

In: Heinz-Dirk Luckhardt, Harald H. Zimmermann (1991): Computergestützte und maschinelle Übersetzung. Dudweiler: AQ-Verlag, 5-8

Harald H. Zimmermann

Computergestützte Übersetzung als ein Beitrag zur Überwindung von Sprachbarrieren in der Fachinformation

Die maschinelle oder computergestützte Übersetzung steht im Blickfeld einer Reihe von Disziplinen. Dementsprechend gibt es eine Vielzahl wissenschaftlicher und auch praxisorientierter Behandlungen des in der vorliegenden Arbeit angesprochenen Themas. Die meisten Übersichten sind vorwiegend unter sprachwissenschaftlichen Aspekten erstellt worden. Zu nennen sind hier v.a. die Arbeiten von Bar-Hillel (1960), Bruderer (1978), Slocum (1984), Hutchins (1986). Unter übersetzungswissenschaftlichen Aspekten wird das Thema in Blatt et al. (1985) behandelt, während Zimmermann (1980, 1987) die Anwendungsproblematik in den Vordergrund stellt.

Den hier vorzustellenden anwendungsorientierten Untersuchungen liegt ein Übersetzungssystem zugrunde, das in langjähriger experimenteller Forschung am Sonderforschungsbereich Elektronische Sprachforschung an der Universität des Saarlandes realisiert werden konnte. Vorphasen (Russisch-Deutsch) reichen in die 60-er Jahre zurück. Zu diesem "Saarbrücker Übersetzungssystem" (SUSY) kann auf relevante Veröffentlichungen verwiesen werden (v.a. Maas 1978 und Luckhardt 1988c sowie die Einführung in Blatt et al. 1985, S. 141-251); auf Komponenten, die für den vorliegenden Zusammenhang wichtig erscheinen, geht Luckhardt 1987 ein.

Erklärtes Ziel des Forschungsprojekts MARIS war es (vgl. Kap. 2.1), bestehende prototypische Entwicklungen zur maschinellen Übersetzung in größerem Stil in eine praktische Anwendungsumgebung hineinzustellen. Für diese Anwendung wurde ein Bereich ausgewählt, der im Mittelpunkt eines langjährigen Förderprogramms der Bundesregierung stand: die Fachinformationsvermittlung über elektronische Datenbanken (vgl. FACHINFORMATION 1985-88). Im Rahmen dieses vom Bundesministerium für Forschung und Technologie geförderten Forschungsprojekts (Förderkennzeichen 1013209/2) sollten "multilinguale Anwendungen" des Systems SUSY im Bereich "referenzorientierter Informations-Systeme" erfolgen: daraus erklärt sich nicht nur das Akronym, sondern grob auch die Zielsetzung: Orientierung auf Fachinformationsbanken, speziell Literaturdatenbanken. Dies soll im folgenden kurz beschrieben werden.

Angesichts der Vielfalt der natürlichen Sprachen steht die Wissenschaft vor einem Kommunikationsproblem: Wissenschaftliche Erkenntnisse sind zunehmend international von Interesse, es ist aber unmöglich, wissenschaftliche Ergebnisse in allen natürlichen Sprachen zu distribuieren. Lösungen unterschiedlicher Art bieten sich an. Hierzu rechnen die Einführung von Klassifikationssystemen (Universal Decimal Classification UDC, Internationale Patentklassifikation u.a.), die Entwicklung von Kunstsprachen (wie Esperanto) und nicht zuletzt die Nutzung einer fachspezifischen Formelsprache (wie in der Mathematik und der Chemie). Im 20. Jahrhundert hat sich zudem zunehmend das Englische als wissenschaftliche Kommunikationssprache herauskristallisiert. Es ist heute weitgehend ein Faktum, dass Veröffentlichungen international nur noch eine Chance haben, beachtet zu werden, wenn sie in Englisch erfolgen. Demgegenüber ist im nationalen Bereich auch in der wissenschaftlichen Kommunikation die Muttersprache üblich, so dass heute jeder Experte seine Fachterminologie - soweit er nicht Englisch als Muttersprache hat - in zwei natürlichen Sprachen beherrschen muss.

Es stellt sich die Frage, inwieweit eine übernational ausgerichtete (d.h. englischsprachige) Fachkommunikation nicht den "nationalen" Transfer des Fachwissens in die praktische Anwendung behindert oder gar verhindert. Die Problematik hat aber auch einen ökonomischen Aspekt: Eine Datenbank - um auf den hier näher interessierenden Bereich einzugehen - lässt sich international besser vermarkten, wenn sie englischsprachig ist. Im nationalen Bereich ist es eher umgekehrt. Die Konsequenz kann also nur heißen: **es muss Ziel einer Fachinformationspolitik sein, Fachinformation in den wesentlichen Komponenten mehrsprachig (d.h. für die Bundesrepublik Deutschland: zumindest Deutsch und Englisch) zu gestalten.**

Der Bereich, in dem dies noch am ehesten möglich erscheint, ohne die Kosten wesentlich zu erhöhen, ist die referenzorientierte Fachinformation. Hierunter sind v.a. bibliographische Datenbanken (Literaturdatenbanken) zu verstehen, d.h. Informationsbanken, die an textuellen (für die Übersetzung relevanten) Komponenten die Titel und ggf. Abstracts (Kurzfassungen) von Veröffentlichungen, nicht aber die vollen Texte der Veröffentlichungen selbst enthalten. Zu den Elementen, die für einen natürlichsprachigen Zugang mehrsprachig aufbereitet sein sollten, sind auch die Stich- und Schlagwortsysteme bzw. die Thesauri zu rechnen. Schließlich erscheint es wünschenswert, dass auch die Interaktionssprache der Retrievalsysteme selbst in verschiedenen natürlichen Sprachen zugänglich ist.

Es ist klar, dass eine Beschränkung von Übersetzungsleistungen auf Titel und Abstracts nur eine Teillösung darstellen kann. Dies ist in erster Linie - zumindest gilt dies für das hier beschriebene MARIS-Projekt - auf finanzielle Restriktionen zurückzuführen, d.h. eine Kostenfrage. Dennoch erscheint schon viel für die nationale Fachinformation gewonnen, wenn für die wichtigsten referentiellen Informationsbanken die textuellen Daten in Deutsch, Englisch und ggf. (d.h. wenn von diesen Sprachen unterschieden) in der Originalsprache verfügbar und zugänglich sind.

In diesem Zusammenhang stellt sich zunächst allgemein die Frage, ob dieses Ziel notwendig mit Hilfe maschineller bzw. maschinengestützter Verfahren erreicht werden muss. Betrachtet man nämlich den Stand der Entwicklungen in diesem Bereich, so ist zunächst ein Mangel an geeigneten Werkzeugen festzustellen. Auch das dem Forschungsprojekt MARIS zugrunde liegende System SUSY macht hier keine Ausnahme, sieht man einmal von der relativ hochstehenden Leistungsfähigkeit der deutschen Indexierungssysteme ab (vgl. dazu Womser-Hacker 1989).

Eine erste - allgemeine - Frage ist daher zunächst: Ist es sinnvoll, den Computer für Übersetzungszwecke heranzuziehen? Wenn man zur Zeit des Projektbeginns (1985) noch prinzipielle Zweifel hatte, so sind diese durch die Entwicklungen der letzten Zeit zumindest reduziert worden. Weltweit werden inzwischen verstärkt Anstrengungen unternommen, den Übersetzungsprozess maschinell zu unterstützen. Auch in der Bundesrepublik kommen bereits am Markt verfügbare Systeme zum Einsatz. Zu nennen sind hier v.a. LOGOS, METAL und SYSTRAN. Eine entsprechende, im Zusammenhang mit MARIS und EUROTRA am IAI veranstaltete Tagung sowie 2 MARIS/STS-Präsentationen in Saarbrücken (vgl. IAI-MT 86) haben die prinzipiellen Leistungsmöglichkeiten der Systeme zur maschinellen und maschinengestützten Übersetzung verdeutlichen können. Mit der zu erwartenden weiteren positiven Entwicklung des Preis-Leistungsverhältnisses im Hardware-Bereich werden die prinzipiellen Anwendungsmöglichkeiten noch zunehmen. Es steht inzwischen also außer Frage, dass sich Übersetzungsabteilungen und -büros dieser Instrumente zunehmend bedienen werden, einmal ganz abgesehen von den Möglichkeiten,

die bereits eine um Glossarzugriffe erweiterte Textverarbeitung im PC-Bereich bietet bzw. bringen wird (vgl. dazu auch Zimmermann 1990).

Eine Frage, die im Rahmen von MARIS beantwortet werden sollte, ist dagegen die der Machbarkeit des Konzepts im vorgegebenen Rahmen. Eine erste Teilfrage in diesem Zusammenhang könnte lauten: Ist es sinnvoll, für die gegebenen Ziele das Saarbrücker Übersetzungssystem SUSY einzusetzen? Dies erschien insofern als ein besonderes Problem, als mit dem Auslaufen des Sonderforschungsbereichs Elektronische Sprachforschung zum Jahresende 1986 eine wesentliche Entwicklungsgrundlage entfallen ist. Zudem liegt mit SUSY-II eine konzeptionelle Fortentwicklung vor. Schließlich stand zu erwarten, dass zumindest mittelfristig mit EUROTRA ein Werkzeug zur Verfügung stehen würde, das in seinen Leistungsmöglichkeiten den Ergebnissen des Übersetzungssystems SUSY überlegen sein sollte. Zu diesen Systemen hatte die Projektgruppe kostenfreien Zugang. Bedenkenswert wäre auch gewesen, am Markt verfügbare Systeme wie SYSTRAN oder METAL heranzuziehen. Dies war jedoch aus Kosten- wie Organisationsgründen zum Zeitpunkt der Projektrealisierung nicht möglich. Da es mit Sicherheit in Zukunft leistungsfähigere softwaretechnologische Entwicklungen geben wird, kann diese Frage nicht absolut gesehen werden, sondern ist wie folgt zu relativieren: Eignet sich SUSY - in der vorliegenden Form - für das Erreichen des Projektziels? Antworten auf die hier gestellten Fragen geben die Beiträge in Kap. 2 dieses Buches.

Alle bestehenden maschinellen Übersetzungshilfen weisen im wesentlichen große Defizite im lexikalischen Inventar auf. Eine Übersetzung steht und fällt aber - das haben vor allem die systematischen Bewertungen von SYSTRAN gezeigt (vgl. van Slype 1979) - mit dem lexikalischen Inventar. Der Terminologie- und Wörterbuchpflege kommt also eine große Bedeutung, wenn nicht gar der Vorrang zu vor einer rein linguistischen (syntakto-semantischen) Verfahrensweise. Dies gilt besonders für den Fall, wenn das Ziel nicht - wie bei SYSTRAN und EUROTRA - die Erreichung einer qualitativ hochstehenden "rein" maschinellen Übersetzung (HQMT - High Quality Machine Translation) ist, sondern die Unterstützung der Übersetzung durch den Menschen. MARIS war auf Forschung und Entwicklung im letztgenannten Bereich ausgerichtet.

Das Konzept des Saarbrücker Translations-Systems (STS) trägt dabei den technologischen Entwicklungen der letzten Jahre Rechnung: Ziel war es, die Voraussetzungen zu schaffen, maschinenlesbar vorliegende Daten so aufzubereiten, dass ein Übersetzer mit möglichst geringem intellektuellen Aufwand (d.h. kostengünstig) hoch qualifizierte, maschinell weiterverarbeitbare (im vorliegenden Fall: in Datenbanken unmittelbar einbringbare) Übersetzungen liefert. Es darf dabei nicht vernachlässigt werden, dass bis heute noch keine Erfahrungen beim systematischen integrierten Einsatz des Computers im Übersetzungsbereich vorliegen, auch wenn - wie z.B. im Bundessprachenamt und beim Übersetzerdienst der Europäischen Gemeinschaft - im terminologischen Bereich Hilfen angeboten werden (Online- und Off-Line-Recherchen in Datenbanken wie LEXIS und EURODICAUTOM). Auf den sich entwickelnden Markt der interaktiven lexikongestützten Übersetzungshilfen (Anwendungen von elektronischen Terminologie- und Lexikonhilfen bei Verfahren wie INDEKS, LC-TOP, TERMEX etc.) kann an dieser Stelle nicht weiter eingegangen werden. Sie stellen in der Regel eine einfache Suchhilfe bereit, wobei der Computer (PC) beim "Blättern" eingesetzt wird (elektronisches Wörterbuch). Zum Zeitpunkt der Projektdurchführung waren diese Systeme zudem nicht verfügbar.

Mit dem STS ist für eine erweiterte Anwendung eine gute Voraussetzung geschaffen. Die hier vorgestellte Verfahrensweise der computergestützten Übersetzung ist naturgemäß nur ein Bei-

trag, eine Variante in der Palette möglicher und notwendiger Maßnahmen zur Überwindung der Sprachbarrieren in der Fachinformation. Es gilt jedoch, die Chance zu nutzen, inzwischen verfügbare Werkzeuge für eine entscheidende Verbesserung der internationalen Wissenschaftskommunikation einzusetzen. Ein "Europa der Forscher" ist ohne die Überwindung der Sprachbarrieren nicht zu leisten.