

## *COMPUTER UND SPRACHE*

Vortrag zum Rahmenthema *Schule im Zeitalter der elektronischen Fachinformation* aus Anlass des 20jährigen Bestehens der Otto-Hahn-Schule in Hanau (4. Otto-Hahn-Forum)

### 1 Sprache und Wissenstransfer

Die Fähigkeit, eine natürliche (Wort-)Sprache in ihrer Virtuosität und Anwendungsvielfalt zu erwerben (erlernen) und zur Verständigung (Kommunikation) zu benutzen, ist ein wichtiges Merkmal des Menschen. Die Sprache wird dabei in vielerlei Hinsicht verwendet. Nicht alles, was gesprochen wird, ist beispielsweise "wortwörtlich" gemeint. Andererseits muss dem sprachlichen Verstehen ein gemeinsamer "Code" (beim Sprecher und beim Hörer) zugrunde liegen, der sprachliche Äußerungen über Regeln mit Bedeutungen verbindet. Wir wissen auch, dass es das "ideale" Verstehen wohl kaum gibt, zumal der Spracherwerb eng verknüpft ist mit dem persönlichen Erfahren der Welt. Daher ist jedes Verstehen natürlichsprachiger Äußerungen auch ein gutes Stück Interpretation, und dies um so mehr, je geringer die gemeinsame Wissensbasis (bzw. Welterfahrung) ist.

Wir wollen im folgenden eine Reihe von Fragestellungen ausklammern, die mit dem verschiedenartigen Gebrauch und der Funktion einer natürlichen Sprache verbunden sind. Hier kann - stellvertretend - auf das lesenswerte Buch von Wolf Schneider ("Wörter machen Leute", München 1976) verwiesen werden. Es darf aber nicht übersehen werden, dass viele der dort geschilderten Faktoren wie Sprachmagie, Sprachartistik, Sprachstilistik bei den folgenden Darstellungen mitschwingen.

Im Mittelpunkt der folgenden Überlegungen steht - gemäß dem vorgegebenen Rahmen - die informationelle Ebene der Sprache, d.h. die Sprache als Instrument des Wissenstransfers. Würde man hier den Argumenten und Schlussfolgerungen von Wolf Schneider folgen, so erschiene dies

als reine Sisyphusarbeit: Er kommt nämlich zu dem (hoffentlich nicht ganz zutreffenden) Schluss, dass diese Funktion einerseits gegenüber anderen Sprachfunktionen (wie Gebet, Zusage, Selbstdarstellung, Manipulation, Utopie und Geschwätz) nur einen geringen Teil des täglichen Sprachaufkommens ausmacht, andererseits auch niemand "wirklich" informieren oder auch informiert werden möchte (vgl. S. 247 ff.).

Nun, wir können ihm hierbei - zumindest zur Ehrenrettung unserer eigenen Wissenschaft - nicht so uneingeschränkt folgen. Doch ist zuzugestehen, dass der Wissenstransfer mit Hilfe der natürlichen Sprache nicht unproblematisch ist. Jede Wissenschaft, die ihre Erkenntnisse möglichst eindeutig (d.h. allgemeinverbindlich und unmissverständlich) darstellen und vermitteln möchte, ist hier nämlich in einem gewissen Dilemma: einerseits wird die natürliche Sprache (zumindest die Muttersprache) von fast jedermann (zumindest jedem Wissenschaftler) ausreichend beherrscht, d.h. das über sie vermittelte Wissen erreicht prinzipiell viele Menschen einer Sprachgemeinschaft, andererseits ist die große Flexibilität und auch die damit verbundene Vagheit, die eine natürliche Sprache braucht, um in einer größeren Gemeinschaft als Vermittlungsinstrument dienen zu können, auch eine Barriere, da auf dem Vermittlungsweg Missverständnisse oder zumindest Unklarheiten auftreten können.

Dies ist (neben anderen, z.B. sprachökonomischen Gesichtspunkten) ein Grund dafür, dass sich *formale* Wissenschafts(teil)sprachen ausgebildet haben, z.B. in der Mathematik und der Chemie. Ein weiterer Grund ist, dass solche Fachkunstsprachen sehr leicht *internationalisiert* werden können, während die natürlichen Sprachen durch ihre Vielfalt noch zusätzlich das Problem (und die Barriere) der Sprachübersetzung mit sich bringen. Dennoch bleibt als Faktum bestehen: die natürliche Sprache war, ist und bleibt auf absehbare Zeit ein wichtiges Instrument des Wissenstransfers und des zwischenmenschlichen Informationsaustauschs.

Angesichts der Probleme, die in ihrer Vielfalt hier nur angedeutet werden konnten, gilt es aber, nach Wegen zu suchen, Barrieren, die dem Gebrauch der natürlichen Sprache(n) in der (weltweiten) Information und Kommunikation - die nicht nur auf Fachinformation beschränkt ist - abzubauen oder zu reduzieren.

## 2 Computer und Information

Nicht erst in den letzten Jahren hat man den Computer als Instrument der Wissensspeicherung, -verarbeitung, -vermittlung und Wiedergewinnung entdeckt. Aber lange Zeit *war er* ein recht kostspieliges, schwer zugängliches und auch nicht für alle Problemlösungen (v.a. wegen der mangelnden Geschwindigkeit) geeignetes Werkzeug. Ansätze zur maschinellen Verarbeitung natürlicher Sprachen bis hin zur maschinellen Übersetzung gibt es bereits seit den 50er Jahren; in den 60er und 70er Jahren gab es schon fast eine Euphorie im Hinblick auf die Anwendung des Computers im Unterricht. Dennoch hatte vieles von dem, was damals konzipiert und labormäßig realisiert wurde, bereits aus ökonomischen Gründen wenig Aussicht auf Erfolg.

Erfolgreicher erschien dagegen - und hier kann die Bundesrepublik Deutschland auch zu den Pionieren gerechnet werden - der Aufbau großer elektronischer Datenbanken, die über weltweite "Netze" elektronisch zugänglich gemacht wurden. Während für die Sprachübersetzung und den computergestützten Unterricht überhaupt erst noch formale *bzw.* formalisierbare Konzepte zu erarbeiten sind, lagen in der traditionellen fachlichen Information und Dokumentation bereits große Datensammlungen "auf Papier" vor, die nur noch maschinell zu erfassen, zu speichern und über Datenbanksysteme (über einfache, relativ leicht erlernbare Dialoge) zugänglich zu machen waren. Etwas vereinfacht ausgedrückt: der Karteikasten und das Bücherregal werden hierbei durch den Computerspeicher ersetzt mit dem Vorteil, dass damit ungeheure Datenmengen konzentriert gespeichert und gleichsam per Knopfdruck in kürzester Zeit am Arbeitsplatz des Technikers, des Wissenschaftlers, des Juristen *usf.* verfügbar werden.

Neben bibliographischen Datenbanken, die v.a. die Titel, Quellenangaben und ggf. auch Kurzfassungen zur Fachliteratur verzeichnen, entstanden zunehmend (Voll-)Textdatenbanken, die z.B. Gesetzestexte und Urteile, aber auch Wirtschaftsnachrichten speichern, daneben in wachsender Zahl auch numerische Datensammlungen oder Faktenbanken, die Statistiken, Messwerte, Formeln *usf.* umfassen. Allein in einer einzigen solchen Datenbank, die inzwischen weltweit zugänglich ist, dem Chemical Abstract Service (CAS), werden pro Jahr rund 500.000 Neuerscheinungen von Fachaufsätzen erfasst und strukturell aufbereitet, um sie z.B. mittels chemischer Formeln *bzw.* Teilformeln abfragen zu können.

Ein Ende dieser Entwicklungen ist noch nicht abzusehen. Hinzu kommt, dass sich die allgemeinen technischen Rahmenbedingungen extrem verbessert haben. Mit dem Personal Computer (PC), der mit relativ geringem Mehraufwand als Endgerät zum Zugang zu Datenbanken verwendet werden kann, ist eine flexible, ausbaufähige technische Schnittstelle entstanden, die praktisch von jedermann benutzt werden kann; mit dem Ausbau der Telekommunikationsnetze (wo mit dem integrierten Datendienstnetz ISDN gerade ein neuer Technologieschub bevorsteht) sind weitere technische Voraussetzungen gegeben, viele Interessenten in der Wirtschaft und Industrie an diese Wissensspeicher heranzuführen.

Auch das in der Bundesrepublik Deutschland z.Zt. wegen einiger technischer Mängel noch etwas am Rande stehende Bildschirmtext-System (BTX) wird - in veränderter Form - in naher Zukunft seine Möglichkeiten erst richtig entfalten: Die französische "Schwester" von BTX, das MINITEL, hat in Frankreich beispielsweise bereits mehrere Millionen Anschlüsse, die den Weg zur elektronischen Information und Kommunikation für den privaten wie den betrieblichen Bereich gleichermaßen bahnen.

Die verfügbaren Endgeräte: der PC oder auch die "professionelle Workstation" sind demgegenüber ebenfalls leistungsfähiger geworden. In wachsendem Maße werden Texte, Graphiken und numerische Daten bereits beim ersten Entstehen (der Generierung) maschinell verfügbar, da komfortable Textsysteme bis hin zum computergestützten Zeichnen schon fast zum betrieblichen Alltag gehören (auch wenn Schulen und Universitäten hier noch einen gewissen Nachholbedarf haben). Mit Festplattenspeichern wachsender Kapazität, aber auch der datenorientierten "Compact Disk" (der sog. CD-ROM) und der graphikorientierten Bildplatte sind "vor Ort" (d.h. am einzelnen Arbeitsplatz) Speicherkapazitäten gegeben, die diese Endgeräte in Verbindung mit den steigenden Geschwindigkeiten zu leistungsfähigen Instrumenten machen, die komplexe Verarbeitungen erstmals ökonomisch zulassen, eine Voraussetzung für alle Entwicklungen in einer freien Marktwirtschaft.

Diese Entwicklungen gehen einher mit weltweiten Anstrengungen, die Computer *intelligenter* zu machen. Ich gebrauche das Wort *Intelligenz* ohne Anführungsstriche, in dem Bewusstsein, dass damit natürlich nicht ein direkter Vergleich mit der menschlichen Intelligenz beabsichtigt ist. Es soll vielmehr zum Ausdruck gebracht werden, dass Computer - übrigens ist dies ein schon länger

andauernder Vorgang - in der Lage sind bzw. in die Lage versetzt werden, immer komplexere Problemlösungen zu bewältigen, die bislang der menschlichen Problemlösungsfähigkeit zugeschrieben wurden. Ein Zeichen dieser Entwicklungen sind die sog. *Expertensysteme*. Wesentliches Merkmal eines Expertensystems ist die Fähigkeit, über Regeln zur Problemlösung auf Wissensspeicher zuzugreifen, wobei weiteres Wissen aus dem vorhandenen Datenmaterial abgeleitet wird. Auf eine sich entwickelnde Fähigkeit dieser Systeme, den Dialog in natürlicher Sprache, wird in einem besonderen Zusammenhang nochmals eingegangen.

### 3 Sprachverarbeitung mit dem Computer

Die Sprachwissenschaft, besonders die sog. Computerlinguistik, beschäftigt sich seit mehreren Jahrzehnten bereits mit der Sprachdatenverarbeitung. Dies hat zunächst auch eine erkenntnistheoretische Seite: Um ein Sprachsystem zu erforschen, kann man große Mengen gesprochener oder geschriebener Sprachäußerungen sammeln und systematisch untersuchen. Hier tut der Computer als Speicher- und Sortierinstrument gute Dienste:

- Er ersetzt beispielsweise die Karteikarte des Wissenschaftlers. Im weitesten Sinne gehört hier die *Sprachstatistik* hinzu. Ein Beispiel sind die Untersuchungen des Instituts für Deutsche Sprache in Mannheim zur Entwicklung des Wortschatzes und der Wortbedeutungen im Vergleich DDR / BRD anhand von Zeitungstexten mit mehreren Millionen laufender Wörter.
- Man kann daneben auch *Modelle über ein Sprachsystem* (genauer gesagt: zu Teilsystemen, etwa der Satzstruktur, der Satz- und Wortbedeutung) entwickeln und mit Hilfe des Computers auf Plausibilität und Adäquatheit testen. Dies ist schon ein weit komplexeres Unterfangen, da hierzu etwa die Entwicklung und Nutzung eines computerverständlichen Formalismus, d.h. Kenntnisse in formalen Sprachen, aber auch sprachwissenschaftliches Know-How gehören. Wegen der Komplexität der natürlichen Sprache (bis hin zu Vagheitsproblemen) können solche Modelle bislang kaum befriedigen, soweit die umfassende Abbildung einer natürlichen Sprache betroffen ist.

Dennoch haben Forschungen mit Computersimulationen ihren Sinn und die Kenntnisse über die natürliche(n) Sprachen erweitern helfen. Zumindest kann hierbei der Computer als ein Instrument gelten, das Schwachstellen in formalen Theorien unerbittlich aufdeckt.

Die genannten Anwendungsfelder sind im weitesten Sinne der Grundlagenforschung zuzurechnen, auch wenn sich die Forschungsziele - etwa zu Fragen der maschinellen Übersetzung - eigentlich immer schon mit Problemen begründeten, die aus der Anwendung kommen. Dennoch waren Fragen der Anwendung von Verfahren der maschinellen Verarbeitung natürlicher Sprachen für die menschliche Information und Kommunikation ein wichtiger Aspekt, auch die Grundlagenforschung zu fördern. Im folgenden sollen - ohne Vollständigkeit anzustreben - einige wichtige Themenstellungen zunächst im Überblick genannt werden, ehe auf einige daraus etwas näher eingegangen wird.

- Auf den wachsenden Markt der (Text-)Datenbanken wurde bereits hingewiesen. Die größte Barriere ist hier die Vielsprachigkeit. Ziel muss es also sein, die vorhandenen Texte in möglichst vielen natürlichen Sprachen verfügbar zu machen. Die Alternative, Englisch (als die z.Z. wichtigste internationale Wissenschafts-Verkehrssprache) zu verwenden, lässt sich auf lange Sicht - wenn überhaupt - nicht verwirklichen, obgleich in einigen Fachgebieten, z.B. in der Physik, ein deutscher Wissenschaftler kaum mehr in Deutsch publiziert. Da es - aus personellen wie ökonomischen Gründen - kaum möglich ist, alle erscheinende Fachliteratur intellektuell zu übersetzen, ist die *maschinelle Übersetzung* ein großes Desiderat.
- Ein Großteil zwischenmenschlicher sprachlicher Kommunikation erfolgt nicht über die geschriebene, sondern über die gesprochene Sprache. Dieser Bereich kann aus der Sicht der weiteren Verarbeitung überhaupt nur sinnvoll bearbeitet werden, wenn es gelingt, gesprochene Sprache zu erschließen. Dieses Problem wird unter dem Begriff der *Spracherkennung* zusammengefasst. Das Spektrum der Anwendung reicht dabei von der Identifikation einer begrenzten Zahl einzelner Wörter bei beliebigen Sprechern (etwa zur Steuerung eines Gerätes oder bei einem Auskunftssystem) über die Sprecheridentifikation (ähnlich den Fingerabdrücken) bis zu der Maschine, die fließend gesprochene Sprache in Text umsetzen kann.

- Ein weiterer, wenn auch weniger spektakulärer Anwendungsbereich ist die Textverarbeitung im privaten wie betrieblichen Bereich. Themen sind hier die automatische Silbentrennung, die Rechtschreibfehlerermittlung und -korrektur, Stil- und Grammatikhilfen. Ein Problem, das in einigen Sprachen (wie dem Deutschen) bereits komplexe sprachanalytische Fähigkeiten voraussetzt, ist beispielsweise die automatische Erkennung und Korrektur von Fehlern bei der Zeichensetzung.

Zu diesem Bereich der betrieblichen Information (wie auch der Fachinformation) sind auch Verfahren der Erschließung für die Ablage und das Wiederfinden von Texten zu rechnen, bei denen die maschinelle Indexierung zur Ermittlung von Schlag- oder Stichwörtern genutzt wird.

- Der Dialog mit dem Computer kann inzwischen auf verschiedene Arten geführt werden. Da immer mehr Menschen - auch solche, die in formal-technischer Hinsicht ungeübt sind - damit konfrontiert werden, ist nach Wegen zu suchen, die diesen technischen Zugang (die sog. Mensch-Maschine-Schnittstelle) verbessern. Die Suchbaumtechnik, eine Art Auswahlmenüverfahren (wie sie z.B. bei Bildschirmtext verwendet wird), ist zwar eine relativ einfache Möglichkeit, setzt aber umfangreiche Vorarbeiten im System voraus und kann Such- und Problemlösungsprozesse verzögern. Ein Zugang zu Datenbanken mittels eines natürlichsprachigen Dialogs ist also ein weiteres Desiderat.

#### 4 Zum Stand der Forschung und Entwicklung

Die Forschungen und Entwicklungen der letzten 20 Jahre vor allem in der Computerlinguistik haben vor allem zwei wichtige Erkenntnisse gebracht: Mit der Schulgrammatik, aber auch mit formallinguistischen Modellen alleine kann man das Problem des Erkennens (genauer: Verstehens) der natürlichen Sprache durch einen Computer bei weitem nicht bewältigen. So wichtig die rein sprachlinguistische Beschreibung (z.B. der Morphologie, also der Wortgestalt und Wortbildung, oder der Syntax, also der Satzstruktur) auch ist: ohne eine Kombination aus Sprachregeln

und Weltwissen sind die "großen" Themen - maschinelle Übersetzung und Verarbeitung gesprochener Sprache - im systematischen Sinne nicht angemessen zu lösen.

Aus dieser Erkenntnis heraus sind weltweit - auch in der Bundesrepublik Deutschland und in Europa - Forschungsprojekte entstanden, die Methoden und Verfahren der Linguistik, Psychologie, Informatik und Logik kombinieren. Aus diesem Themenkreis soll stellvertretend nur ein Schlagwort eingebracht und kurz erläutert werden: das der *Partnermodellierung* beim Mensch-Maschine-Dialog. Unter Partnermodellierung (vgl., dazu u.a. K. Morik, C.-L. Rollinger: Partnermodellierung im Evidenzraum. In: GWAI-83, ed. B. Neumann, Berlin/Heidelberg 1983, ISBN 3-540-12871-9, S. 158-168) versteht man ein System, das über explizites Wissen über seine Benutzer und evtl. Benutzertypen verfügt und dieses Wissen zur Steuerung seines Verhaltens benutzt. Dieses Wissen kann z.B. auch im Dialog mit dem Benutzer ergänzt bzw. verändert werden. Auf diesem Wissen lassen sich z.B. Beratungsdialoge (Hotelbuchung, Wohnungssuche, Geldanlage ...) steuern. Einen Anfragenden, von dem man z.B. weiß (d.h., im Dialog erfragt hat), dass er schulpflichtige Kinder hat, wird anders beraten als ein pensionierter Wohnungssuchender. Entsprechendes gilt für die Sprachkriterien (z.B. zur Auswahl des "Jargons", aber auch einfacherer Elemente wie Ellipsierung, Anaphorik).

Es muss an dieser Stelle eindringlich davor gewarnt werden, diese im Modell und für Modelluntersuchungen entwickelten Verfahren unüberlegt in die Praxis zu übernehmen. Schon Joseph Weizenbaum hat sich in seinem ebenfalls sehr lesenswerten Buch "Die Macht der Computer und die Ohnmacht der Vernunft" (dt. 1978, Frankfurt; orig. engl. 1976) ausführlich mit dem Thema Computer und Sprache befasst und dabei auch - aus eigener Erfahrung - vor unüberprüften, d.h. in ihren Folgen nicht übersehbaren Umsetzungen gewarnt. Selbst wenn man nämlich nicht der Ansicht Weizenbaums folgt, dass die (heutigen) Computer strukturell nicht in der Lage sind, menschliches Problemlösungsverhalten (auch beim Sprachverstehen) angemessen abzubilden: Die Komplexität des Verstehensvorgangs ist so groß, dass Abbildungen dieses Prozesses 'heutzutage nur Modellcharakter haben können, in sensiblen Bereichen aber, wie etwa bei Beratungssituationen und in Diagnosesystemen (ganz zu schweigen von psychotherapeutischen Gesprächen), nicht praxisrelevant erscheinen.



Aus der Sicht der Anwendung stehen Verfahren der Sprachtechnik, wie ich diesen Bereich in Anlehnung an Begriffe wie Medizintechnik oder Informationstechnik bezeichnen möchte, noch am Anfang einer langwierigen Entwicklung. Trotz einiger Einschränkungen, die eher prinzipieller oder auch verfahrenstechnischer Natur sind, kann jedoch aufgezeigt werden - und damit wird sich v.a. der nächste Abschnitt befassen - dass bestimmte praxisrelevante Themen wie z.B. die maschinelle Sprachübersetzung, aber auch Funktionseinheiten wie automatische Silbentrennung und Rechtschreibkontrollen, Computerwörterbücher und deren Anwendung beim Vokabeltraining, einen Stand erreicht haben, der zumindest den grundlegenden Anforderungen der Praxis standhalten.

Die Kriterien der Praxis unterscheiden sich - wie in fast allen Bereichen - grundlegend von den theoretischen Anforderungen, auch wenn man (angesichts mancher Fehlentwicklungen v.a. in der maschinellen Übersetzung) ein Mindestmaß an Integration theoretischen Grundlagenwissens voraussetzen muss.

Die Fragestellungen der Praxis lassen sich demgegenüber im wesentlichen auf folgende Punkte reduzieren:

- Ist der Einsatz eines maschinellen Systems - über alles gerechnet - kostengünstiger als ein entsprechender intellektueller Arbeitsschritt?
- Bringt der Einsatz eines maschinellen Systems gegenüber der intellektuellen Verarbeitung zeitliche Vorteile?
- Führt der Einsatz eines maschinellen Systems zu Lösungen, die bislang nicht oder mit vertretbarem Aufwand nicht realisiert werden konnten?

Diese Kriterien können sich im günstigen Falle ergänzen, aber jedes für sich genommen kann im Einzelfall schon ausreichen. Am Beispiel der maschinellen Übersetzung soll dies einmal verdeutlicht werden:

Es gibt bis heute kein am Markt verfügbares maschinelles Übersetzungssystem, das eine mögliche hochqualitative Übersetzung erreicht oder ihr sehr nahekommt. Dennoch sind nicht überall zu jedem Thema menschliche Übersetzer (sog. Humanübersetzer) verfügbar; auch bei der intellektuellen Übersetzungsqualität muss man häufig genug Abstriche machen, schließlich kann die maschinelle Übersetzung als Rohübersetzung rein informationell betrachtet (gleichsam als Wegwerfübersetzung) ausreichend sein, und letztlich entwickelt sich die maschinelle Übersetzung in Verbindung mit einer intellektuellen Nachkorrektur zu einer kostengünstigen Alternative (vgl., die Beispiele Bild 9ff. im Anhang).

Entsprechende Beispiele lassen sich im Bereich der Textverarbeitung, der Dokumentablage und der Nutzung mehrsprachiger elektronischer Wörterbücher aufzeigen. Dies soll in der sich anschließenden praktischen Vorführung weiter veranschaulicht werden, so dass hier nur einige punktuelle Rahmensetzungen erfolgen können:

- Die *vollautomatische Silbentrennung* ist heute eine Standardfunktion guter Textverarbeitungssysteme. Gerade die deutsche Silbentrennung stellt jedoch bereits relativ hohe Anforderungen an ein elektronisches System, die in Einzelfällen (etwa bei Mehrdeutigkeiten wie Wach-stube / Wachs-tube; Erd-rücken / Er-drücken, Druckerzeugnis / Druckerzeugnis) eine Sinn-Erfassung des Satzes oder Textes benötigen. Sieht man jedoch von solchen Fällen ab - sie machen statistisch betrachtet nur einen minimalen Prozentsatz aus - so tun heute automatische Trennverfahren sehr gute Dienste - zur Entlastung des Sachbearbeiters wie der Schreibkraft (vgl. Bild 1: Beispiele für die Behandlung von Ausnahmefällen bei einer algorithmischen Lösung).

<u>Wortformen:</u>	
wor-an	(# = unästhetische Trennung
wor-un#ter	% = c-k -- > k-k
black-out	& = Konsonantenverdopplung)
<u>Wortanfänge:</u>	
zi-vi-li-	(für „zi-vi-li-siert“ ...)
mon-go-lo-	(für „mon-go-lo-id“ ...)
we#sent-	(für „we#sent-lich“ ...)
schif&fahrts-	(für „Schiff-fahrts-amt“ ...)
<u>Wort und Wortzusammensetzung:</u>	
wis-sens(-)	(für Wis-sens-er-werb ...)
wel-len(-)	(für wel-len-ar-tig ...)
buc%kel(-)	(für Buk-kel-an-satz ...)

*Bild 1. Beispiele aus der Ausnahmeliste zur deutschen Silbentrennung (System PAST)*

- Deutlich höhere Ansprüche stellen Problemlösungen zur *Rechtschreibhilfe* an ein System. Es genügt in der Regel kaum, einzelne Wörter korrekt zu identifizieren, obwohl die Erkennung von Tippfehlern schon eine große praktische Hilfe sein kann. Fehlerhafte Wörter sollten nach Möglichkeit gleich korrigiert werden können (vgl. Bild 2: Beispiele für wissensorientierte Fehlerkorrekturen). Sprachen wie das Deutsche verlangen nach morphologischen Verfahren, die auch Wörter als korrekt bestätigen, wenn sie nach den Regeln der Wortbildung gestaltet sind (vgl. Bild 3: Kodierungen zur Erkennung von Wortbeugung, -ableitung und Zusammensetzung; Bild 4: Beispiel für eine Wortidentifikation).
- Texterfassung kostet Zeit (und Geld); auch wenn es irgendwann einmal möglich sein sollte, gesprochene Sprache in Schrifttext umzusetzen: Es ist kaum vorstellbar, dass dies für beliebige Sprecher und beliebige Texte gilt. Mit der Erfindung der Schreibmaschine ist ein relativ brauchbares Instrument der Dateneingabe geschaffen worden (und Kenntnisse im Maschinenschreiben sind - nebenbei bemerkt - eigentlich von jedem zu fordern,

der mit dem Computer umgehen möchte). Dennoch kann man sich vorstellen, dass man mit modernen Verfahren den Erfassungsvorgang noch beschleunigt (mit dem Ziel, eine *Maschinenstenographie* zu schaffen, mit deren Hilfe man in der Lage ist, in Sprechgeschwindigkeit Texte zu erfassen und vollautomatisch in Normalschrift umzusetzen, Ansätze verschiedener Art (unter Verwendung einer speziellen Akkordanschlags-Tastatur, aber auch durch Wortkürzel) sind dazu in Entwicklung (vgl. Bild 5: Beispiel für eine Kürzelschrift).

- *Grammatik- und Stilhilfen*, aber auch Verfahren zur *automatischen Texterschließung und Retrieval* setzen v.a. entsprechende elektronische Wörterbücher voraus. Dies gilt für Begriffsrelationierungen (vgl. Bild 6) wie auch für die Ermittlung sinntragender Wörter in einem Text (vgl. Bild 7). Es ist heute kein großes Problem mehr, derartige Wörterbücher auf einem Personal Computer bereitzustellen und entsprechende Verfahren in die Texterschließung und -suche zu integrieren. Für die Praxis sind die (weitgehend noch ungelösten) Fragen der Wortmehrdeutigkeit meist weniger wichtig, da in den relevanten Fachgebieten nur selten daraus Probleme resultieren.
- Probleme der *Wort- oder Textübersetzung* können heute schon durch PC-Lösungen sinnvoll behandelt werden. Dies gilt einmal für die (wortorientierte) automatische Blätterfunktion in einem elektronischen Übersetzungswörterbuch, um den professionellen Übersetzer, aber zunehmend auch den Endnutzer zu unterstützen. Dies kann aus dem Text heraus geschehen (vgl. Bild 8: Einfache - wortorientierte - Übersetzungshilfe), aber auch durch elektronisches Nachschlagen, ggf. mit Übernahme gefundener Äquivalente in einen zu übersetzenden Text.

Abonement -- > Abonnement	Rythmus -- > Rhythmus
Adresse -- > Adresse	Weisbrot -- > Weißbrot
arhythmisch -- > arrhythmisch	achsial -- > axial
Balett -- > Ballett	Apartheit -- > Apartheid
Dahrlehen -- > Darlehen	Balaleika -- > Balalaika
Diphthong -- > Diphthong	chifriert -- > chiffriert
emfindlich -- > empfindlich	Dedektiv -- > Detektiv
Hypothenuse -- > Hypotenuse	Embrio -- > Embryo
Interesse -- > Interesse	Entgeld -- > Entgelt
Paralele -- > Parallele	Idiologe -- > Ideologe
Rechtanwalt -- > Rechtsanwalt	Lybien -- > Libyen
richtungsweisend -- > richtungweisend	Prophezeihung -- > Prophezeiung
	Reperatur -- > Reparatur
	sodaß -- > so daß
	Widerstand -- > Widerstand

*Bild 2. Beispiele für wissensorientierte Fehlerkorrekturen*

haus	Substantiv;	Endungen: 0, e, es; Fuge: 0 (leer)
häu-ser	Substantiv;	Endungen: 0, n; Stamm: Haus; Fuge: 0
son-ne	Substantiv;	Endungen: 0, n; Fuge: n
se-hen	Infinitiv;	Endungen: 0, s, d, de, den, ...
sah	Verbstamm;	Endungen: 0, st, t, en, et; Stamm: sehen
bist	Wortform;	Stamm: sein (Verb)
schön	Adjektiv;	Endungen: 0, e, es ..., er, ere, ..., ste ...

Das der Softex-Rechtschreibung zugrundeliegende elektronische Wörterbuch mit Markierungen dieser Art umfaßt z. Z. rund 90 000 Einträge, in einer Ausbauversion zur Indexierung (vgl. unten) bereits über 140 000 Elemente (Speicherbedarf auf Festplatte: zwischen 1,2 und 2,4 Megabyte je Anwendungsfall). Damit wird inklusive der Dekompositions- und Derivationsverfahren für beliebige (Zeitungs-)Texte bereits eine Trefferquote (identifizierte Textwortformen) von über 98% erreicht; bei den nicht identifizierten Wörtern handelt es sich zumeist um Abkürzungen und Eigennamen.

*Bild 3. Beispielkodierungen Rechtschreibwörterbuch PRIMUS  
(in expliziter Darstellung)*

„stammformenorientierte Worlanalyse“						
Stamm	/ form	/ en	/ orientiert	/ e	// Wort	/ analyse
Sub. Fuge: 0	Sub. Fuge:en	Fuge	Ptz. II Endg. 0, e...	Endg.	Sub. Fuge:0	Sub. Endg:0

Bild 4. Beispiel der Wortidentifikation der Rechtschreibhilfe

Langform	Kurzform	proz. Einsparung (Schreibaufwand/ Text)
die	d	1,61
der	r	1,57
und	u	1,34
unter	ut	0,85
nicht	ni	0,60
...		
arbeit	rb	–
mensch	mx	–
...		
mit freundlichem Gruß	mfg	–

Bild 5. Beispiel für MKS-Kürzel

Ausgangswort	Relation	Relatum
Ministerpräsident	KOM	Minister
Ministerpräsident	KOM	Präsident
systemintern	KOM	System
systemintern	KOM	intern
Wohnung	DER	wohnen
Wohnung	DER	bewohnbar
Fachinformation	KOM	fachbezogen
Fachinformation	KOM	fachlich
Fachinformation	KOM	Information
...		

*Bild 6. Beispiele zur morphosyntaktischen Relationierung*



„... Von Ungeträumtem geätzt,  
wirft das schlaflos durchwanderte Brotland den Lebens-  
berg auf. ...“

Deskriptoren: ungeträumt / ätzen / aufwerfen / durchwan-  
dern / Brotland / Lebensberg

(neue) Relationierungen:

ungeträumt	-- träumen	schlaflos	-- schlafen
durchwandern	-- wandern	aufwerfen	-- werfen
Brotland	-- Land	Brotland	-- Brot
Lebensberg	-- Leben	Lebensberg	-- Berg

*Bild 7: Auszug aus der automatischen Texterschließung  
(Text aus dem Gedicht „Atemwende“ von P. Celan)*

<p><u>Beispieltext:</u>          „Spezial-Überwendlichschnellnäher mit gerader Nadel und 3 Fäden (Stichtyp 504), extra hohem Hub ... mit Schnellverstellung und Feineinstellung, Druckknopfstichsteller, Kantenbeschneideeinrichtung, ...“</p> <p><u>Translationshilfe PRIMUS-TRS (Deutsch-Englisch):</u>          Spezial-Überwendlichschnellnäher &gt;&gt; special high speed overedger          Nadel &gt;&gt; needle          Fäden - &gt; Faden &gt;&gt; thread          Stichtyp &gt;&gt; stich type          hohem - &gt; hoch &gt;&gt; high          Hub &gt;&gt; stoke / travel          Schnellverstellung &gt;&gt; top feed          Feineinstellung &gt;&gt; quick-change          Druckknopfstichsteller &gt;&gt; push-botton feed regulator          Kantenbeschneideeinrlchtung &gt;&gt; edge trimming mechanism</p>
---

Bild 8. Einfache Übersetzungshilfe PRIMUS-TRS (Auszug)

## 5 Perspektiven

Zum Abschluss soll versucht werden, eine kleine Vorausschau zu den Entwicklungen und Möglichkeiten der nächsten Jahre zu geben. Dabei will ich mich auf realistische Themen beziehen, d.h. Lösungen, wie sie heute schon experimentell oder punktuell vorliegen, in der Annahme, dass sich die Rahmenbedingungen im jetzigen Trend weiter entwickeln.

- Die maschinelle und computergestützte Übersetzung wird - getragen von den technischen Entwicklungen im PC-Bereich und der Datenfernverarbeitung - einen deutlichen Anteil gewinnen. Die *Telettranslation* wird es möglich machen, über Mailboxen Übersetzungsaufträge an Computerzentren abzuliefern und mit Hilfe von dezentralen, auf die Postedition von Computerübersetzung bzw. die Nutzung von Terminologiebanken spezialisierten Übersetzerbüros kostengünstig zu bearbeiten.

- Bei dem Zugang zu weltweiten Textdatenbanken werden Terminologie- und Übersetzungshilfen Sprachbarrieren überwinden helfen. Dies gilt v.a. für die Fachinformation, etwa im Patentwesen.
- Wörterbücher und Enzyklopädien werden auf PC - in umfangreichen Sammlungen auch über CD-ROM und Bildplatte - "vor Ort" zur Verfügung sein; entsprechendes gilt für Grammatik-, Rechtschreib- und Stilhilfen.

Vor diesem Hintergrund wird sich auch die Ausbildungssituation entsprechend ändern. So wie heute der Taschenrechner längst zu einem unentbehrlichen Hilfsmittel geworden ist (wobei man allerdings in manchen Fällen bedauern könnte, dass die Fähigkeit des Kopfrechnens aus Bequemlichkeit schon fast verloren zu gehen scheint), wird der Textcomputer zu Hause oder auch in der Schule ein alltägliches Instrument sein.

Manch einer wird diese Entwicklung bedauern, etwa unter dem Aspekt, dass ein so wichtiger Teil der menschlichen Kreativität wie die Sprachfähigkeit nun auch noch wenigstens in Teilbereichen von "Maschinen" übernommen wird. Manch einer - und zu diesen rechnet sich auch der Autor - wird allerdings auch auf die Möglichkeit hinweisen, dass auf diesem Wege manchem zu neuer Kreativität verholfen werden kann, da die nicht immer nur positiv einzuschätzenden Normen der Rechtschreibung und Silbentrennung - um nur die einfachen Beispiele zu nennen - keine zusätzlichen Barrieren mehr darstellen. Auf lange Sicht werden zudem die komplexeren Probleme des Sprachverstehens nicht durch Computerintelligenz ersetzt werden können.

Bei all dem ist zu berücksichtigen, dass der Computer weitgehend nur im instrumentellen Teil der sprachlichen Kommunikation, d.h. im Bereich des (internationalen) Wissenstransfers, sinnvolle Anwendung finden wird. So wie Wörterbücher und Grammatiken nur eine dienende Funktion haben, sollte der Sprachcomputer ein handhabbarer Gebrauchsgegenstand sein, ganz unserem *Zeitalter der Information* angemessen. Dadurch, dass er uns von trivialen Tätigkeiten entlastet, kann er uns den nötigen Freiraum und auch die Zeit geben, die komplizierten Themen anzugehen, die uns Gegenwart und Zukunft bringen.