Kapitel 3

Technische Dokumentation und Mehrsprachigkeit

3.1 Aspekte technischer Dokumentation

Technische Dokumentationen sind nicht nur eine eher untergeordnete Ergänzung eines technischen Produkts, sondern wichtiger Bestandteil, der in vielen Fällen seine Verwendung erst ermöglicht (vgl. Abb. 3.2). Im europäischen Binnenmarkt wird die Bedeutung technischer Dokumentationen noch zunehmen, da hier Produkthaftungsvorschriften u.a. das Vorliegen aller technischen Unterlagen in der Muttersprache des Konsumenten fordern werden.

Die EG–Produkthaftpflicht–Richtlinie vom 25. Juli 1985 hat durch die folgende Definition des Produktfehlers in Artikel 6 großen Einfluß auf die technische Dokumentation gewonnen:
„a. Ein Produkt ist fehlerhaft, wenn es nicht die Sicherheit bietet, die man unter Berücksichtigung aller Umstände, insbesondere
b. der Darbietung des Produkts,
c. des Gebrauch, mit dem billigerweise gerechnet werden kann,
d. des Zeitpunkts, zu dem das Produkt in den Verkehr gebracht wurde,
zu erwarten berechtigt ist.“

Abbildung 3.1: Die EG–Produkthaftpflicht–Richtlinie

3.1.1 Umfang

Technische Dokumentation kann enormen Umfang annehmen. Einige Beispiele:

- Die Betriebsanleitung für ein Atom–U–Boot soll gedruckt 18.000 kg wiegen (nach „ZEIT“ Nr. 2, 18. Januar 1993, S. 32).
- Die Flugzeugabteilung (‘Commercial Airplanes’) von Boeing soll im Jahr 1987 allein 18.564.000 neue Seiten an Dokumentation produziert haben. Die Seiten,

die im gleichen Jahr von Boeing an ihre Kunden ausgeliefert wurden, hätten aufeinandergestapelt den Mount Everest überragt (zitiert nach [Riehm 1992], der die Informationen wiederum von [Becker 1988] übernommen hat).

- Das Gewicht der Dokumentation zu einem einzelnen Flugzeug übersteigt offenbar gelegentlich das Gewicht des Flugzeugs selber (nach [Levine et al. 1991]).

3.2 Mehrsprachigkeit

3.2.1 Mehrsprachigkeit im Informationszeitalter

Die Frage der Mehrsprachigkeit ist nicht nur von hoher ökonomischer Relevanz, sondern hat insbesondere im europäischen Rahmen auch eine kulturelle Dimension. Dies wird von der EG–Kommission immer wieder deutlich betont, so z.B. in den vorgelegten Vorschlägen für ein europäisches Forschungs- und Technologieentwicklungsprogramm im Gebiet der sogenannten Sprachtechnologien (language technologies) [LT-1992].
Die drei Hauptziele dieses Programms sind:

- Erhöhung der Wettbewerbsfähigkeit der europäischen Industrie, Dienstleistungen und Verwaltungen,
- die Unterstützung der europäischen sozialen Strukturen durch Überwindung der Sprachbarrieren,
- die Erhaltung der kulturellen Vielfalt Europas durch den Schutz der verschiedenen europäischen Sprachen im elektronischen Zeitalter.

Aus diesen Zielen werden zwei Hauptlinien abgeleitet:

- Die Schaffung einer Sprachtechnologie–Infrastruktur in Europa dergestalt, daß alle Sprachen der Gemeinschaft in zukünftigen elektronischen Informationssystemen und -diensten angeboten werden können.
- Die Stimulation von Forschung, Technologieentwicklung und Demonstrationen von Sprachtechnologie in ausgewählten Anwendungsgebieten.

Zu den Anwendungsgebieten zählen

- Integrated Document Creation and Management,
- Multilingual Computer Services,
- Person–to–Person Communications,
- Translation Services and Language Learning.

Die ISO/IEC–Richtlinie 37 definiert die Gebrauchsanweisung folgendermaßen:

> Gebrauchsanweisungen sind Kommunikationsmittel wie Texte, Worte, Zeichen, Symbole, Diagramme, Illustrationen, hörbare und sichtbare Informationen, die einzeln oder in Kombination verwendet werden können; sie sind Mittel zur Informationsübertragung an den Benutzer, wie er ein Produkt sicher und bestimmungsgemäß verwenden soll.
>
> Solche Anweisungen können auf dem Produkt selbst und/oder seiner Verpackung und/oder in Begleitmaterialien (z.B. Prospekte, Handbücher, Audio- und Videokassetten, Mikrofilm/Mikrofiche, Disketten, Laserdisks/CDROM, WORM) gegeben werden.

Abbildung 3.2: Die ISO/IEC–Richtlinie 37

Für diese Anwendungsgebiete wurden vier sogenannte strategische Kernprojekte ausgewiesen. Dazu gehören:

- Integriertes System zur Software–Dokumentation,
- Interaktives multilinguales Multimedia–System,
- Pan–europäisches multilinguales Konferenzsystem,
- Integriertes multilinguales Online–Akquisitionssystem.

In der Sicht dieses Programms wird Sprachtechnologie wie folgt definiert:

> *'Language Technology (LT) therefore provides the means by which communication and information system services and products can accomodate different languages and therefore satisfy the needs of multilingual environment within the Community.'* ([LT-1992])

3.2.2 Technische Unterstützung von Mehrsprachigkeit

Mehrsprachigkeit wird bereits heute in Schnittstellen zu verschiedensten technischen Systemen unterstützt. Das reicht von einer sehr simplen Form des Wechsels der Sprache der vorgefertigten und kurzen Display–Anzeigen bei Kartentelefonen bis zu anspruchsvolleren Vorhaben. So soll es zukünftig möglich sein, beim digitalen Verkehrsrundfunk (RDS–TMC) Staumeldungen und Umleitungshinweise, die in einer codierten Form ausgestrahlt werden, im Autoradio jeweils in die eigene Muttersprache umgesetzt zu bekommen, unabhängig davon, wo man in Europa gerade unterwegs ist.

3.3 Behandlung mehrsprachiger Dokumente

3.3.1 Der herkömmliche Ansatz

Die große Mehrheit technischer Dokumente wird derzeit zunächst in einer Sprache erstellt. Diese Ausgangsversion wird dann an Übersetzer gegeben, um daraus Versionen für die Zielsprachen zu erstellen. Dadurch ergeben sich auf jeden Fall Zeitverzögerungen zwischen den Versionen für die unterschiedlichen Sprachen; ferner entsteht das Problem, wie die Verwendung einheitlicher Terminologie gewährleistet werden kann, insbesondere, wenn an einem größeren Dokument verschiedene Übersetzer gleichzeitig beteiligt sind.

Die Probleme mit mehrsprachiger Terminologie können entschärft werden, wenn Firmen für ihre Produkte Terminologie–Datenbanken aufbauen und diese dann in den Übersetzungsprozeß einbezogen werden. Solche Verfahren scheinen sich als Unterstützung menschlicher Übersetzer bereits durchzusetzen [Freibott 1992].

Anforderungen an die Übersetzung

Beim Stichwort „Übersetzung“ denken sicherlich die allermeisten zunächst an die Übersetzung literarischer Werke. Die Anforderungen an die Leistung des Übersetzers sind dabei besonders hoch; manche gehen sogar so weit, bei anspruchsvollen literarischen Werken eine gelungene Übersetzung in die Nähe einer „Neuschaffung“ des Werkes zu rücken. Übersetzer sehen sich in diesem Umfeld oft auch als Vermittler zwischen Kulturen.

Diese zu Recht hohe Wertschätzung der literarischen Übersetzung hat nicht zuletzt damit zu tun, daß es nicht nur um die Übertragung von Inhalten geht, sondern zusätzlich auch noch versucht werden muß, den Sprachstil, die sprachliche Gestaltung und andere, ein literarisches Werk auszeichnende Merkmale nachzubilden. Dies kann besonders dann ein anstrengendes Unterfangen werden, wenn die Sprachen große Unterschiede in ihren Möglichkeiten aufweisen oder wenn große kulturelle Distanzen vorliegen (wenn zum Beispiel, um die Handlungsweisen von beteiligten Personen zu verstehen, ein großes Hintergrundwissen über ihre Gesellschaftsordnung, die Traditionen usw. benötigt wird).

Die Übersetzung „schöner Literatur“ spielt in der Praxis allerdings nur eine marginale Rolle. Für das Jahr 1986 wird als weltweite Übersetzungsleistung die Zahl von 500 Millionen Seiten angegeben, wobei mehr als 100 Millionen auf Europa entfallen. Davon betrifft weniger als 1 Prozent schöne Literatur; 30 % sind von staatlichen Stellen, 50 % von Industrie und Handel beauftragt. Im letzteren Fall handelt es sich überwiegend um die Übersetzung technischer Dokumentation (Zahlen nach Vortragsunterlagen Prof. von Hahn).

An technische Texte und damit auch ihre Übersetzungen werden ganz andere Anforderungen als an literarische gestellt. Sie haben meist eine klare Funktion: Der Leser soll ein technisches Gerät nutzen, warten, gegebenenfalls reparieren können,

ohne Gefahren für sich und andere bzw. für das Gerät einzugehen. Bei der Übersetzung in eine andere Sprache ist die Erhaltung des Sprachstils usw. nicht erforderlich. Wichtig ist, daß die im Vordergrund stehende Botschaft des Textes klar, knapp und unzweideutig vermittelt wird. Sprachliche „Originaltreue“ ist in diesem Zusammenhang kein anzustrebendes Ziel.

3.3.2 Einsatz maschineller Übersetzungssysteme

Obwohl die Entwicklung von maschinellen Übersetzungssystemen immer wieder Perioden großer Erwartungen und anschließend folgender Enttäuschungen durchlaufen hat, sind Systeme am Markt und auch im Einsatz. Für keines dieser Systeme ist gewährleistet, daß ein vollautomatischer Übersetzungsvorgang stattfindet.

Lediglich in ganz eingeschränkten Gebieten ist dies bisher erreichbar: So wird vom System METEO, das beim Kanadischen Wetterdienst englische Wetterberichte ins Französische überträgt, von einem vollautomatischen Einsatz in dem Sinne berichtet, daß vom System übersetzte Texte unmittelbar ohne menschliche Intervention über Nachrichtenagenturen weiterverbreitet werden [Kittredge, Goldberg 1986]. Lediglich die vom System zurückgewiesenen Texte (weniger als 10 % des Textmaterials) werden von Menschen übersetzt und danach in die Nachrichtenkanäle eingespeist.

Von der Ausnahme METEO abgesehen, ist bei allen anderen am Markt befindlichen Systemen normalerweise mit einem nicht unerheblichen Aufwand für Edition vor und/oder nach dem maschinellen Übersetzungsschritt zu rechnen.

3.3.3 Das Problem der Wissensbasierung von MÜ

Unabhängig davon, welches der gängigen Modelle der maschinellen Übersetzung einem System zugrundeliegt (direkt oder Transfer–basiert oder Interlingua–basiert), stellt sich allen Ansätzen zur maschinellen Übersetzung das Problem, daß für die Behandlung mehrdeutiger Strukturen der Ausgangssprache und die Entscheidung über Übersetzungsäquivalente in der Zielsprache ein mehr oder weniger tiefes Verständnis des Inhalts unumgänglich ist.

Man kann diesem Problem zwar auszuweichen versuchen, indem man sich benüht, Mehrdeutigkeiten unaufgelöst in die Zielsprache zu übertragen. Dies funktioniert in gewissen Grenzen und verlagert das Problem der Analyse und Disambiguierung auf den Leser.

Ein Beispiel: *„Die Soldaten feuerten auf die Frauen. Einige von ihnen fielen.“* im Kontrast zu *„ ... Einige von ihnen zielten über die Köpfe.“* bzw. *„ ... Einige von Ihnen weinten.“* Falls die Zielsprache bei Plural–Pronomen ebenfalls das Genus nicht ausweisen muß, läßt sich die Mehrdeutigkeit von „ihnen“ in den Beispielen übertragen. Dies trifft etwa auf Englisch zu. Für Französisch versagt diese Technik; hier ist es für die korrekte Übertragung notwendig zu wissen, auf welchen der beiden Personenkreise (Soldaten, Frauen) referiert wird.

Es ist sicher nicht zu pessimistisch, wenn postuliert wird, daß, um *beliebige Texte* adäquat übersetzen zu können, das oft als „Weltwissen“ apostrophierte gesamte menschliche Wissen verfügbar sein müßte.

3.3.4 FAHQTUT

Das langfristige Ziel der maschinellen Übersetzung wird mit dem Akronym 'FAHQTUT' zusammengefaßt, d.h. 'Fully–Automatic High Quality Translation of Unrestricted Texts'. Von diesem Langfristziel ist man immer noch weit entfernt.[1]

Mit einer ganzen Reihe von Einschränkungen wird versucht, die Komplexität des Problems so zu reduzieren, daß anwendbare Systeme realisiert werden können. Die wichtigsten dieser Ansätze zur Komplexitätsreduktion sind:

- Einschränkung des Themenbereichs,
- „Kontrolliertes Schreiben“,
- Verzicht auf volle Automatisierung,
- Verzicht auf 'high quality'.

Einschränkung des Themenbereichs

Wird der Themenbereich eingeschränkt, so sind viele Mehrdeutigkeiten — beispielsweise von Wörtern — nicht mehr gegeben; dies erleichtert die Analyse und hilft, die adäquate Übersetzung zu bestimmen.

Das folgende Beispiel unterstreicht diesen Zusammenhang. Der deutsche Satz *„Die Leitung wurde umgelegt“* findet je nach Fachgebiet die folgenden Entsprechungen:

- im Bereich Telekommunikation: *'The line was rerouted'*,
- im Bereich der Elektrotechnik: *'The cable was repositioned'*,
- und im umgangssprachlichen Wortschatz könnte eine Bedeutung die sein, die durch *'The management was killed'* übertragen wird.

Für Texte aus Gebieten mit einer genügend starken Einschränkung ist es aussichtsreich, das benötigte Gebietswissen zu modellieren und für den Analyse–Prozeß zu benutzen. Ein früher Vertreter dieses Ansatzes sind die von Schank und Schülern durchgeführten Experimente, auf der Basis von Conceptual Dependency–Repräsentationen maschinelle Übersetzungen von einfachen Zeitungstexten zu ermöglichen [Schank, Abelson 1977]. Der aktuell wichtigste Vertreter dieses Ansatzes ist das KBMT–Projekt an der Carnegie Mellon University (vgl. unten und S. 44).

[1] Und es ist zu Recht Gegenstand kontroverser Debatten, inwieweit dieses Ziel überhaupt jemals erreicht werden kann.

„Kontrolliertes Schreiben“

Eine andere Technik ist, den Wortschatz und die erlaubten syntaktischen Konstruktionen einzuschränken (sogenanntes ‘controlled writing’). Einige Firmen (z.B. Xerox, und Caterpillar) haben für ihre eigenen Publikationen Schreibvorschriften erlassen und verbieten darin schwierig zu verarbeitende Konstruktionen. Für die Luftfahrt hat ein internationaler Verband Standards entwickelt nicht nur für die sprachlichen Formen, sondern auch für die Inhaltsauswahl und -anordnung in Handbüchern (AECMA).

Eine solche kontrollierte Sprache kann nicht nur den Vorteil haben, automatisch syntaktisch und gegebenenfalls auch semantisch leichter zu analysieren zu sein, sondern führt — so wird zumindest berichtet — zu besser verständlichen Texten auch in der Quellsprache. Eine offene Frage ist, ob und wie sichergestellt werden kann, daß sich die Autoren auch tatsächlich an die Stilvorgaben halten. Erfahrungen bei natürlichsprachlichen Schnittstellen haben gezeigt, daß es Benutzern schwer fällt, syntaktische Einschränkungen zu kontrollieren und einzuhalten [Zoeppritz 1984].

Ein Projekt, das wissensbasierte Ansätze und kontrollierte Sprache für die automatische Übersetzung in großem Stil zu verbinden sucht, ist das auf den KBMT–Vorarbeiten (s.o.) aufsetzende Caterpillar–Projekt an der Carnegie Mellon University. Gebrauchsanleitungen und weitere Informationen zu Caterpillar–Produkten sollen in kontrolliertem Englisch erstellt und mit Hilfe des zu entwickelnden Systems in ein gutes Dutzend Zielsprachen automatisch übersetzt werden [Nyberg, Mitamura 1992].

Verzicht auf volle Automatisierung

Eine weitere Möglichkeit stellt der Verzicht auf die volle Automatisierung dar. Nimmt man Vor- und Nachbearbeitung oder beides in Kauf oder sieht vor, daß steuernd in den Übersetzungsvorgang eingegriffen werden kann (sogenannte ‘interaktive Übersetzung’), dann handelt es sich um maschinell unterstützte Übersetzung. Eine typische Arbeitsteilung hierbei ist die, daß das System standardmäßig leicht zu behandelnde Fälle automatisch korrekt überträgt und der Übersetzer dann bei kritischen Stellen ins Spiel kommt.

Verzicht auf ‘high quality’

Eine weitere Reduktion ist der Verzicht auf den Anspruch ‘high quality’. Rohübersetzungen, die Wesentliches vom Inhalt zu übertragen in der Lage sind, reichen z.B. dann aus, wenn es darum geht, ein grobes Bild von bestimmten Inhalten zu bekommen und nur im Fall großen Interesses dann eine zeitaufwendige und teure Vollübersetzung zu veranlassen.

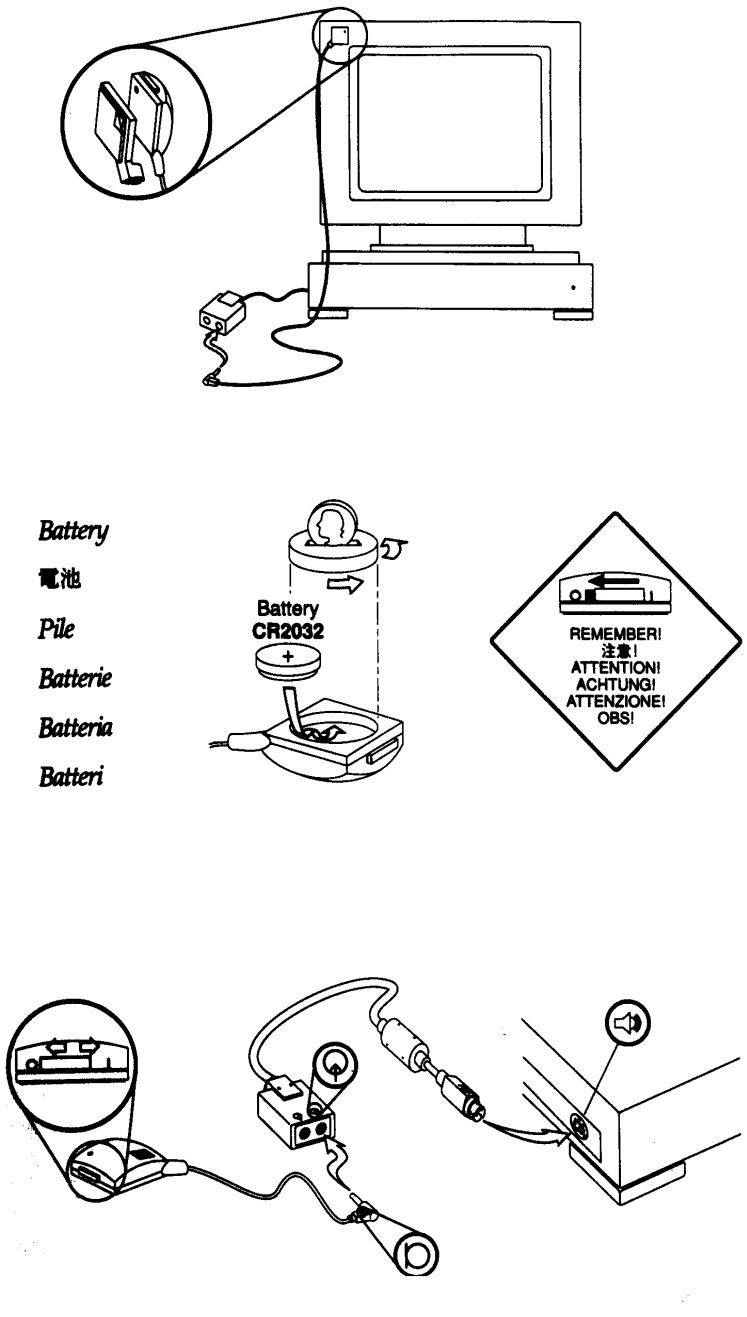

Abbildung 3.3: Eine Installationsanleitung (fast) ohne Worte

3.3.5 Piktogramme statt Mehrsprachigkeit?

Die radikalste Lösung für das Problem der Mehrsprachigkeit wäre die, ganz auf einzelsprachliche Texte zu verzichten und zu weitestgehend visuell gestalteten technischen Dokumentationen mit Bildern und Piktogramm–Darstellungen überzugehen. Ein

THEMEN DER ZEIT

„Du sollst dir kein Bild machen!" verlangt die Bibel. Über Piktogramme sagt sie nichts. Zum Jahreswechsel, der ja immer auch ein Zeiten- und ein Zeichenwechsel ist, ein Beitrag der jungen Kölner Gestalterin Juli Gudehus

DIE SCHÖPFUNG

Abbildung 3.4: Die GENESIS als Piktogramm

Beispiel hierfür — eine Installationsanleitung für das Mikrofon einer SUN — zeigt Abb. 3.3.

Piktogramme, Zeichnungen und andere visuelle Mittel haben den Vorteil, sprachunabhängig und damit inhärent multilingual zu sein. Wird Text verwendet, ist dagegen für jede weitere Zielsprache eines technischen Dokuments zusätzlicher Aufwand

VERGLEICH	
piktogrammbasierte Informationspräsentation	*mehrsprachige Dokumente*
– inhärent multilingual	– Aufwand für jede weitere Sprache
– Problem der Ausdrucksmächtigkeit temporale Beziehung? kausale Beziehungen?	– prinzipiell uneingeschränkt aber: Schwächen bei räumlichen Beziehungen

Abbildung 3.5: Vergleich: Piktogramme vs. Texte

(für Erstellung, Übersetzung, Verwaltung usw.) erforderlich. Räumliche Beziehungen lassen sich zudem durch visuelle Techniken (wie Zeichnungen oder Fotos) einfacher verdeutlichen, sie mit sprachlichen Mitteln zu kommunizieren, ist oft nicht in der gleichen Einfachheit und Klarheit möglich.[2]

Einige Einwände gegen eine rein visuell basierte, international verwendbare Darstellungsform:

- Ist sichergestellt, daß Piktogramm–Auffassung und Piktogramm–Verstehen (also das visuelle Vokabular, wenn man so will) universal sind und nicht auch von Kultur und Erfahrungshintergrund abhängen?
- Es ist unklar, inwieweit komplexe Beziehungen zwischen Propositionen (Kausalität, Vorbedingung, temporale Bezüge usw.) rein visuell adäquat ausgedrückt werden können.

Ein illustratives Beispiel in diesem Zusammenhang, das hier ebenfalls ausschnittweise wiedergegeben wird, ist der in der „ZEIT“ zum Jahreswechsel 1992/93 erschienene Versuch einer „Übersetzung“ des ersten Teils der Genesis in „Piktogramm–Sprache“ (vgl. Abb. 3.4).

3.4 Der Ansatz in TECHDOC

Im Vergleich zu den Ansätzen zur maschinellen Übersetzung kann der in dieser Arbeit entwickelte und im Projekt TECHDOC prototypisch realisierte Ansatz so charakterisiert werden, daß versucht wird, gänzlich auf die Analyse natürlichsprachlicher Quellen zu verzichten und alle damit verbundenen Probleme zu vermeiden. Es wird stattdessen versucht, von Einzelsprachen unabhängige Inhaltsrepräsentationen (verbunden mit Textstruktur–Information) aufzubauen und aus diesen dann Texte in verschiedenen Sprachen zu generieren.

[2]Mit Möglichkeiten zur integrierten Nutzung von visuellen und textuellen Mitteln in Benutzerschnittstellen beschäftigt sich [Strothotte 1989].

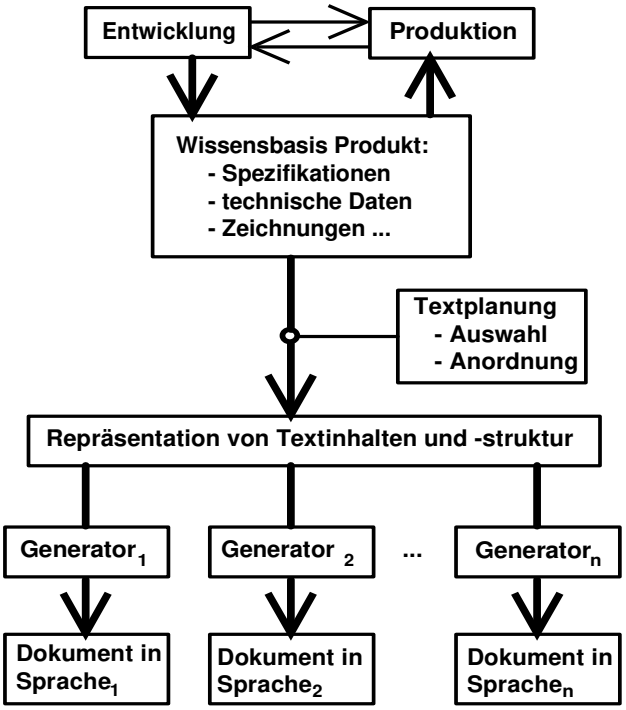

Abbildung 3.6: TECHDOC–Szenario: multilinguale Generierung technischer Dokumente

Der Aufwand beim Aufbau solcher Wissensbasen ist, wie auch die Erfahrungen bei der sog. wissensbasierten maschinellen Übersetzung zeigen, nicht gering; er ist allerdings auch nicht größer als bei Ansätzen zu wissensbasierter MÜ in realistischen Anwendungsfeldern. Der Aufwand kann voraussichtlich zukünftig geringer werden, weil im Umfeld von Design, Planung, Produktion, Qualitätssicherung und anderen Ingenieur–Aktivitäten umfangreiche, strukturierte Informationsbestände (z.B. Produktdatenmodelle, Prozeßmodell) im Aufbau befindlich sind und (dies bleibt allerdings noch Gegenstand weiterer Untersuchungen) ggf. für die Bedürfnisse von Systemen wie TECHDOC nutzbar gemacht werden können.

Umgekehrt kann davon ausgegangen werden, daß die für TECHDOC aufgebauten Wissensbasen auch für andere Zwecke nutzbar gemacht werden können. Ein erstes Beispiel dafür ist TECHDOC–I, ein System, das interaktiv Hinweise zu Wartungsarbeiten geben kann und dabei den Kenntnisstand des Benutzers mit einbezieht [Peter 1992]. Dieses System operiert auf der Wissensbasis, die auch für die Produk-

VERGLEICH

wissensbasierte MÜ	*Generierung aus Wissensbasen*
– Parsing	—
– Disambiguierung	—
– Aufbau großer Wissensbasen erforderlich	– Aufbau großer Wissensbasen erforderlich
– Inhalt aus Quelltext	– Inhalt in Wissensbasis

Abbildung 3.7: Vergleich: Generierung vs. wissensbasierte MÜ

tion der TECHDOC–Texte verwendet wird. Exemplarisch wurde auch gezeigt (vgl. 7.5), wie die Wissensbasis von TECHDOC zur Simulation verwendet werden kann.

3.4.1 Diskussion

Die Ergebnisse bei einem Ansatz wie in TECHDOC müssen nicht die volle Sprachfähigkeit und Variationsmöglichkeit eines menschlichen Autors erreichen, um nützlich zu sein. Gegebenenfalls ist es sogar ein Vorteil, wenn eine zu große „Kreativität“ in technischen Texten vermieden wird.

VERGLEICH

MÜ mit kontrollierter Eingabe	*Generierung aus Wissensbasen*
– Wie können Vorgaben „erzwungen“ werden?	—
– Parsingproblem (leicht gemildert)	– Problem Wissensakquisition
– ein Quelltext muß vorliegen	– Texte verfügbar, sobald WR–Struktur aufgebaut
– Modell des Anwendungsgebietes?	– Modell des Anwendungsgebietes zentral

Abbildung 3.8: Vergleich: Generierung vs. MÜ mit kontrollierter Eingabe

Viel mehr ins Gewicht dürfte fallen, daß der Entstehungsprozeß der technischen Dokumente viel stärker bereits in die Produktentwicklung hinein verlagert und mit dieser besser verzahnt werden kann. Es ist vorstellbar, daß mit Hilfe eines geeig-

Vorteile

- Aktualisierung und Versionskontrolle
- auf Bedarf verfügbar („papierarmes Büro“)
- Verteilung vereinfacht
- an Zielgruppe anpaßbar
- an Situation anpaßbar
  (z.B. Integration mit Sensorikinformation)
- klare Schnittstelle für weitere Sprachen

Abbildung 3.9: Mehrsprachige Generierung aus Wissensbasen

net ausgebauten Autorenwerkzeugs (vgl. 8.1) die an der Produktentwicklung und -gestaltung beteiligten Ingenieure selbst die Funktion der bisherigen technischen Redakteure übernehmen.

Von Vorteil ist weiter, den Inhalt der Dokumente modelliert und abgelegt zu haben, so daß bei kleinen inhaltlichen Veränderungen auf die bereits abgelegten Inhalts- und Strukturinformationen zurückgegriffen werden kann und ein entsprechend variierter Text sehr leicht und in allen unterstützten Sprachen zur gleichen Zeit produziert werden kann.

Eine sprachunabhängige Repräsentation der Dokumentinhalte zu besitzen und die einzelsprachlichen Ausprägungen lediglich als Sichten darauf zu sehen, erleichtert auch die Versionskontrolle. Allerdings sollte nicht vergessen werden, daß gelegentlich (z.B. aufgrund unterschiedlicher Zulassungsvorschriften oder aus Gründen des Wettbewerbs) für einzelne Länder oder ganze Ländergruppen Unterschiede bei den Produkten vorliegen, die dann für die für diese Länder geltenden Sprachversionen auch ihren Niederschlag finden müssen. Mit anderen Worten: Die gemeinsame, sprachunabhängige Repräsentation des sozusagen „virtuellen“ Dokuments muß solche Unterschiede, die sich dann für eine oder mehrere der Sprachen auch in den Texten aktualisieren, mit berücksichtigen können.

Liegen mit den repräsentierten Textinhalten und den repräsentierten Strukturinformationen die Grundlagen für die Entscheidung über das Layout vor, so kann das tatsächliche gewählte Layout sehr flexibel unterschiedlichsten Ausgabe–Medien oder geänderten Anforderungen an die Gestaltung angepaßt werden. Für Systeme ohne diese Information entsteht für solch „oberflächliche“ Aufgaben ein erheblicher Zusatzaufwand. So berichtet ein Anwender des maschinellen Übersetzungsystems METAL über einen extrem hohen Anteil von De- und Reformatierungsarbeiten und zeigt

auf, welch hoher Zeitanteil vom gesamten Übersetzungsaufwand in diese Aufgabe investiert werden muß [Paillet 1992].

3.5 Generierung und Multilingualität

Warum sollte man sich mit multilingualer Generierung beschäftigen? Zu möglichen Motiven zählen:

- praktische Anforderungen,
- linguistische Interessen,
- Fragen des Generatorentwurfs.

Praktische Anforderungen wurden bereits oben ausreichend erläutert.

Aus linguistischer Sicht können Arbeiten in multilingualer Generierung eine interessante Übung in kontrastiver Linguistik sein. Die Perspektive der Generierung unterscheidet sich deutlich von anderen Zugängen. Die zentrale Frage ist hier: ein Inhalt sei gegeben — wie drücken die Sprachen, die betrachtet werden, diesen Inhalt aus?

Für den Entwurf von Generierungssystemen geben Schritte in Richtung Multilingualität immer einen Anstoß, deutlicher zwischen sprachunabhängiger „Generierungsmaschinerie“ und sprachspezifischen Daten zu unterscheiden.

3.5.1 Multilingualität in Generierungssystemen

Multilingualität war nie der zentrale Diskussionspunkt im Bereich natürlichsprachlicher Generierung, spielte aber bereits bei ganz frühen Generatoren eine Rolle. Goldmans BABEL–Programm konnte nicht nur Ausgaben in Englisch produzieren, sondern in einer ganzen Zahl von Sprachen, z.B. Spanisch, Mandarin, Deutsch. BABEL ging dabei von einer semantischen Repräsentation aus (den sogenannten ‘Conceptual Dependency’–Graphen), die entworfen worden waren mit dem Anspruch, so universal wie möglich zu sein [Goldman 1975].

Arbeiten im Hinblick auf Multilingualität aus der jüngsten Zeit können grob in die folgenden Kategorien klassifiziert werden:

- Experimente, einen taktischen Generator auf eine andere Sprache zu übertragen,
- Experimente, Textgeneratoren ihre Ausgabe in verschiedenen Sprachen produzieren zu lassen,
- Bemühungen, Multilingualität bereits in das Design von Anfang an einzubeziehen.

Zur dritten Kategorie kann man u.a. das Projekt POLYGLOSS aus Stuttgart [Emele, Zajac 1990] und das multilinguale PENMAN–Projekt in Sydney [Bateman et al. 1991] zählen.

Experimente zur Anpassung eines taktischen Generators an eine andere Sprache sind oft mehr aus pädagogischen Gründen ausgeführt worden als mit dem Ziel, zu tatsächlich einsetzbaren Versionen zu kommen. Über diese Experimente wird nicht oft berichtet. Die Diskussion von Erfahrungen bei einem Versuch, das für Englisch entwickelte System MUMBLE auf Deutsch zu übertragen, durch [Jacob, Maier 1988] ist eine eher seltene Ausnahme.

In die zweite Kategorie gehören Reportgeneratoren, zu denen eine Version in einer anderen Sprache entstanden ist. FRANA [Contant 1986] ist z.B. die französische Version von Kukichs ANA–Programm [Kukich 1983]. ANA lieferte Börsenberichte aus Dow-Jones Daten. Sowohl SEMTEX als auch GEOTEX, beides Textgeneratoren, die auf SEMSYN aufsetzen [Rösner 1986], gibt es mit englischer Version. SEMTEX–E produziert englische Versionen von Berichten über Arbeitsmarktentwicklungen [Rambow 1988]. Bei GEOTEX wurden die deutschen und englischen Versionen in ein Programm integriert, das simultan geometrische Konstruktionen in beiden Sprachen beschreiben konnte [Kehl 1987].

3.6 Andere Arbeiten zur Generierung von Instruktionstexten

Generierung hat sich erst in den letzten zehn bis zwölf Jahren als Gebiet innerhalb der Computerlinguistik und sprachorientierten KI konstituiert. Die älteren Arbeiten sind oft eher grundlagenorientiert als anwendungsbezogen. Anwendungen von Generierung, die früh angegangen wurden, waren die innerhalb von Dialogsystemen oder im Rahmen maschineller Übersetzung.

Arbeiten, die ausschließlich die Generierung von Instruktionstexten aus einer Wissensbasis behandeln, sind erst jüngeren Datums:

- Im COMET–Projekt stehen, wie in den Projekten WIP und dem Nachfolgeprojekt PPP [Wahlster et al. 1993], die Fragen der Integration und wechselseitigen Bezugnahme von Text und Graphik im Mittelpunkt. COMET arbeitet über einer Wissensbasis zur Bedienung eines Funkgeräts [Feiner, McKeown 1990].

- Im WIP–Projekt wurden, von Präsentationsparametern gesteuert, unterschiedliche Anleitungen für die Bedienung einer Espresso–Maschine und für die Wartung und Nutzung eines Rasenmähers produziert. WIP hat als einen Präsentationsparameter auch die Sprache. Das System kann zwischen Deutsch und Englisch wechseln.

- Anleitungen zur Benutzung schnurloser Telefone hat das IMAGENE–Projekt [Linden et al. 1992] als Gegenstandsbereich. Die Texte wurden, insbesondere

im Hinblick auf die Diskursstruktur und die verwendeten RST–Relationen (vgl. 4.2) hin analysiert und mit einem systemischen Ansatz reproduziert.

- Das System EPICURE [Dale 1990] generiert Kochrezepte. Dale hat sich dabei vor allem auf die Frage nach den geeigneten referierenden Nominalphrasen konzentriert. Von der Problematik der Wissensrepräsentation her ging es bei der Arbeit auch um die Fragestellung, wie sich das Entstehen neuer Substanzen (z.B. Teig durch Vermischung der Zutaten) oder die Zustandsänderung von festen und zählbaren Größen in substanzartige modellieren lassen.

- Im IDAS–Projekt (Intelligent Document Advisory System, [Levine et al. 1991]; s.a. 6.4.2) werden zur Unterstützung der Benutzer sogenannter 'automatischer Testeinrichtungen' (ATEs) Antworten auf (über Menü eingegebene) Fragen generiert, wobei auf die Situation und ein Benutzermodell zurückgegriffen wird, um den Detailliertheitsgrad zu steuern. In IDAS wird Generierung mit dem Hypertext–Paradigma verbunden. Gibt es von Elementen einer Antwort Bezüge zu anderen Einheiten der Wissensbasis, so können diese durch Aktivierung entsprechender Hypertext–Links ebenfalls erreicht werden.

- Instruktionstexte für Caterpillar–Maschinen sind der Gegenstandsbereich eines großen Projekts der CMU und der Carnegie–Group (vgl. S. 35). Dieses Projekt ist der hauptsächliche Vertreter der sogenannten „wissensbasierten maschinellen Übersetzung“ (Knowledge–Based Machine Translation oder KBMT). Mit Hilfe einer Modellierung des Sachgebiets und einer entsprechend ausgefeilten englischen Grammatik wird aus Quelltexten, die in kontrolliertem Englisch abgefaßt sind, eine semantische Zwischenstruktur erzeugt, aus der dann in fast einem Dutzend Zielsprachen Instruktionen zur Wartung und Handhabung von Baumaschinen automatisch erzeugt werden sollen. Der Umfang der jährlich zu bearbeitenden Seiten ist enorm [Nyberg, Mitamura 1992].

- Im Rahmen des von der EU seit kurzem geförderten GIST–Projekts (Generation of Instructional Texts) wird untersucht, wie aus einer Inhaltsrepräsentation Instruktionstexte für das Ausfüllen von Formularen (z.B. für Finanz- oder Sozialamt) und andere Informationen für den Umgang mit Behörden in mehreren Zielsprachen (u.a. Italienisch, Deutsch, Englisch) erzeugt werden können.