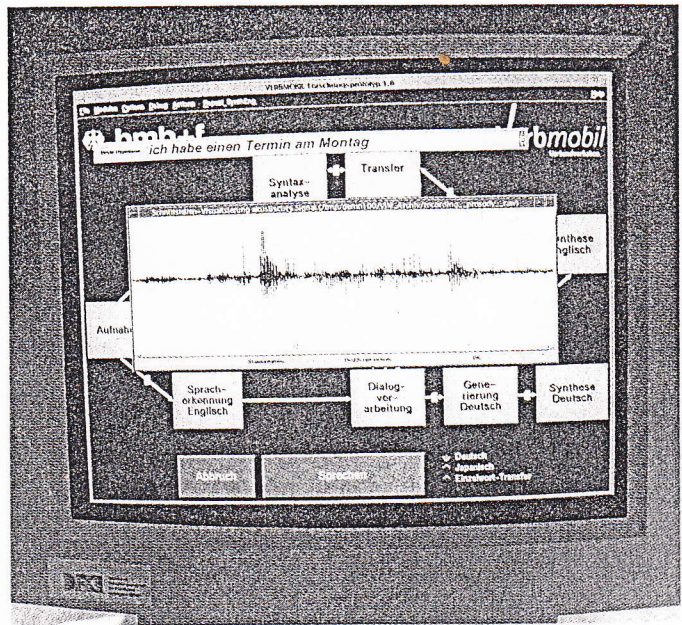
Melodische Syntax

Bevor das Erkennungsmodul sein Ergebnis an die Syntaxanalyse weitergibt, wird das Resultat noch durch die von einem Prosodie-Modul (Silbenmessung) gefundenen Informationen zu Satz- und Phrasengrenzen bereichert. Der Clou ist hierbei die Verwertung von Dauer und Stärke des Signals sowie von Intonation und Pausen, um die erforderliche Information abzuleiten. Nur so läßt sich auf sprachlicher Ebene entscheiden, ob die Eingabe beispielsweise 'Ja, vielleicht geht es auch am Dienstag,' oder 'Ja, vielleicht. Geht es auch am Dienstag?' lautete. Die Syntaxanalyse überprüft ihren Input auf grammatikalische Korrektheit und liefert ihrerseits das Ergebnis bei der Semantik ab: dieses kann immer noch Mehrdeutigkeiten enthalten. Ob mit dem Wörtchen 'sie' die Anrede einer einzelnen Person oder einer Gruppe bezweckt ist oder um welche 'Stock' (Stockwerk oder Axt?) es sich handelt, muß letztendlich die Semantik ermitteln. Unterstützt wird die linguistische Analyse sowie das Transfermodul zusätzlich von einem Dialogmodul, das sogenannte Dialogakte durch

statistische Methoden zu erkennen versucht. Beispiele für Dialogakte, die im wesentlichen Kernaussagen darstellen, sind *suggest_date*, *reject_date* und *accept_date*. Eben diese Dialogakte werden bei der 'flachen' Verarbeitung ausgenutzt, um aus für die herkömmliche Tiefenanalyse unverständlichen Sätzen trotzdem relevante Informationen zu gewinnen.

Ein Satz wie 'Hmm, ja, dann würde ich mal äh sagen äh vorschlagen, daß wir uns äh also dann äh am dritten Juni treffen,' wird von der Tiefenanalyse nicht zufriedenstellend bearbeitet, die flache Verarbeitung hingegen erkennt den Dialogakt *suggest_date*, filtert die zugehörige Datumsangabe 'dritter Juni' heraus und macht stur nach Schema F einen Übersetzungsversuch. Dieser Ansatz mag zwar auf den ersten Blick wie das Ei des Kolumbus aussehen, bleibt aber erfolglos, sobald ein entscheidender Dialogakt nicht oder falsch erkannt wurde, und täuscht zudem über Schwächen bei der Tiefenanalyse hinweg.

Das Verbmobil-System versucht deshalb auch, beide Ansätze miteinander zu verknüpfen, indem die gefundenen Ergebnisse intern bewertet und gegeneinander abgewogen werden – die dazu herangezogenen Kriterien darf man allerdings nicht überschätzen, denn, falls das System zuverlässig entscheiden könnte, welcher Vorschlag der angemessener ist, würde es ja gerade bessere Ergebnisse liefern. Das Resultat der flachen Verarbeitung wird übrigens auch verwendet, wenn die Tiefenanalyse nach 35 Sekunden kein brauchbares Ergebnis gefunden hat.



Akustische Signale lassen sich auch visualisieren.

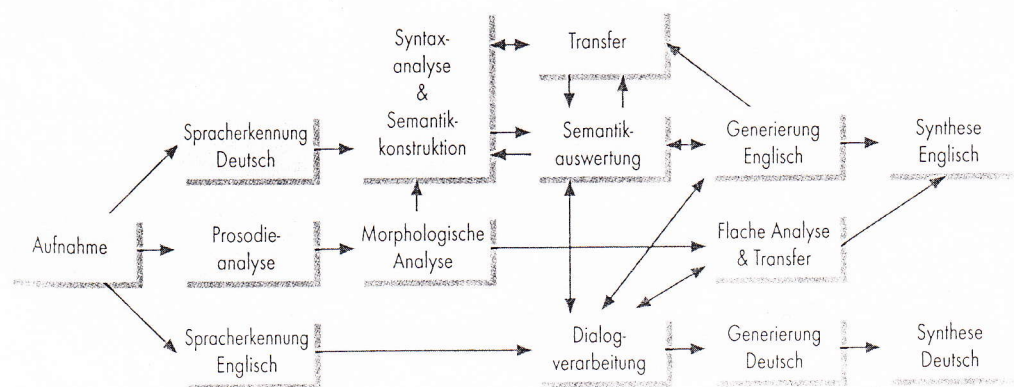
Begriffsstutzig

Abschließend erzeugen Transfer- und Generierungsmodule die Ausgabeinformation für das Synthesemodul. Hier müssen Verschleifungen (zum Beispiel: 'geht's') realisiert werden, damit die generierte Sprache nicht zu roboterhaft klingt, und ebenso muß das Geschlecht des Sprechers umgesetzt werden. Daß das nicht immer klappt, zeigte ein Mitarbeiter des Projektes unabsichtlich auf einer Präsentation im vergangenen Herbst, quasi als Transvestit wider Willen. In der 1997 angelaufenen zweiten Phase, in der laut Spiegel [1] wieder ein ähnlich hoher Etat bereitgestellt werden wird, will man unter anderem Wortschatz, Domäne (Anwendungsgebiet) und Sprachpaare erweitern und das System auf einem Sprachser-

ver eventuell der Öffentlichkeit zur Verfügung stellen. Obwohl Verbmobil inkorrekte Datumsangaben wie den '31. Juni' erkennen soll, mutet das laut Gesamtprojektleiter Prof. Dr. Wolfgang Wahlster vom DFKI implementierte Weltwissen verschwindend gering an: Finden sich in der Eingabe Anredeformen wie 'Herr Walter' oder 'Frau Schmidt', dann hat der Sprecher Pech gehabt – Verbmobil kennt nur drei Namen. Man scheint schlichtweg übersehen zu haben, daß das auf die Anrede folgende Wort geradezu als Name prädestiniert ist. Unseren Testsatz 'Der Konferenzraum ist im fünften Stock' verhunzte der Zauberlehrling zu 'The conference of woman this five o'clock'. Trotz der offensichtlich für die Domäne Terminverhandlung als unwichtig erachteten und deshalb unbekannteren Vokabel 'Konferenzraum' und der eklatanten grammatikalischen Fehler sollte zumindest eines klar sein: Eine Konferenz mit nur einem Teilnehmer ist ein Widerspruch an sich. Vielleicht hat man sich aber lediglich einen der ungeschriebenen Grundsätze der KI zu sehr zu Herzen genommen: Die Maschine ist dann intelligent, wenn sie Humor zeigt. (ae)

Literatur

- [1] Mit Weltwissen gefüttert. Der Spiegel, Ausgabe 5/97, S. 164 **ct**



Ein Vorteil des Komponentenaufbaus: funktional gleichwertige Module, jedoch mit unterschiedlicher Konzeption können leicht gegeneinander ausgetauscht werden.