

PIK – Praxis der Informationsverarbeitung und Kommunikation 10 (1987) 1, S. 38-45, Bürokommunikation  
Carl Hanser Verlag, München 1987

H. Zimmermann, Fachrichtung Informationswissenschaft / Universität des Saarlandes

### **Textverarbeitung im Rahmen moderner Bürokommunikationstechniken**

Zimmermann, H., Studium Germanistik/Geschichte (Staatsexamen), Promotion Sprachwissenschaft/Informatik (Saarbrücken). Von 1974 bis 1980 Professor für Nicht-numerische Datenverarbeitung (Regensburg), seit 1980 ordentlicher Professor für Informationswissenschaft an der Universität des Saarlandes. Seit 1984 zugleich Leiter des Instituts der Gesellschaft zur Förderung der angewandten Informationsforschung e.V. an der Universität des Saarlandes (IAI). Wiss. Geschäftsführer der Softex, Software-Institut für maschinelle Textverarbeitung GmbH.

Die angemessene Ver- und Bearbeitung natürlicher Sprache mit dem Computer ist trotz nunmehr jahrzehntelanger Forschung und Entwicklung noch ein großes Desiderat. Als Zielpunkte kann man heute (nach wie vor) die qualitativ hochstehende maschinelle Übersetzung beliebiger Texte (HQMT - High Quality Machine Translation) und daneben die sinnbezogene Erkennung beliebiger fließend gesprochener Sprache (Speech Recognition) bezeichnen. Mit den Forschungen zu natürlichsprachigen Mensch-Maschine-Schnittstellen bei Informationssystemen, insbesondere mit Bezug auf den natürlichsprachigen Dialog bei Expertensystemen, hat die Sprachdatenverarbeitung einen neuen Akzent erhalten, indem hier nicht mehr die Bearbeitung beliebiger Daten, sondern „nur“ eines anwendungsbezogenen Ausschnitts im Vordergrund steht. Auch dies aber ist in vielen Fällen scheinbar uferlos.

Mit der technischen Entwicklung hat sich auf der anderen Seite für die Bürokommunikation, insbesondere die Textverarbeitung im betrieblichen (aber auch individuellen) Rahmen, inzwischen ein breites Anwendungsfeld erschlossen, für das nach wie vor dringend (Teil-)Lösungen in der Verarbeitung natürlichsprachiger Daten (Texte) gesucht werden. Es stellt sich dabei auch die Frage, inwieweit die bisherigen Ergebnisse der jahrzehntelangen Forschung und Entwicklung nutzbringend in die Praxis transferiert werden können. Eine Hauptschwierigkeit, der es zu begegnen gilt, ist dabei eher psychologischer Art. Während man vom Taschenrechner - mit Bezug auf seine Funktionen -

stets zuverlässige, „exakte“ Problemlösungen erwarten kann, sind „perfekte“ Lösungen in der Bearbeitung beliebiger Texte nicht möglich, zudem ist das Wissen der Benutzer in Einzelfällen den Systemleistungen überlegen. Die "Systemfehler" - die meist in der systematischen, aber verständnislosen Anwendung formaler Grammatiken liegen - werden vom Benutzer in der Regel nicht durchschaut, so dass er auch der Autorität des Systems in solchen Fällen misstraut, in denen dieses die korrekte Lösung anbietet.

Unter diesen Rahmenbedingungen ist es schwer, verlässliche Prognosen über die Entwicklung der Textverarbeitung in der Bürokommunikation zu geben. Ziel dieses Beitrags ist daher, den derzeitigen Stand in einigen Themenbereichen wiederzugeben. Ausgeklammert werden insbesondere die experimentellen Anwendungen in der „Künstlichen Intelligenz“, aber auch die Bereiche der Erkennung gesprochener Sprache und der Sprachsynthese. Schwerpunktmäßig soll die Strategie verdeutlicht werden, mit der man - meist in kleinen Schritten - akzeptable Lösungen erreichen kann. Dies geschieht exemplarisch unter Rückgriff auf Erfahrungen des Autors, insbesondere auch bei der Entwicklung und dem Einsatz eigener Produkte.

## 1 Der technische Rahmen

Sprachdatenverarbeitung im Büro geschieht nicht „absolut“, sondern ist eine Funktion, die eingebettet ist in verschiedene Problemlösungsanforderungen eines (meist vorgegebenen) Rahmensystems. Der Rahmen kann ein Editor sein (ein dediziertes Textsystem, ein Textsoftwarepaket), aber auch ein Archivierungs- und Retrievalsystem, schließlich - mehr spezialisiert - ein Übersetzerarbeitsplatz, ein Text-to-Voice-System.

Der Rahmen muss geeignete Schnittstellen aufweisen, die einerseits eine (häufig notwendige) Interaktion mit dem Benutzer ermöglichen, andererseits muss er über Funktionen verfügen, die die Ergebnisse des Prozesses in vollem Umfang auszuwerten erlauben. Ein noch so guter Silbentrennalgorithmus wird im Deutschen z. B. wenig nützen, wenn der Editor nicht in der Lage ist, die Differenzierung bei „ck“ („k-k“) und der Konsonantenverdoppelung („Schiffahrt“ -- > „Schiffahrt“) beim Zeilenumbruch bzw. Druck zu

realisieren oder unästhetische Trennstellen zu eliminieren. Stil- und Grammatikhilfen müssen im aktuellen Fall (bei der Textgenerierung) angesprochen, eine Recherche mit Text- oder Schlagwörtern muss durch ein Thesaurussystem unterstützt werden können. Nicht zuletzt spielen auch die allgemeinen technischen Rahmenbedingungen eine (wenn auch zunehmend vernachlässigbare) Rolle: die bestehenden Lösungen sind vielfach gerade durch diese Restriktionen (mangelnde Geschwindigkeit, geringer Speicher- raum) bei Bürosystemen geprägt.

Die Rahmenbedingungen stehen daneben in Wechselwirkung zu den funktionalen Lö- sungen der Sprachdatenverarbeitung: Da es bislang eine Reihe von Schwachstellen technischer wie inhaltlicher Art bei der Sprachanalyse selbst gab, sind die Rahmenlö- sungen nicht auf komplexere Verfahren eingestellt; umgekehrt behindern unzureichende Schnitt- und Abnahmestellen eine adäquate Umsetzung der inzwischen angebotenen bzw. prinzipiell erreichbaren Lösungen.

## 2 Die Benutzer

Bestimmte Funktionen der Sprachdatenverarbeitung wie vollautomatische Silbentren- nung und Rechtschreibhilfen, aber auch Worterklärungs-, Stil- und Grammatikhilfen sind ein Desiderat für jeden Textarbeitsplatz. Überall da, wo heute neben dem elektronischen Schreibplatz ein (gedruckter) „DUDEN“ steht, ist prinzipiell ein Bedarf für derartige Hilfen gegeben. Die Sekretärin, aber auch der Sachbearbeiter (und nicht zuletzt der Manager) sind also - weitgehend unabhängig von ihrer individuellen Qualifikation - Zielgruppen der textbezogenen Bürokommunikation. Dies ist bereits ein sehr weites Feld, hinzu kommen im betrieblichen Bereich in aller Regel branchenspezifische Spezialisierungen. Daneben sind Funktionen für hochspezialisierte Tätigkeiten von Interesse, so z. B. eine Maschi- nenkurzschrift für Vielschreiber oder Übersetzungshilfen für die fremdsprachige Kom- munikation und den professionellen Übersetzer.

Die vorliegenden Erfahrungen zeigen inzwischen, dass man in diesem Bereich trotz an- fänglicher Skepsis der „Praktiker“ die elektronische Unterstützung annimmt. Da es keine „perfekte“ Lösung gibt, kann als Faustregel gelten: Die Funktionen müssen letztendlich

eine geringere Fehlerrate aufweisen, als sie im Alltag der normalen (d. h. intellektuellen) Textbearbeitung auftritt. Der Zeitaufwand muss - bezogen auf den Bearbeitungsprozess der Textgenerierung und Kontrolle - etwa im Rahmen des intellektuellen Gesamtaufwands liegen. In diesen Gesamtzusammenhang muss auch eine Bewertung der inzwischen vorliegenden Verfahren gestellt werden, „linguistische“ Bewertungen helfen kaum weiter.

### 3 Die Lösungen

Bei den verfügbaren Lösungen zur schreibunterstützten Textverarbeitung ist zu unterscheiden zwischen integrierten Verfahren und externen Anbindungen. Integrierte Lösungen sind auf die Funktionen des Editors oder Textsystems voll abgestimmt; sie erlauben eine kontrollierte Interaktion während des Textbearbeitungsprozesses oder werden selbsttätig bei erforderlichen Systemaktionen (z. B. Zeilenumbruch) ausgelöst. Externe Lösungen überlagern entweder den Textbearbeitungsprozess oder werden nach Beendigung der Texterfassung vom Betriebssystem her als eigenständiger Prozess aufgerufen. Bei der integrierten Lösung ist ein Wechsel des Editors in der Regel mit einer Anpassung an ein neues funktionales Konzept verbunden; die externe Lösung steht meist für verschiedene Editoren zur Verfügung.

#### 3.1 Vollautomatische Silbentrennung

Die vollautomatische Silbentrennung sollte eigentlich zu den Standardfunktionen einer elektronischen Textbearbeitung gehören. Es ist heute z. B. möglich, im Deutschen - das ja in dieser Hinsicht nicht gerade zu den einfachen Sprachen gehört - schon bei algorithmischen Lösungen eine Trenngenauigkeit von über 99 Prozent - bezogen auf die Alltagssprache im Büro - zu erreichen. Dies ist qualitativ mehr, als die „Durchschnittsschreibkraft“ leistet. Blocksatz und Proportionalschrift sind überhaupt nur voll nutzbar unter Integration vollautomatischer Trennfunktionen. Dennoch verfügt eine Reihe von Textsystemen und Editoren nicht über eine derartige Funktion, sondern versteht unter „automatischem Randausgleich“ einen Zeilenumbruch ohne jede Trennung oder aber nur nach einer intellektuellen Interaktion. Es kommt hier übrigens weniger auf die korrek-

te Behandlung der „ck“-Lösung oder der Konsonantenverdopplung an: Im deutschen Standardtext tritt das „ck“-Problem etwa bei einem Prozent der Fälle, die Konsonantenverdopplung in weniger als einem Tausendstel der Belege auf, so dass mangelnde Druck- und Umbruchfunktionen durch ein „Verbot“ der Trennung an derartigen Stellen leicht „überspielt“ werden können. Problematisch ist vielmehr die korrekte Behandlung der Kompositionen und Wortableitungen. Als Nutzer sollte man zudem darauf achten, die meist systemseitig vorhandene Ausnahmeliste ergänzen zu können. Denn gerade hier gilt: selbst das komplexeste System ist nicht fehlerfrei.

Als Beispiel für ein modernes Verfahren kann die algorithmische Lösung der automatischen Silbentrennung von Softex (System PAST) gelten. Sie arbeitet mit einer systemseitig vorgegebenen, nutzerseitig pflegbaren (d. h. ergänzbaren) Ausnahmeliste. Die Ausnahmeliste zum Deutschen umfasst dabei (bei derzeit rd. 3000 Elementen) auch Wortanfänge und Flexionsformen, die „an sich“ vom Algorithmus selbst regelhaft korrekt getrennt würden, aber in Wortzusammensetzungen und -ableitungen an den Teilwortübergängen zu Fehlern führen würden. So ist beispielsweise „firmen“ als Flexionsform in der Ausnahmeliste verzeichnet, damit „Firmenangestellter“, „Firmenauskunft“ ... nicht fälschlich getrennt werden. Es hängt dabei entscheidend von der Strategie (d. h. dem Zusammenspiel zwischen Regelanwendung, Trennwahrscheinlichkeitsverfahren und Ausnahmelistenanwendung) ab, wenn eine hohe Trenngenauigkeit erzielt werden soll.

<u>Wortformen:</u>	
wor-an	= unästhetische Trennung
wor-unter	≡ 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-ke-l-an-satz ...)

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

Bei der hohen Trenngenauigkeit der Softex-Trennverfahren (im Deutschen bei Fließtext rd. 99%) genügt es meist für Anwender, weitere Wortformen der Liste hinzuzufügen.

### 3.2 Rechtschreibhilfverfahren

Da heute vielfach mit dem Slogan „Der DUDEN im Computer“ u. ä. geworben wird, muss zunächst einmal eine Klarstellung erfolgen: Gedruckte Wörterbücher und Grammatiken haben meist eine andere Aufgabe als elektronische Systeme. Es geht in der elektronischen Textbearbeitung nicht einfach nur darum, gedruckte Wörterbücher elektronisch verfügbar zu machen (obwohl dies an sich auch schon ein interessanter, mit der Integration der CD-ROM sich sicherlich noch weiter entwickelnder Aspekt sein wird). Im Vordergrund der elektronischen Anwendung steht die unmittelbare Nutzung sprachdatenverarbeitender Prozesse. Das System soll so weit wie möglich selbst entscheiden, ob ein Wort den sprachlichen Regularitäten entspricht und im vorliegenden Textzusammenhang korrekt gebraucht ist. Um ein Beispiel zu geben: Die Wörter „Nachname“ und „Nachnahme“ sind - für sich betrachtet - beide korrekt, in den Verbindungen „sein Nachname ist Müller“ oder das Buch kommt per Nachname“ aber falsch. Die gedruckten Wörterbücher geben hierzu eine kurze Worterklärung, überlassen die korrekte Anwendung aber dem Leser. Umgekehrt geben sie dem Benutzer (warum eigentlich?) keine Hilfe zur Korrektur (z. B. durch „Falscheinträge“ im Wörterbuch mit Verweis auf die richtige Schreibform), wenn er ein Wort aus Mangel an Wissen falsch schreibt, sondern verlassen sich auf seine „Suchintelligenz“.

Dagegen ist es verhältnismäßig leicht, dem Computer beizubringen, wie ein falsches Wort zu korrigieren ist (um z. B. aus „Lybien“ automatisch „Libyen“, aus „richtungsweisend“ „richtungweisend“ zu machen u. a. m.)

Es ist also nicht sinnvoll, einfach ein Wörterbuch wie den DUDEN als „elektronische Liste“ zu verwenden, wenn man eine deutsche Rechtschreibhilfe programmiert: vielmehr sind Strategien und Funktionen zu entwickeln, um möglichst Fehler gleich zu korrigieren, um Wortzusammensetzungen, die man nicht in Wörterbüchern findet, als korrekt zu akzeptieren, um eine Wortidentifikation gleich zur Silbentrennung mitzunutzen usf. Ähnliches gilt für Worterläuterungen, Stil- und Grammatikhilfen.

Abonement -- > Abonnement	Rythmus--> Rhythmus
Adresse -- > Adresse	Weisbrot -- > Weißbrot
arhythmisch -- > arrhythmisch	achsial -- > axial
Balet -- > Ballett	Apartheit -- > Apartheid
Dahrlehen -- > Darlehen	Balaleika -- > Balalaika
Diphthong --> Diphthong	chifriert-- > chiffriert
emfindlich -- > empfindlich	Dedektiv -- > Detektiv
Hypothenuse -- > Hypotenuse	Embryo -- > Embryo
Interresse --> Interesse	Entgelt --> 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*

Die einfachste Strategie für eine Rechtschreibhilfe ist, einen Text mit einer Liste von Wortformen abzugleichen und alle Wörter als korrekt zu akzeptieren, die in dieser Liste sind. Mit ein paar Dutzend hochfrequenter Wörter hat man rasch eine Trefferquote von 50 Prozent, unter Nutzung von Frequenzlisten erreicht man relativ leicht im Deutschen 80 Prozent (in anderen Sprachen, die wenig komponierend sind, auch über 90 Prozent). Dieses Vorgehen ist meist mit dem Verfahren verbunden, dem Benutzer eine einfache Möglichkeit zu geben, die elektronische Wortformenliste (das „Wörterbuch“, „seinen“

„DUDEN“) anhand seiner Texte zu erweitern. Da hierbei auf stark komprimierende Speichertechniken zurückgegriffen werden kann und die Identifikationsgeschwindigkeit aufgrund des einfachen Abgleichs sehr hoch ist, sind diese Verfahren in der Praxis von einiger Bedeutung. Für Sprachen mit starker Flexion und Komposition (wie das Deutsche) erscheinen sie jedoch - v. a. auch gemessen an den zunehmenden Geschwindigkeiten der Rechner und den wachsenden Speicherkapazitäten - wenig zukunftsorientiert. Der Benutzer muss zudem (ständig) erhebliche „Wörterbucharbeit“ investieren, um die Rechtschreibhilfe auf seine Texte anzupassen.

Es kann jedoch davon ausgegangen werden, dass im flexionsmorphologischen Bereich inzwischen ausgereifte und z. T. auch praxiserprobte linguistische Modelle vorliegen, die auf die praktischen Anforderungen der Rechtschreibkontrolle umgesetzt werden können. Zur Verdeutlichung der inzwischen verfügbaren Möglichkeiten soll wiederum auf das Beispiel der Softex-Lösung zurückgegriffen werden. Die Softex-Rechtschreibhilfe mit integrierter Silbentrennung ist am PC-Markt unter dem Namen PRIMUS als externe Anbindung für eine Reihe von Editoren (WordStar, WordStar 2000, WordPerfect, Euroscript ...) für verschiedene Sprachen (Deutsch, Englisch, Französisch) verfügbar und wird daneben intern bei NIOS-TOP-EDIT (Nixdorf) sowie WordStar Extra (MicroPro) eingebunden.

PRIMUS liegt ein umfangreiches elektronisches Stammwörterbuch zugrunde. Zu jedem Eintrag sind u. a. markiert: die Wortklasse, die Bindefähigkeit bei Wortzusammensetzungen und die Flexions- oder Beugemöglichkeiten. Zusätzlich ist die Silbentrennung explizit vermerkt.

Der Identifikationsalgorithmus (Programmiersprache „C“) schlägt jeweils im Wörterbuch nach, überprüft die zugelassenen Endungen, prüft ggf. auf Wortzusammensetzungen und -ableitungen sowie in begrenztem Umfang auch auf korrekte Groß- oder Kleinschreibung. Die lexikalisch verzeichneten Silbentrennungen werden analog den sprachlichen Regeln (v. a. bei Wortzusammensetzungen und -ableitungen, aber auch bei der Flexion) angepasst und für die Weiterverarbeitung zur Verfügung gestellt. Der Benutzer ist zudem in der Lage, über entsprechende Menüs das Wörterbuch zu erweitern; dies ist



v. a. für individuelle Abkürzungen und Eigennamen nötig, aber gelegentlich auch bei extrem fachspezifischen Anwendungen. Eine wichtige Funktion stellt die (oben bereits angesprochene) Integration der Korrektur sog. wissensorientierter Rechtschreibfehler dar: Dieses inzwischen im Deutschen auf über 3000 Einträge angewachsene Inventar erleichtert die Texterstellung und -korrektur erheblich, da viele sprachlich gängige typische Schreibfehler systemseitig schon vorgemerkt sind, der Benutzer außerdem seine persönlichen „Schwächen“ ein- für allemal festhalten kann und somit ein leidiges Problem der Unsicherheit los wird.

Es ist hier nicht möglich, alle Problemfälle zu schildern, die in diesem Bereich noch einer Lösung bedürfen. Offenkundig ist jedoch, dass eine Reihe von Rechtschreibproblemen über eine Einzelwortanalyse nicht gelöst werden können. Dies betrifft die Groß-/Kleinschreibung, z. B. in Wendungen wie „in Bezug auf“ | „mit Bezug auf“, aber auch bestimmte Fälle der Zusammen- und Getrennschreibung. Hier können - wenn überhaupt - erst komplexere Analysen helfen.

haus	Substantiv;	Endungen: 0, e, es; Fuge: 0 (leer)	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 (Zeitungstexte) bereits eine Trefferquote (identifizierte Textwortformen) von über 98% erreicht; bei den nicht identifizierten Wörtern handelt es sich zu meist um Abkürzungen und Eigennamen.
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 ...	

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

„stammformenorientierte Wortanalyse“

Stamm	form	en	orientiert	e	Wort	analyse
Sub.	Sub.	Fuge	Ptz. II	Endg.	Sub.	Sub.
Fuge: 0	Fuge:en		Endg. 0	e...	Fuge:0	Endg:0

Bild 4. Beispiel der Wortidentifikation der Rechtschreibhilfe

Eine interessante Variante, die sich aus der Technik der Korrektur wissensorientierter Rechtschreibfehler ergibt, ist die Integration einer Art „Maschinenkurzschrift“ mit der normalen Schreibmaschine: Hierbei stützt sich das System auf eine Liste (Teilwörterbuch) speziell abgekürzter Wörter, die bei der Texterstellung wissentlich benutzt und später automatisch in die Normal- bzw. Langform umgesetzt werden. Das von Softex entwickelte MKS-System ist so gestaltet, dass es bis auf einbuchstabige (hochfrequente) Wortkürzungen mit den Normalschreibweisen voll kompatibel ist, so dass praktisch eine stufenlose Verwendung dieser Kürzel bzw. Textkonstanten möglich ist. Es lässt sich leicht errechnen, dass sich bereits bei der systematischen Kürzung hochfrequenter Wörter deutliche Zeitersparungen erzielen lassen: 10% Reduktion des Erfassungs- und Speicheraufwands sind bei der Verwendung von gut zwei Dutzend Abkürzungen erzielbar; ca. 200 Kürzungen von Wörtern und Wendungen führen zu einer Aufwands- und Speicherreduktion von etwa 20%, im günstigsten Falle lassen sich bei konsequenter Anwendung dieses Kürzelsystems rd. 50% Zeiteinsparung - wohlgemerkt unter Verwendung der normalen Schreibmaschinentastatur und bei Integration in den normalen Wortschatz – erreichen (vgl. Brenner/Zimmermann 1986).

Langform	Kurz-	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

Hierzu ein kleines Beispiel für einen entsprechend abgekürzten Text und die diesbezügliche „Auflösung“:

Sgdsgh, we Y ä öm gtyn Tlfgräc xo wisn, kön w ön Ftag vm 10. dm gträcl öhhn. -

->

Sehr geehrte Damen, sehr geehrte Herren, wie Sie aus unserem gestrigen Telefongespräch schon wissen, können wir unseren Auftrag vom 10. dieses Monats beträchtlich erhöhen.

Es bleibt weiteren Experimenten überlassen zu prüfen, inwieweit ein derartiges Verfahren einmal praxiswirksam werden kann.

### 3.3 Grammatik- und Stilhilfen

Wenn man sich weitergehende Verfahren überlegt, muss man sich bewusst sein, dass es bislang kein maschinelles System gibt, das beliebige Texte auf grammatische oder semantische Korrektheit hin überprüfen kann. Umgekehrt setzen vielmehr alle allgemein bzw. breiter anwendbaren Analyse- oder Übersetzungssysteme die vorherige korrekte (intellektuelle) Anwendung sprachlicher Regeln voraus, z. B. inklusive der Regeln der Zeichensetzung. Zudem sind komplexe Systeme gegenwärtig noch sehr rechenaufwendig, so dass man trotz gesteigerter Leistungen im Bereich der Bürokommunikation (bzw. des Home-Computers) auf längere Sicht noch keine Wunderdinge erwarten darf. Elektronische Grammatik- und Stilhilfen haben daher in erster Linie eine Beratungs- und weniger eine Kontrollfunktion.

Die Benutzung herkömmlicher (gedruckter) Wörterbücher und Grammatiken geht von vergleichbaren Strategien aus: In der Regel dient ein Stichwort (ein textrelevantes Problemwort oder auch ein grammatischer Begriff) als Indikator. Grammatikalische (wortbezogene) Problemfälle können sein:

Rechtschreibung (Dehnung, Umlaut ...), z. B. „Mahl“ / Mal“

Worterklärung (Fremdwort, Mehrdeutigkeit), z. B. „Etikette“  
Groß-/Kleinschreibung, z. B. „achtel“/„Achtel“  
Genus, z. B. der/die/das „Dschungel“  
Pluralbildung, z. B. „die Synonyme/Synonyma“  
Zusammen- od. Getrennschreibung, z. B. „abendlang/Abende lang“  
Deklination/Konjugation, z. B. „Biß“ - des Bisses, die Bisse  
Silbentrennprobleme (Konsonantenausfall), z. B. „die Pappuppe“  
Zeichensetzung, z. B. Komma vor „aber“.

Bei der bei Softex in Entwicklung stehenden elektronischen (wortbezogenen) Grammatikhilfe kann der Benutzer im Einzelfall die Hilfe (evtl. auf einen Problemtyp eingeschränkt) aktivieren oder aber sich systematisch beraten lassen: die Hilfe wird in Form eines Kurztexes geliefert. Ggf. lassen sich (system- oder auch benutzerseitig) „Schwierigkeitsgrade“ einführen, die sich auf den Bedarf der individuellen Nutzung oder aber den (erwarteten) Kenntnisstand einer Zielgruppe (z. B. „Schüler“) ausrichten. Bei gleichartigen Fällen benutzen gedruckte Wörterbücher in den meisten Fällen Verweise im Wörterbuch. Eine Problemlösung kann somit oft erst nach längerem Blättern gefunden werden. Ähnlich wird mit einem elektronischen Wörterbuch verfahren, doch lässt sich in der „elektronischen Hilfe“ dieses Blättern automatisieren: ein systeminterner Verweis von „viertel“ auf „achtel“ erlaubt es beispielsweise, die unter dem Stichwort „achtel“ gespeicherte Information beim Textwort „viertel“ anzustoßen u. a. m.

Ähnliches gilt für den Bereich der (wortorientierten) Stilhilfe: Hierbei sind analog zum traditionellen Verfahren in einem elektronischen Wörterbuch Angaben zur Sprachschicht (gehobene Sprache, Standardsprache, Vulgärsprache ...), zur Sprachepoche (z. B. „veraltet“, selten“) und zur Präferenz (z. B. Fremdwort gegenüber seiner Eindeutschung) verzeichnet. Hierzu gehören auch Begriffsrelationierungen (Synonyme, Ober-/ Unterbegriffe) und -erläuterungen (z. B. zu Fremd- oder Fachwörtern).

Auch hier ist eine unreflektierte Übernahme der Einträge aus gedruckten Wörterbüchern keineswegs sinnvoll. Während die traditionellen Inventare auf eine möglichst umfassende Darstellung der sprachlich-systematischen Möglichkeiten zielen, steht bei elektroni-

schen Wörterbüchern die Ausrichtung auf den praktischen Gebrauch (im Büro) im Vordergrund.

Elektronische Stil- und Grammatikhilfen werden vorwiegend bei der Erstellung von Texten zum Einsatz kommen. Die realisierbaren Möglichkeiten sind trotz der derzeitigen Grenzen der automatischen Texterschließung bereits derart vielfältig, dass heute noch nicht abgesehen werden kann, wo letztlich die Einsatzschwerpunkte liegen werden. Je ausgedehnter die elektronische Kommunikation v. a. in der Fachinformation wird, desto intensiver wird z. B. die Nutzung auch beim Verstehen elektronisch überspielter fremd-erstellter Texte sein. Die Entwicklung dieser Funktionen steht zudem erst am Anfang.

### 3.4 Indexierung und Retrieval

Die Thematik der Erschließung und Aufbereitung von Texten zum Zwecke der Ablage und des Wiederfindens kann an dieser Stelle nicht umfassend dargelegt werden. Hierzu sei auf die Studien von Zimmermann!/Kroupa/Keil 1983 und Krause 1986 verwiesen.

Im Bereich der Fachinformation auf Mainframe-Ebene sind inzwischen Erfahrungen v. a. zum sog. Freitext-Retrieval gemacht worden. Bekannte Beispiele sind die Anwendungen von STAIRS (IBM), etwa bei der Datenbank LEXINFORM der DATEV, und GOLEM-PASSAT (SIEMENS), etwa bei JURIS und der Patentdatenbank PATDPA. Inzwischen sind auch auf PC-Ebene ähnliche Systeme verfügbar, so z. B. das Verfahren FULL SEARCH der Fa. LASEC, Berlin.

Es stellt sich die Frage, inwieweit komplexere Verfahren der Freitexterschließung auf den Bereich der Bürokommunikation (und noch spezieller: auf die PC-Welt) übertragen werden können. Dabei sind zwei Anwendungsschwerpunkte vorstellbar: Einerseits soll über eine „vor Ort“ verfügbare Texterschließungs- und Retrievalkomponente ein Inhouse-Informationssystem (oder eine arbeitsplatzbezogene Textdatenbank) besser und leichter zugänglich werden, andererseits soll das Retrieval auf Fremdsystemen (Hosts) verbessert werden.

Mit dem von Softex entwickelten Verfahren PRIMUS-IDX, das inzwischen in Pilottests steht, soll der Benutzer v. a. von trivialen Operationen entlastet werden. Zu den neu entwickelten (unter MS-DOS und UNIX einsetzbaren) Funktionen gehören:

- die Reduktion von Textwortformen auf die Grundform  
(„gefundenes" -- > „finden"; „Häusern" -- > „Haus" ...)
- die (zusätzliche) Zerlegung von Zusammensetzungen in sinnvolle Teile  
(„Grammatikregel" --> Regel; Grammatik ...)
- die morphologische Relationierung  
(„wohnen" -- „Wohnung" -- „bewohnbar" -- „bewohnen" ...)
- die bedeutungsmäßige Relationierung  
(PKW -- Personenkraftwagen; PKW -- Auto; ...)
- die Mehrwortidentifikation und -relationierung  
(„juristisch" + „Person" -- > „juristische Person"

Die relevanten elektronischen Wörterbücher sind dabei - soweit nötig - um Grundformenverweise ergänzt (v. a. bei unregelmäßigen Formen), daneben ist ein morphologisch-syntaktisch orientiertes Relationenwörterbuch entwickelt worden, das Verweise von Wortzusammensetzungen und -ableitungen auf sinnvolle Teilwörter enthält und beim Retrieval später als „Thesaurus" zugeschaltet werden kann (eine Erweiterung dieses Relationenwörterbuchs um semantische Beziehungen ist vorgesehen).

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*

Für die praktische Anwendung der Indexierung im Deutschen stehen inzwischen bei PRIMUS-IDX ein elektronisches Wörterbuch mit rd. 140 000 (Stamm-) Einträgen und ein

Wörterbuch mit Beziehungsverweisen der geschilderten Art mit über 100 000 Relationierungen zur Verfügung. Das Verfahren selbst ermöglicht inzwischen prinzipiell auch die Erkennung von Mehrwortbegriffen („juristische Person“; „katholische Kirche“ ...) und diskontinuierlicher Verbsegmente („fasste .., zusammen“ -- > „zusammenfassen“).

Ein entsprechendes Update- und Wörterbuchpflegeverfahren stellt sicher, dass die Ergänzungen v. a. des Relationenwörterbuchs textabhängig (bei Bedarf vollautomatisch) erfolgen. Wörter, die für die Ablage und das Wiederfinden für nicht relevant gehalten werden, können durch eine benutzerseitig ergänzbare Selektionsdatei als Deskriptoren ausgeschlossen werden.

Für die Ablage und das Wiederfinden von Texten / Dokumenten im Büro ist es eine große Hilfe, wenn auch mit Begriffen gesucht werden kann, die im Text eines Dokuments (d. h. eines Briefes, Berichts, Protokolls, einer Notiz) vorkommen. Andererseits ist es nicht trivial, thematisch relevante Stichwörter aus einem Text herauszufiltern, so dass mit Ballast zu rechnen ist, auch wenn sog. Stoppwörter (wie DER, UND, MIT ...) noch leicht zu eliminieren sind. Ein weitgehend ungelöstes Problem ist die Behandlung der Mehrdeutigkeit (Bank – Geldinstitut / Sitzgelegenheit ...).

„... Von Ungeträumtem geätzt, wirft das schlaflos durchwanderte Brotland den Lebensberg auf....“

Deskriptoren: ungeträumt / ätzen / aufwerfen / durchwandern / 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)*

### 3.5 Translationshilfen

Die Sprachübersetzung ist ein kleineres, wenn auch nicht unbedeutendes Marktsegment der Bürokommunikation. Aus der Sicht der Anwendung lassen sich hier zwei Schwerpunkte unterscheiden: Einerseits die maschinelle Unterstützung der fremdsprachigen Textgenerierung („Übersetzerarbeitsplatz“), andererseits die allgemeine Hilfe beim Verstehen fremdsprachiger Texte, insbesondere fremdsprachiger Korrespondenz. Der letztgenannte Bereich wird an Bedeutung gewinnen, je weiter die elektronische Kommunikation (z. B. über TELETEX) vordringt: wo heute das gedruckte Übersetzungswörterbuch am Arbeitsplatz steht, wird die elektronische Übersetzungshilfe relevant.

Es ist sinnvoll, eine Übersetzungshilfe nicht isoliert zu sehen, sondern mit den übrigen Funktionen zu verbinden: Rechtschreib- und Stilhilfe sowie die Indexierung (Grundformen- und Mehrwortermittlung) bilden wichtige Teilschritte für eine Übersetzungshilfe. Angesichts der möglichen Vielfalt von Übersetzungsäquivalenten zu einem Textwort sind Fachgebietskennungen und Bedeutungsdifferenzierungen wichtige Voraussetzungen für eine praxiswirksame Anwendung. Für den Fachübersetzer werden zudem Funktionen benötigt, die ihm eine intensive Pflege des Inventars und eine spezifische Ausrichtung auf seinen professionellen Bedarf ermöglichen, während für den Gelegenheitsnutzer eher ein umfangreicher Allgemeinwortschatz - wie ihn die gedruckten Wörterbücher aufweisen - elektronisch bereitgestellt werden muss.

Eine mögliche erste Stufe der Übersetzungshilfe stellt der textuelle Abgleich eines im Bürosystem vorliegenden Quelltextes mit einem mehrsprachigen elektronischen Wörterbuch dar. Hierbei werden - evtl.. nach Vorauswahl von Fachgebieten und unter Nutzung einer Selektionsdatei - parallel zu Elementen des Quelltextes die möglichen zielsprachigen Entsprechungen angeboten. Ein derartiges einfaches Verfahren ist z. Z. bereits in PRIMUS-TRS realisiert:

Beispieltext:

„Spezial-Überwendlichschnellnäher mit gerader Nadel und 3 Fäden (Stichtyp 504), extra hohem Hub ... mit Schnellverstellung und Feineinstellung, Druckknopfstichsteller, Kantenbeschneideeinrichtung,...“



Translationshilfe PRIMUS-TRS (Deutsch-Englisch):

Spezial-Überwendlichschnellnäher » special high speed overedger

Nadel » needle

Fäden - > Faden » thread

Stichtyp » stitch type

hohem - > hoch » high

Hub » stoke / travel

Schnellverstellung » top feed

Feineinstellung » quick-change

Druckknopfstichsteller » push-botton feed regulator

Kantenbeschneideeinrichtung » edge trimming mechanism

*Bild 8. Einfache Übersetzungshilfe PRIMUS-TRS (Auszug)*

Weitere Funktionen können sein:

- Elektronisches Nachschlagen eines Wortes
- Übernahme eines Übersetzungsäquivalents in einen (zu erstellenden) zielsprachigen Text
- Interaktive Wörterbuchergänzung durch den Nutzer.

Komplexe Übersetzungshilfen auf Büro- oder PC-Ebene stehen erst am Anfang ihrer Entwicklung. Es wird international zudem an Verfahren gearbeitet, die es ermöglichen, über Datenfernübertragung von Texten an ein Übersetzungs-Servicezentrum sog. Rohübersetzungen zu erhalten. Ein erstes Beispiel ist der Übersetzungsservice über das französische MINITEL/TELETEL (etwa vergleichbar mit dem deutschen BTX-System) mit Hilfe des SYSTRAN-Verfahrens. Es ist vorstellbar, dass sich weltweit eine Reihe solcher Service-Zentren entwickelt, die über Postnetze wie Datex-P (oder später ISDN) erreichbar sind und für verschiedene Sprachrichtungen entsprechende Dienstleistungen anbieten. Auch eine Entwicklung, deren Möglichkeiten erst angedeutet werden können.

#### 4 Perspektiven: Buch und Computer

Ein Blick hinter die Kulissen der Buchproduktion zeigt, dass die elektronische Datenverarbeitung längst wesentliche Anteile am Fertigungsprozess von Büchern, Zeitschriften, Zeitungen usf. besitzt. Der Bleisatz hat weitestgehend ausgedient, die moderne Produktionstechnik schließt die Nutzung elektronischer Texteditoren und -systeme bei der Generierung (z. T. schon durch die Autoren) wie auch den Photo- bzw. Laserdruck (DESKTOP-Publishing ist das neueste Schlagwort) mit ein. Diese Art des elektronischen Publizierens, an deren Ende ein Papierprodukt steht, wird kaum mehr diskutiert. Hierbei dient der Computer in erster Linie als Rationalisierungsinstrument.

Die prinzipiellen Möglichkeiten, das „Buch“ (Papierprodukt) durch elektronische Speicher, Datenfernübertragung und Bildschirm (besonders in der Bürokommunikation) zu ersetzen, sind bekannt. Offen ist, wann dieser Ersetzungsprozess in Richtung auf das „papierarme Büro“ praxiswirksam wird und mit welchen Marktanteilen die neuen Medien zu Lasten der Papierprodukte rechnen können. Die fortschreitende Miniaturisierung und das sich positiv entwickelnde Preis-Leistungsverhältnis am Computermarkt werden jedoch in naher Zukunft deutlich Wirkung zeigen. Es erscheint dabei müßig, exakte Prognosen aufzustellen, wie es auch wenig sinnvoll ist, eine Buch- und Lesekultur allein auf der Basis einer Speicher- und Distributionstechnik zu motivieren. Ein Fortschreiten der Technik in diesem Bereich ist (aus verschiedenen Gründen) nicht zu verhindern, wenn auch das Buch bestimmt nicht aus den Studier- und Wohnzimmern verschwinden wird.

Von dieser Entwicklung ist v. a. der professionelle Bereich, d. h. das Verlagswesen und der Buchhandel, betroffen. Dies gilt im o. a. Themenfeld besonders für lexikographische und enzyklopädische Verlage. Hier stellt sich zunächst ein Problem: die bestehenden gedruckten Datensammlungen sind - bei aller Fülle des gespeicherten Wissens - nicht auf die elektronische Nutzung ausgerichtet. Eine einfache Umsetzung (die seitens der Verlage in einigen Fällen versucht wird) ist - wie dargelegt - wenig erfolgversprechend und wird v. a. den Möglichkeiten der Datenverarbeitung nicht gerecht, die ja gerade in der Integration dieser „Wissensbasen“ in umfassendere Textbe- und -verarbeitungsprozeduren bestehen. In aller Regel sind daher die vorhandenen (auch edv-gespeicherten)

Materialien zunächst auf die neuen Verarbeitungsmöglichkeiten hin zu strukturieren und formalisieren. Daneben sind sie auf die möglichen Funktionen hin auszurichten: die aufgeführten Anwendungsbereiche sind hierfür ein Indiz.

Es ist unbestritten, dass die Erfahrungen der Verlagsredaktionen, die Marktkenntnisse der Verleger und auch die verfügbaren Materialien eine gute Ausgangsbasis für die anstehenden Entwicklungen bilden. Mit den sich entwickelnden technischen Möglichkeiten (Datenbanken, Netzwerke, Speichertechniken wie CD-ROM, Graphik-Systeme, um nur einiges zu nennen) wird ein geeignetes technisches Instrumentarium geboten, das für neue Produkte hervorragend geeignet ist.

Allerdings bedarf es noch erheblicher Investitionen sowohl in die Entwicklung der technischen Funktionen, die diese Daten in geeignete Produkte einbinden, als auch der (lexikalischen) Inventare selbst. Dabei sind in Zukunft in weiteren Bereichen, v. a. bei Enzyklopädiën und Wörterbüchern, auch Lösungen vorstellbar, bei denen eine allgemeine elektronische Wissensbasis für die „parallele“ Nutzung auf direkte Art (Papier/CDROM), im Lehr- und Lernbereich (elektronischer Vokabeltrainer) und im Büro (als Rechtschreib-, Stil-, Grammatik- und Übersetzungshilfe ...) Verwendung findet.

Es steht zu erwarten, dass mit der Realisierung dieser oder vergleichbarer Entwicklungen, auch und gerade in Verbindung mit der weiteren Miniaturisierung, der „sprachgestützten“ elektronischen Textbearbeitung im Büro etwa die gleiche allgemeine Bedeutung zukommt wie heute dem Taschenrechner. Es erscheint durchaus möglich, den Schreibenden von trivialer „Arbeit“ zu entlasten und dennoch eine hochstehende Sprachnorm zu bewahren. Es dürfte auch psychologisch interessant sein, inwieweit die Reduktion der „Angst“ vor Schreib- und Stilfehlern bei der Nutzung der Textverarbeitung gerade diejenigen zum „Mehr-Schreiben“ (oder überhaupt zum Schreiben) motiviert, den heute entsprechende (vermeintliche oder tatsächliche) Schwächen eben am Schreiben hindern.

Der Taschenrechner machte den Mathematiklehrer auch nicht überflüssig, im Gegenteil. Wenn eines Tages die Textverarbeitung in der Schule ebenso selbstverständlich sein

wird wie der Taschenrechner heute, dann werden Instrumente der Art, wie sie hier vorgestellt wurden, zum Standard des Schreibplatzes gehören. Eine Voraussetzung gilt es aber zu nennen: Die Fertigkeit des Maschinenschreibens ist inzwischen zu einer so wesentlichen Komponente für alle späteren Berufe geworden, dass sie eigentlich als „Kulturtechnik“ in den Lehrplan einer jeden Schulform gehört. Die „alte“ Schreibmaschine hat schon viel bewirkt, aber sie hat hier eindeutig ihre Grenzen. Die Zukunft gehört den komfortablen, schreibunterstützenden elektronischen Text- und Bürosystemen.

#### Literatur

Brenner, L.; Zimmermann, H.: Deutsche Einheitskurzschrift - Maschinenkurzschrift (MKS). Ein Vergleich. In: Forum Bürowirtschaft 1986 (1) S. 2-4; 1986 (2) S. 2-4.

Krause, J.: PADOK: Test und Vergleich von Texterschließungssystemen für das Deutsche Patent- und Fachinformationssystem. Regensburg 1986

Zimmermann, H.; Kroupa, E., Keil, G. (1983): CTX - ein Verfahren zur computergestützten Texterschließung. Saarbrücken.