

## 4.2 Perspektiven der maschinellen Übersetzung in der Fachinformation

Nach der Erörterung spezifischer Fragen der maschinellen bzw. computergestützten Übersetzung kehren wir hier zur allgemeineren Frage der Mehrsprachigkeit von Fachinformation zurück. Dabei sollen auch Fragen einer möglichen, das Problem der Mehrsprachigkeit einbeziehenden Fachinformationspolitik in der Bundesrepublik Deutschland und in der Europäischen Gemeinschaft gestellt und kurz begründet werden. Die angesprochenen Themen erheben keinesfalls den Anspruch auf Vollständigkeit, auch bedeuten sie keine Kritik an den laufenden Aktivitäten der Bundesregierung bzw. der Europäischen Gemeinschaft, sondern gewichten diese unter dem allgemeinen Aspekt der Überwindung von Sprachbarrieren.

### 4.2.1. Die deutsche Fachinformation muss mehrsprachig werden.

In der Fachinformation muss auf Dauer zwei Kriterien Rechnung getragen werden:

- International ist **Englisch** - bis auf wenige Ausnahmen - die Sprache wissenschaftlicher Kommunikation. Nicht in Englisch publizierte Fachliteratur ist nur national - wenn überhaupt - distribuierbar.
- National ist die **Muttersprache** (hier: Deutsch) das natürliche Kommunikationsinstrument auch in der Fachkommunikation. Es wäre ein (zu) großes Risiko, wegen der Notwendigkeit der internationalen Kommunikation allein auf Englisch auszuweichen. Die Akzeptanz von Forschung und Entwicklung im eigenen Land, der Transfer wissenschaftlicher Ergebnisse in die Praxis würde zusätzlich blockiert.

Daraus folgt: in allen Fällen, in denen sowohl nationale als auch internationale Zielgruppen erreicht werden sollen, müssten die Daten mehrsprachig - im Mindestfall Deutsch-Englisch (in Europa nach aller Erfahrung eher Deutsch-Englisch-Französisch) - vorliegen.

Dieses Ziel kann naturgemäß nicht umfassend realisiert werden. Folgende Elemente/Stufen erscheinen jedoch sinnvoll:

- Es sollte keine international angebotene bibliographische Datenbank (mehr) geben, deren "deutscher" Anteil (Titel, Abstracts) nicht in Deutsch und Englisch verfügbar ist. (Für Faktenbanken ist die Problematik ebenfalls gegeben, erscheint jedoch vergleichbar einfacher lösbar).

Begründung: Vordergründig ist dadurch international eine Umsatzsteigerung zu erwarten. Wichtiger noch ist aber, dass die Chance, am internationalen Informationsmarkt überhaupt "bemerkt" zu werden, für die deutsche Wissenschaft und Forschung steigt.

- Internationale Datenbanken sollten zumindest terminologisch mit deutschen Begriffen (Deskriptoren) recherchiert werden können.

Begründung: Damit wird zumindest partiell die Möglichkeit auch für den Nicht-Spezialisten eröffnet, fremdsprachige (i.W. englische) Daten über muttersprachliche Begriffe zu erreichen. Die sprachwissenschaftliche Forschung hat längst gezeigt, dass das passive Sprachverstehen i.d.R. weit umfassender ist als der aktive Sprachgebrauch. Diesem Umstand trägt diese Komponente besonders Rechnung.

- Es müssen die Anstrengungen bedeutend verstärkt werden, Volltexte in Deutsch und Englisch verfügbar zu machen. Dies kann durch den Ausbau bestehender Einrichtungen (z.B. an der TIB Hannover), die Förderung der Etablierung entsprechender Abteilungen in den Informationszentren und ggf. durch den Aufbau von Service-Zentren geschehen.

Begründung: Mit dem Aufbau von Volltext-Datenbanken ist der erste (technische) Schritt getan. Aber - um ein Beispiel zu geben -: was nutzt eine deutschsprachige Patentdatenbank, wenn ihr Inhalt international nicht "verstehbar" ist?

Verfahren: In einigen Bereichen - so der Patent- und Normeninformation - müsste umgehend mit der Entwicklung einer englischen Variante begonnen werden. Für Proceedings internationaler Kongresse gilt ähnliches. Hierzu sind zunächst Modelle zu entwickeln und Evaluierungen (d.h. Effizienzbewertungen) durchzuführen.

#### 4.2.2 Die thematische Inhaltserschließung muss verbessert werden.

Begründung: Die Klassifikations- und Thesaurusforschung ist lange Zeit (und eigentlich bis heute) bezüglich ihres Potentials nicht ausreichend gewürdigt worden. Der Grund ist z.T. darin zu sehen, dass sie sich vielfach präskriptiv und nicht deskriptiv verhalten hat, zudem wurde einer Normierung häufig genug mit mehr oder minder ideologischen Argumenten das Wort geredet. Die Vielfalt von Klassifikationssystemen trägt auch nicht gerade zur Verbesserung der Situation bei, ganz abgesehen davon, dass ihre Nützlichkeit angesichts der Problematik, dass der Nutzer zunächst eine künstliche Systematik (Sprache) erlernen muss (Aufwand), häufig in keinem Verhältnis zu ihrem Ertrag (bessere Ergebnisse) stand.

Dennoch sind derartige Instrumente, sinnvoll angewandt (Beispiel: IPC), auch für die Zukunft unerlässlich.

Verfahren: Was häufig fehlt, sind systematische Prüfungen/Untersuchungen bezüglich der Konsistenz, aber auch Verfahren, die auf automatischem Wege (zudem kostengünstig) Textklassifikationen (d.h. Clusterbildungen) zur Ermittlung der Identität bzw. Ähnlichkeit von Texten ermöglichen. Der Ausbau bzw. die Entwicklung sprachlich motivierter Klassifikationssysteme, Untersuchungen zu Textkondensierungen im Verbund mit intellektuellen Aufbereitungen bieten eine gute Chance, zum Nutzen der Experten brauchbar-praktische Lösungen zu realisieren.

Dass die Retrieval- und Datenbanksoftware verstärkt Hilfestellungen (Einbeziehung mehrsprachiger Thesauri; Clusterverfahren, Ranking) anbieten muss, ist eine Konsequenz dieser Überlegungen.

#### 4.2.3 Fachinformation muss zu einer "Europäischen Angelegenheit" (zumindest der EG) werden.

Die Fachinformationspolitik der EG-Mitgliedstaaten, insbesondere Frankreichs und der Bundesrepublik Deutschland, ist weitgehend national bestimmt. Es gibt zwar Ansätze, im europäischen Rahmen zusammenzuarbeiten, dies beschränkt sich aber auf wenige systematische Ansätze (etwa ist hier das Projekt EUROTRA zu nennen). Eine große europäische Linie einer

gemeinsamen Fachinformationspolitik ist weder informationstechnologisch noch inhaltlich zu erkennen. Wenn dennoch (z.B. gerade in Frankreich und in der Bundesrepublik Deutschland) relative Erfolge zu verzeichnen sind: wie viel stabiler würde bei konsequenter Verfolgung europäischer Interessen ein gemeinsames Konzept sein!

Natürlich muss man sich z.B. dazu durchringen, "preußische" mit "italienischer" Mentalität zu verbinden - keine leichte Aufgabe, auch angesichts der Sprachbarrieren. Aber eines dürfte sicher sein: auf die Dauer lassen sich in einem sich vereinigenden Europa diese Probleme sowieso nicht ausklammern, umgekehrt kann eine erfolgreiche, d.h. beharrlich-ausdauernd betriebene, die fundamentalen Einzelinteressen berücksichtigende und dennoch nach innen wie außen abgestimmte europäische Fachinformation gerade die Integration weiter fördern. Dabei sollte der Zusammenschluss nicht gleich nach außen (Japan, USA ...) eine neue "gemeinsame" Barriere errichten, sondern dadurch auch die internationale Kooperation fördern, dass ein verlässlicher Partner (statt einer Vielzahl heterogener Interessen) bei der internationalen Vernetzung von Fachinformationssystemen gefunden wird.

Dass sich hier das Sprachbarrierenproblem (fast) potenziert, da unterschiedliche Interessen harmonisiert, dass Alternativen verfolgt und auch unterschiedliche "Level" berücksichtigt werden müssen, versteht sich. Dies sollte im Innern wie im Äußern aber ein Ansporn sein.

Im einzelnen ist hier zu wünschen:

- Die Einrichtung eines aus kompetenten Partnern sich konstituierenden Gremiums, das in Zusammenarbeit mit der EG-Kommission ein europäisches Fachinformationsprogramm erarbeitet.
- Die Realisierung von mehrsprachigen Modellprojekten (z.B. im Bereich Patente, Industrie-Normen).
- Die Entwicklung einer europäischen Aus- und Fortbildungskonzeption, die den Austausch von Studierenden und Wissenschaftlern fördert.
- Die stufenweise Vernetzung bestehender Informationsdienste unter Kompatibilisierung der Daten und Integration multilingualer Aspekte (evtl. mit Englisch als Switching Language).

#### 4.2.4. Ausblick

Schaut man in eine ferne Zukunft, so scheint sicher zu sein, dass sich die professionelle Übersetzung im Normalfall der MT bedienen wird. Dies wird mehr oder weniger von den "geeigneten" Sprachenpaaren abhängen, von der Vollständigkeit der Maschinenwörterbücher, von der "richtigen" technischen Systemumgebung und nicht zuletzt auch von den Kosten der Systemnutzung.

Daneben werden auch CAT-Systeme als Vokabel- und Terminologiehilfen - in Textverarbeitungssysteme integriert oder an sie angeschlossen - eine wichtige Rolle spielen, insb. als Unterstützung bei der Erstellung eines fremdsprachigen Textes durch den Autor selbst oder um einen über Telekommunikation (TELETEX) übermittelten Text zu verstehen. Im Büro wird CAT hauptsächlich mit bilingualen Wörterbüchern auf CD-ROM oder Festplatte als Stil- oder Grammatikhilfe benutzt werden.

Im Gegensatz z.B. zu Hutchins (1986, S. 331-334) muss man nicht die Notwendigkeit eines besonderen Übersetzerarbeitsplatzes sehen, aber es wird besondere Übersetzungshilfe-Funktionen geben, die - in leistungsfähige Autoren-Arbeitsplätze integriert - die Nutzung von Fortschritten im Bereich von Computerlexika und -thesauri möglich machen.

Wie werden im Hinblick auf - im weitesten Sinne - computergestützte technische Übersetzungen die Arbeitsbedingungen der professionellen Übersetzer aussehen? Man kann MT-Systeme als besondere "Expertensysteme" ansehen, wobei angemerkt werden muss, dass sie so gut wie nie die Qualität guter spezialisierter Übersetzer erreichen werden. Um jedoch solche Übersetzungs-Expertensysteme zu entwickeln, zu verbessern und zu unterhalten, werden "Sprachtechnologien" (linguistic knowledge engineers) gebraucht, die in der Lage sind, mit diesen Systemen umzugehen. Was die Komplexität der natürlichen Sprache angeht, so wird es zur Zusammenarbeit zwischen System und Übersetzer kommen. Und es besteht die gute Chance, dass - falls Übersetzungen billiger und schneller werden - die Nachfrage nach ihnen steigt und letzten Endes die Übersetzer - unter veränderten Bedingungen - nach wie vor eine wichtige Rolle spielen werden.