

Klaus Schubert

Zum gegenwärtigen Stand der maschinellen Übersetzung

1. Die Übersetzungsmaschine

1.1. Zwei alte Träume

Ähnlich wie das Fliegen gehört zu den alten Träumen der Menschheit auch die Idee, es müsse doch möglich sein, die Barriere zwischen Menschen verschiedener Sprache mit Hilfe technischer Geräte zu überwinden. Ideen und mehr oder weniger ausgearbeitete Pläne für Systeme der mechanischen Übersetzung tauchen in Europa seit dem 17. Jahrhundert immer wieder auf (vgl. Hutchins 1986: 21-23). Allerdings ist wohl nie eine solche Maschine gebaut worden.

Es besteht ein interessanter Zusammenhang zwischen den frühen Überlegungen über Möglichkeiten des mechanischen Übersetzens und einem anderen alten Traum, den zu derselben Zeit aufkommenden Plänen für Universal- oder Plansprachen. In manchen Fällen kommen Projekte beider Richtungen von denselben Autoren, und bisweilen entsteht der Entwurf einer Plansprache im Hinblick auf

Formen der Textwiedergabe in einem Zwischenstadium eines mechanischen Übersetzungsvorgangs, oder es wird umgekehrt die Idee des mechanischen Übersetzens als naheliegende Anwendung einer Plansprache vorgeschlagen (vgl. z.B. Klubkov 1984).

Der vorliegende Beitrag knüpft hier an. Er soll einen Einblick in die Arbeitsweise heutiger maschineller Übersetzungssysteme geben und die Rolle des Esperanto als der einzigen Plansprache beleuchten, die in der maschinellen Übersetzung (und allgemeiner in der Sprachtechnologie) Einsatz gefunden hat. Ich diskutiere unter diesem Blickwinkel sowohl gelöste als auch bisher ungelöste und einige aller Voraussicht nach unlösbare Teilaufgaben im automatisierten Übersetzungsvorgang und illustriere dies mit Rohübersetzungsergebnissen zweier gängiger Systeme.

1.2. Sprache und Computer

Erst mit der Erfindung des Computers entsteht eine reale Möglichkeit, den Traum von der Übersetzungsmaschine Wirklichkeit werden zu lassen. Man kann sogar mit einer gewissen Berechtigung behaupten, daß der Computer – zumindest in einem der Entwicklungsstränge, die zu dem geführt haben, was wir heute Computer nennen – von Anfang an eine Sprachmaschine ist. Bezeichnungen wie *Computer* oder *Rechner* lassen oft die Vorstellung vom Computer als einer besonders großen oder besonders schnellen Rechenmaschine entstehen. Tatsächlich ist aber Rechnen nur eine der Aufgaben, für die der Computer entwickelt worden ist.

Statt aus der Werkstatt eines einzigen Erfinders ist der Computer aus den Arbeiten mehrerer Gruppen von Wissenschaftlern und Ingenieuren hervorgegangen. Eine der beiden Entwicklungsrichtungen, die zum Durchbruch geführt haben, ist mit dem Namen des Mathematikers Alan Turing verbunden. Sie wurde im Zweiten Weltkrieg vom britischen Marinennachrichtendienst betrieben und diente der Decodierung verschlüsselter deutscher Funkprüche. Hier ging es also von Anfang an um Sprache. Die Funkprüche wurden abgehört, maschinell entschlüsselt und dann – natürlich manuell – aus dem Deutschen ins Englische übersetzt. Die Idee, nicht nur die Chiffrierung, sondern auch die fremde Sprache maschinell zu "entschlüsseln", lag auf der

Hand. Turing, der sehr weitblickende Vorstellungen von den Einsatzmöglichkeiten des Computers hatte und heute als Gründervater der Künstlichen Intelligenz gilt, hat sehr früh an die Möglichkeit der maschinellen Übersetzung gedacht (Turing 1969: 13 [posthum], Hodges 1983/1992: 382).

Als nach dem Zweiten Weltkrieg zivile Geldgeber für die Weiterentwicklung des Computers gesucht wurden, wandte sich Andrew Booth von der Universität London an die amerikanische Rockefeller Foundation und erhielt eine Auflage, die für die Entwicklung des Computers ebenso weitreichende Konsequenzen bekommen sollte, wie Booths Lösungsvorschlag für die Sprachtechnologie. Sein Gesprächspartner, das Vorstandsmitglied der Rockefeller Foundation Warren Weaver, machte die Förderung nämlich davon abhängig, daß das neue, kostspielige Gerät nicht bloß zum Rechnen dienen dürfe. Die nichtrechnerische Anwendung, die Booth in Erinnerung an Gespräche mit Turing nannte, war das Übersetzen. Dies fiel in New York auf fruchtbaren Boden, denn Weaver war selbst an der Idee des maschinellen Übersetzens interessiert und hat sich später auch zu diesem Thema geäußert. In seiner Denkschrift von 1949 (veröffentlicht Warren Weaver 1955) findet sich dann auch die Metapher vom Entschlüsseln der Ausgangssprache. (Dieser Absatz folgt Hutchins 1986: 24-25.)

2. Grundlegende Systemarchitekturen

Im ersten Enthusiasmus glaubte man, in kurzer Zeit übersetzende Computerprogramme ausarbeiten zu können. Dabei ging man mit Selbstverständlichkeit davon aus, daß die maschinelle Übersetzung ohne weiteres Eingreifen des Menschen verlaufen und

ein gebrauchsfertiges Resultat abliefern würde. Inzwischen hat man durch mühevoll erarbeitete Erfahrung gelernt, vollautomatische Qualitätsübersetzung, wie man dies heute nennt, als grundsätzlich unerreichbar zu betrachten.

2.1. Direkte Übersetzungssysteme

Die ersten experimentellen Übersetzungsprogramme verfolgten einen (aus heutiger Sicht) naiven Ansatz. Man ließ das Programm ganz einfach jedes Wort des Ausgangstextes – der Kalte Krieg und der Sputnikschock hatten in den USA statt deutschsprachiger jetzt russische Texte in den Mittelpunkt des Interesses gerückt – durch ein Wort der Zielsprache ersetzen. Hierzu benötigt das System ein Wörterbuch, das nicht viel mehr war als eine Liste russischer Wörter mit englischen Übersetzungsäquivalenten.

Beispiele für die so erreichbare Qualität zeigt John Hutchins (1986: 27-28). Die Ursachen, durch die diese Form der Wort-für-Wort-Übersetzung unbefriedigend bleibt, sind mehrfacher Art. Zum einen entsprechen sehr oft einem ausgangssprachlichen Wort mehrere Wörter der Zielsprache, die nicht etwa synonym sind, sondern aus denen je nach Struktur und Inhalt des zu übersetzenden Textes das richtige auszuwählen ist. Dies ist das Problem der lexikalischen Übertragung. Zum anderen stehen die Wörter im englischen Satz nicht in derselben Reihenfolge wie im russischen, und zum dritten gibt es viele englische Wörter, deren Anwesenheit im Satz gar nicht durch ein bestimmtes russisches Wort, sondern durch eine grammatische Form oder Konstruktion, eine bestimmte Wortfolge oder ein anderes

Merkmal des Ausgangstextes verursacht ist. Dies ist das Problem der Strukturübertragung oder Metataxe. Wie ein Blick auf die veröffentlichten Übersetzungsbeispiele zeigt, sind wortweise Übersetzungen fast durchweg unbrauchbar, selbst wenn schon gewisse morphologische und syntaktische Verschönerungen angebracht wurden. Daß man mit einer solchen Methode dennoch bisweilen Ergebnisse erzielen kann, die immerhin so gut sind, daß die Geldgeber hoffnungsvoll gestimmt bleiben, wie etwa bei dem berühmten Georgetown-IBM-Experiment von 1954 (Hutchins 1986: 36-37), liegt zu einem wesentlichen Teil daran, daß viele Entwickler solcher Systeme der Versuchung der Abstimmung des Systems auf die Testtexte erliegen, wie Alan Melby (1995: 19-22) ausführlich beschreibt.

Dieser erste, naheliegendste Ansatz der maschinellen Übersetzung wird heute direkter Ansatz genannt. Er setzt stillschweigend voraus, daß sich der Sinn des Gesamttextes von selbst wieder ergibt, wenn man die Wörter einzeln, isoliert und unabhängig von ihrem Kontext übersetzt und diese Wortübersetzungen dann wieder in derselben Reihenfolge aneinanderreihet, in der ihre Entsprechungen im Ausgangstext stehen. Wie sich schnell erweisen sollte, trifft diese Voraussetzung nicht zu.

2.2. Transfersysteme

Da die direkte Übersetzung sehr unbefriedigende Ergebnisse liefert, entschloß man sich, nicht mehr wort- sondern satzweise zu übersetzen. Die Wörter eines Satzes wurden als Elemente einer syntaktischen Struktur betrachtet. Hierdurch entstand die Notwendigkeit, das Übersetzungssystem mit einem Parser auszustatten, einem Modul, das die syntaktische Struktur von Sätzen erkennt und notiert.

Hierdurch entsteht der nächste Ansatz: der Transferansatz. Hier verläuft der Übersetzungsvorgang in drei recht deutlich getrennten Teilschritten: der Analyse, der Übertragung und der Synthese. (Die Übertragung heißt auch Transfer und die Synthese auch Generierung.)

Die Analyse ist ein Bearbeitungsgang, der sich aus-

schließlich mit dem Text in der Ausgangssprache beschäftigt. Er beginnt mit dem Text in normaler linearer Form (etwa einer Textdatei aus einem Textverarbeitungssystem), zerlegt ihn in Sätze und kennzeichnet die einzelnen Wörter im Satz mit Etiketten, die ihre syntaktische Funktion angeben. Üblicherweise wird ein so analysierter Satz in Form einer Baumstruktur wiedergegeben. Wie der Baum genau aussieht, hängt von dem zugrundegelegten Syntaxmodell ab.

Der Übertragungsschritt ist dann eine Umsetzung der syntaktischen Baumstruktur des ausgangssprachlichen Satzes in eine Baumstruktur, die der Syntax der Zielsprache entspricht. Dies ist die Strukturübertragung. Gleichzeitig (oder vor- oder

nachher) werden die Wörter der Ausgangssprache durch Wörter der Zielsprache ersetzt. Dies ist die lexikale Übertragung. Da menschliche Sprachen nicht eins zu eins aufeinander abbildbar sind, da also ein zielsprachliches Wort Übersetzungsentsprechung eines Mehrwortausdrucks in der Ausgangssprache sein kann und umgekehrt, führen nicht nur die Regeln der Strukturübertragung, sondern auch die der lexikalen Übertragung zu Bedingungen für die Struktur des zielsprachlichen Satzes, die es aufeinander abzustimmen gilt. All dies ist Aufgabe des Übertragungsschritts in der maschinellen Übersetzung nach dem Transfermodell.

Auf die Übertragung folgt die Synthese. Das Modul, das diesen Schritt ausführt, erhält als Eingabe die vom Übertragungsschritt erzeugte syntaktische Baumstruktur eines Satzes in der Zielsprache. Die

Aufgabe der Synthese ist es, den Baum in einen linearen Satz umzuwandeln. Hierbei müssen die Wörter in die ihrer syntaktischen Funktion entsprechende morphologische Form gebracht werden. Redundante Regelmäßigkeiten wie Kongruenz und Rektion sind zu berücksichtigen, und schließlich müssen die Wörter in der richtigen Reihenfolge angeordnet werden. Je nach der grammatischen Grundkonstellation der betreffenden Sprache sind bei der Erzeugung der richtigen Wortfolge redundante sprachinterne Regeln zu beachten (dies ist etwa im Englischen sehr ausgeprägt), während in anderen Sprachen (z.B. im Russischen und in geringerem Maße im Deutschen) Wortfolgeregeln wirksam werden, die satzübergreifender Art sind und den Textzusammenhalt und die kommunikative Gliederung des Satzes steuern.

2.3. Zwischensprachsysteme und isomorphe Grammatiken

Die meisten heute verfügbaren maschinellen Übersetzungssysteme folgen dem Transferansatz. Das Merkmal, dem er seinen Namen verdankt, ist der Transfer- oder Übertragungsschritt, bei dem syntaktische Strukturen der Ausgangssprache in solche der Zielsprache umgesetzt werden. Es gibt zwei Ansätze, die hier anders verfahren. Beiden liegen Überlegungen des kombinatorischen Aufwands zugrunde, den man zu treiben hat, wenn man ein Übersetzungssystem mit mehreren Sprachen entwickeln und dieses System überdies erweiterbar halten möchte. Möchte man beispielsweise zwischen den Sprachen Deutsch, Englisch und Französisch in jeder Richtung übersetzen, braucht man nicht für jede der sechs Kombinationen alle drei Module (Analyse, Übertragung, Synthese) neu zu entwickeln. Da der Analyse- und der Syntheseschritt rein einsprachig verlaufen, kann dasselbe Analysemodul für Deutsch im Übersetzungssystem Deutsch-Englisch und im System Deutsch-Französisch verwendet werden. Auch wenn man später Spanisch, Russisch und Japanisch hinzufügt, braucht das deutsche Analysemodul nicht verändert zu werden. Dasselbe gilt für die Synthesemodule. Die Übertragungsmodule arbeiten jedoch sprachenpaarweise, so daß man bei zwei Sprachen zwei Übertragungsmodule braucht (z.B. Deutsch-Englisch und Englisch-Deutsch), bei drei Sprachen sechs, bei vier Sprachen zwölf, bei fünf Sprachen zwanzig und so weiter. Die Zahl der Übertragungsmodule steigt also sehr viel schneller an als die der Sprachen. (Für n Sprachen sind $n \times (n-1)$ Übertragungsmodule nötig.) Diesem Dilemma der kombinatorischen Vielfalt kann man entkommen, wenn man an die Stelle der direkten Übertragung zwischen Ausgangs- und Zielsprache ein einheitliches, zentrales Zwischenglied X setzen kann, so daß aus jeder Ausgangssprache in X und aus X in jede beliebige Zielsprache über-

setzt wird. Gelingt dies, so sind keine Übertragungsmodule mehr nötig, sondern nur noch für jede beteiligte Sprache ein Analyse- und ein Synthesemodul, wobei allerdings Analyse und Synthese sehr viel weiter zu gehen haben als in einem Transfer-system.

Eine derartige Lösung kann nur dann funktionieren, wenn die Zwischenform X bei der Übersetzung aus dem Deutschen in X und aus X ins Englische nicht anders auszusehen braucht, als wenn aus dem Deutschen in X und aus X ins Französische übersetzt wird. Umgekehrt darf das X , aus dem ins Französische übersetzt wird, nicht davon abhängen, ob vorher aus dem Deutschen oder aus dem Englischen in X übersetzt wurde. Wäre dies nicht gewährleistet, müßte das Synthesemodul, das beispielsweise aus X ins Französische übersetzt, mit unterschiedlichen Regeln ausgestattet sein, je nachdem, ob die Ausgangssprache Deutsch oder Englisch war. In einem System mit n Sprachen wären wieder gerade $n \times (n-1)$ solcher Regelsysteme erforderlich. Da das Zwischenglied X jedoch eingeführt werden sollte, um diese kombinatorische Vielfalt zu vermeiden, muß also die Bedingung gelten, daß das Zwischenglied von Ausgangs- und Zielsprachen unabhängig ist.

Ein solches X heißt Zwischenrepräsentation. (Auch die Bezeichnung Interlingua ist gängig; sie ist jedoch hier, in einem interlinguistischen Zusammenhang, verwirrend.) Damit eine Zwischenrepräsentation von den Ausgangs- und Zielsprachen unabhängig sein kann, muß sie über eigene lexikale und strukturelle Elemente verfügen, also ein eigenes Sprachsystem besitzen. Am einfachsten findet man ein solches System, indem man eine vorhandene Sprache wählt. In diesem Falle, wenn es sich also bei der Zwischenrepräsentation um eine Sprache (und nicht um ein künstliches Symbolsystem)

handelt, spricht man von einer Zwischensprache. Wenn der Übersetzungsvorgang Ausgangs- und Zielsprache nie direkt miteinander in Kontakt bringt, sondern immer die Zwischensprache dazwischenschaltet, fallen wie erwähnt dem Analyse- und dem Synthesevorgang umfangreichere Aufgaben zu als in einem Transfersystem. Wenn die Zwischenrepräsentation eine Sprache ist, ist leicht zu sehen, welcher Art diese Aufgaben sind: es muß eine vollständige Strukturübertragung und eine vollständige lexikale Übertragung aus der Ausgangs- in die Zwischensprache stattfinden, so daß die zwischensprachliche Form des zu übersetzenden Textes keinerlei (grammatische oder lexikale) Hinweise mehr darauf enthält, aus welcher Ausgangssprache der Text stammt.

Maschinelle Übersetzungssysteme mit einer echten Zwischensprache, wie das System DLT, von dem unten (Abschnitt 6.1.) die Rede ist, haben den Übertragungsschritt also im Grunde nicht beseitigt, sondern führen ihn zweimal aus. Ein solches System ist ein doppeltes Transfersystem, in dem erst ein Übertragungsschritt aus der Ausgangssprache in die Zwischensprache und dann ein zweiter Übertragungsschritt aus der Zwischensprache in die Zielsprache durchgeführt wird (Schubert 1988a). Der angestrebte Gewinn, der darin besteht, daß für n Sprachen nicht $n \times (n-1)$ Systeme, sondern nur $2n$ erforderlich sind, ist allerdings hierdurch durchaus erreicht, wenn auch die erforderlichen Module umfangreicher sind als die Analyse- und Synthesemodule in einem einfachen Transfersystem.

Während eine Zwischenrepräsentation über eigene, von Ausgangs- und Zielsprachen unabhängige strukturelle und lexikale Elemente verfügen muß, ist auch versucht worden, eine Teillösung zu verwirklichen, bei der die lexikalischen Elemente jeweils den Ausgangs- und Zielsprachen entstammen, die Analyse des Ausgangssatzes jedoch so weit fortgeführt wird, bis eine syntaktische Struktur entsteht, die für übersetzungsäquivalente Sätze aller am System beteiligten Sprachen dieselbe ist. Ein solcher Ansatz setzt voraus, daß die Syntaxen der Ausgangs- und Zielsprachen auf einen gemeinsamen Nenner gebracht werden können, der dann auf der Ebene der Struktur als Zwischenglied fungiert. Das (vor Marktreife abgebrochene) Forschungs- und Entwicklungsprojekt des Philips-Konzerns, Rosetta, ging von einem solchen Ansatz isomorpher Grammatiken aus (Landsbergen 1987, 1994, Hutchins 1988: 38-39). Es wurde experimentell für die Sprachen Niederländisch, Spanisch und Englisch entwickelt. Ein wesentliches Problem dieses Ansatzes liegt darin, daß die isomorph gemachten Grammatiken aller beteiligten Sprachen jedesmal überarbeitet werden müssen, wenn eine weitere Sprache hinzugefügt wird. Im Falle einer Sprache, die den bereits im System vorhandenen typologisch weniger nahesteht, dürfte der Aufwand hierfür erheblich sein.

Auch der Übersetzungsalgorithmus Petr Trojanskij aus den 30er Jahren (vgl. Abschnitt 6.1. unten) beruht im Grunde auf der Vorstellung von der Isomorphie der Grammatiken verschiedener Sprachen.

3. Der maschinelle Übersetzungsvorgang

Um der Frage nachzugehen, was maschinelle Übersetzung heute leisten kann und was in absehbarer Zukunft von ihr erwartet werden kann, gebe ich zunächst einen kurzen Einblick in die Funktionsweise dieser Systeme.

Die meisten auf dem Markt verfügbaren maschinellen Übersetzungssysteme sind nach dem Transferansatz aufgebaut, wobei einige der kleineren Systeme vereinfachte Lösungen und Abkürzungswege verwenden dürften, die Anleihen bei dem direkten Ansatz machen. Nicht in allen Fällen ist die genaue Funktionsweise der einzelnen Systeme bekannt, da die sprachtechnologischen Einzel-

heiten oft Teil des Betriebsgeheimnisses sind. Bei großen Systemen, deren Entwicklung sich auf langjährige, z.T. mit öffentlichen Mitteln geförderte Forschungsprojekte stützt, gibt es oft recht detaillierte Veröffentlichungen. Bei einigen kleineren PC-Systemen ist die Funktionsweise hingegen meist nur nach dem Black-Box-Verfahren zu ermitteln. Die folgende Darstellung beschreibt daher nicht ein bestimmtes System, sondern beschreibt in allgemeiner Form die wichtigsten Bearbeitungsschritte im maschinellen Übersetzungsvorgang bei Systemen mit Transferarchitektur.

3.1. Parsing, Strukturanalyse und syntaktische Synthese

Wie beschrieben verläuft der Übersetzungsvorgang bei Transfersystemen in drei großen Bearbeitungsgängen, die in weitere Teilschritte unterteilt sein können. Der Analyseschritt formt den ausgangssprachlichen Satz in eine Baumstruktur um, in der

die syntaktischen Funktionen der Wörter und Wortgruppen angegeben sind. Er ist einer der schwierigsten Schritte in der maschinellen Übersetzung. Die wesentlichsten Teilschritte sind folgende:

Eigentliches Parsing

Umformung einer linearen Kette von Wortformen in eine zusammenhängende syntaktische Struktur. Parsing setzt eine formalisierte Syntax der Ausgangssprache voraus. Eine Syntax beschreibt die Kombinierbarkeit von Wortarten und anderen Klassen von Wörtern wie Valenzklassen u.ä. Daher setzt die Anwendung einer Syntax die Zuordnung der Wörter des zu analysierenden Satzes zu Wortarten voraus.

Wortarterkennung

Zuordnung von Wörtern zu Wortarten. Wortarterkennung setzt ein syntaktisches Wörterbuch voraus, das jedem Wort der Sprache eine Wortart zuordnet. Sie setzt weiter eine morphologische Reduktion flektierter Wortformen auf Grundformen und die Erkennung von Komposita voraus, die im Wörterbuch nicht selbst vorkommen, aber aus bekannten Wörtern zusammengesetzt sind.

Morphologische Reduktion

Zuordnung von flektierten Wortformen zu Grundformen.

Analyse von Komposita

Zerlegung zusammengesetzter Wortformen in bekannte Elemente. Setzt Regeln über die Kombinierbarkeit von Wörtern, Wortstämmen, Morphemen und Fugenelementen voraus.

Jeder dieser Bearbeitungsschritte kann auf Mehrdeutigkeiten stoßen, wobei die Mehrdeutigkeiten verschiedener Analyseschritte ineinandergreifen können.

Analyse von Komposita

dt. *Postfacharbeiter*

Die Analyse eines Kompositums ist im Analyseschritt nur erforderlich, wenn es nicht als Ganzes im Wörterbuch steht. Wenn *Postfacharbeiter* im Wörterbuch fehlt, die häufigeren Komposita *Postfach* und *Facharbeiter* jedoch vorhanden sind, entsteht Mehrdeutigkeit: *Post + Facharbeiter* oder *Postfach + Arbeiter*? Für das Ergebnis des Parsings ist die Entscheidung unerheblich, da beide Lösungen ein Wort ergeben, das die morphologischen und syntaktischen Merkmale des Wortes *Arbeiter* übernimmt.

Die anschließende lexikale Übertragung kann jedoch in die Irre gehen, wenn die falsche Analyse gewählt wird (vgl. de Wachter 1995).

Morphologische Reduktion

engl. *saw*

- Wortart: Verb, Grundform: *see* 'sehen', Tempus: Präteritum
- Wortart: Verb, Grundform: *saw* 'sägen', Tempus: Präsens, Numerus/ Person: nicht 3. Sg.
- Wortart: Substantiv, Grundform: *saw* 'Säge', 'Weisheit'...!, Numerus Singular, Kasus: nicht Genitiv

Wortarterkennung

engl. *present*

- Wortart: Adjektiv ('jetzig', 'anwesend'...)
- Wortart: Substantiv ('Geschenk'...)
- Wortart: Verb ('überreichen', 'schenken', 'präsentieren'...)

Durch die Mehrdeutigkeiten der zuarbeitenden Analysen und durch erhebliche Mehrdeutigkeiten in der Syntax selbst verläuft der eigentliche Parsingvorgang im allgemeinen in vielen alternativen Analysewegen. Ein Parser erzeugt dabei wesentlich mehr syntaktisch korrekte, aber im aktuellen Text nicht richtige Analysen eines Satzes als ein Mensch, der dieselbe Aufgabe hätte. Ein Mensch schließt im allgemeinen eine Vielzahl formal möglicher Analysen als unsinnig aus. Da gängige Parser keinen Zugriff auf die Bedeutung von Sätzen oder Texten haben, sondern streng nach formalen grammatischen Regeln der Morphologie, Wortbildung und Syntax verfahren, können sie solche Entscheidungen zwischen Sinnvollem und Unsinnigem nicht treffen. Die syntaktische Mehrdeutigkeit oder Ambiguität ist damit eine der größten Fehlerquellen in der maschinellen Übersetzung.

Der Analyseschritt hat vor allem mit Strukturambiguität zu kämpfen. Er ist deswegen in bisherigen maschinellen Übersetzungssystemen der aufwendigste Bearbeitungsgang. Die nachfolgende Strukturübertragung erfordert ebenfalls ein komplexes Regelsystem, das auf zahlreiche Ambiguitäten stößt und eine Vielzahl von Auswahlentscheidungen zu treffen hat. Soweit die Strukturübertragung jedoch davon ausgehen kann, daß die ihr zur Bearbeitung vorgelegte syntaktische Struktur richtig ist, arbeitet sie im Gegensatz zur Analyse mit deutlich nach den im System verwendeten Regelwerken strukturiertem Ausgangsmaterial und verläuft somit reibungsloser als der Analyseschritt. Die Strukturübertragung leidet jedoch darunter, daß ein Teil der zu treffenden Auswahlentscheidungen entweder von

der Bedeutung des Textes abhängen würde, auf die das maschinelle Übersetzungssystem normalerweise keinen Zugriff hat, oder von übergeordneten textstrukturierenden Kohärenzmechanismen, die zur Zeit mangels entsprechend klar formalisierbarer Forschungsergebnisse noch nicht im wünschenswerten Maße zur Verfügung stehen.

Die syntaktische Synthese ist jener Schritt, bei dem die von der Strukturübertragung erzeugte syntaktische Baumstruktur in der Zielsprache in einen normalen linearen Text umgewandelt wird. Hier geht es darum, die in der Baumstruktur enthaltenen syntaktischen Funktionsangaben (z.B. indirektes Ob-

jekt) in solche Kennzeichen umzusetzen, wie die Zielsprache sie verwendet. Hierdurch erhalten die Wörter und Wortgruppen ihre Stellung in der Wortfolge des Satzes und ihre morphologische Formkategorie (z.B. Dativ). Anschließend muß die entsprechende Form (im Beispiel der Dativ) von dem vorliegenden Wortstamm gebildet werden. Die Synthese erfordert nur in geringem Maße Entscheidungen, die wegen des fehlenden Zugangs zur Semantik nicht gefällt werden können. Sie bleibt allerdings zu einem gewissen Grade unterbestimmt, weil textgrammatische Mechanismen fehlen (Schubert 1992a).

3.2. Lexikale Übertragung

Die Analyse ist bisher in maschinellen Übersetzungssystemen der aufwendigste Arbeitsgang. Die Ursache hierfür liegt in den erheblichen Problemen, die die Entscheidung syntaktischer Ambiguitätsfälle verursacht. Eine weitere Ursache für das Übergewicht der Analyse ist jedoch darin zu suchen, daß ein anderer Vorgang, der noch wesentlich aufwendiger sein könnte, bisher vernachlässigt wurde: die lexikale Übertragung.

Bei der lexikalen Übertragung geht es, vereinfacht gesagt, um die Wahl zwischen alternativen Übersetzungsäquivalenten für einzelne Wörter oder Wortgruppen der Ausgangssprache. Es ist dasselbe Problem, dem sich jeder Wörterbuchbenutzer gegenüber sieht, der ein Wort in einem zweisprachigen Wörterbuch nachsieht und mehr als einen Übersetzungsvorschlag angeboten bekommt.

Die lexikale Übertragung ist damit jener Arbeitsgang der maschinellen Übersetzung, der versucht, mit der altbekannten Tatsache fertigzuwerden, daß die Sprachen das Kontinuum der Bedeutung, über die Menschen kommunizieren, nicht in derselben Weise aufteilen, wodurch die Wörter zweier Sprachen einander fast nie eins zu eins entsprechen. Es ist bekannt, daß dieses Phänomen bei häufig verwendeten Wörtern mehr ins Gewicht fällt als bei seltenen, wodurch die lexikale Übertragung in der Gemeinsprache schwieriger ist als in der Fachsprache. Bisweilen wird (z.B. von Melby 1995: 49-51) die These vertreten, dies sei ein absoluter Unterschied zwischen Fach- und Gemeinsprache, wodurch vollautomatische Übersetzung bei fachsprachlichen Texten möglich, bei gemeinsprachlichen dagegen unmöglich sei. Da jedoch auch fachsprachliche Texte immer gemeinsprachliche Elemente enthalten, ist der Unterschied zwischen beiden Textsorten nach meiner Erfahrung gradueller Natur, so daß maschinelle Übersetzung von Fachtexten unter Bestimm-

ten Bedingungen einfacher, aber nicht in irgendeiner absoluten Weise anders verläuft als bei gemeinsprachlichen Texten (vgl. Schubert 1996, demn.). Im Hinblick auf mögliche Unterschiede zwischen Fach- und Gemeinsprache sei die lexikale Übertragung hier mit den beiden Beispielen *drive* und *condensation* illustriert. Sie entstammen dem Beispieltext, auf den ich in 4.2. näher eingehe. Die auf der folgenden Seite gezeigten Übersetzungsäquivalente entnehme ich einem gemeinsprachlichen und einem fachsprachlichen Wörterbuch (Langenscheidts enzyklopädisches Wörterbuch 1992, Ernst 1985)².

Die Wörterbucheinträge für die Stichwörter *drive* und *condensation* zeigen, aus welcher Vielfalt an Übersetzungsäquivalenten die lexikale Übertragung auszuwählen hat. Die Beispiele sind hier rein illustrativ gemeint – selbstverständlich muß ein maschinelles Übersetzungssystem in den speziellen Fachgebieten, in denen es eingesetzt werden soll, mit detaillierteren Einträgen ausgerüstet sein, als es das allgemein auf industrielle Technik ausgerichtete Wörterbuch von Ernst (1985) leistet. Andererseits kann ein solches System auf manche gemeinsprachliche Übersetzungsäquivalente verzichten, die hier aus *Langenscheidts enzyklopädischem Wörterbuch* (1992) gezeigt werden. Beim Verzicht auf gemeinsprachliche Übersetzungsäquivalente ist jedoch Vorsicht geboten. Das Beispiel *drive* zeigt dies. Fachsprache tendiert zu substantivischem Stil, wobei der Hauptmitteilungsgehalt in den Substantiven liegt und die Verben und Funktionsverbgefüge semantisch blaß sind. Dies führt dazu, daß in fachsprachlichen Texten gerade besonders häufig Allerweltsverben vorkommen. Daher kann ein maschinelles Übersetzungssystem nicht unbesehen auf die gemeinsprachlichen Übersetzungsäquivalente eines Eintrags wie *drive* verzichten.

drive, Substantiv (Langenscheidts enzykl. Wörterbuch 1992)

1. ~~Fahrt~~, bes. Spazierfahrt, Ausflug.
2. Treiben (*Vieh, Holz etc.*).
3. Zusammentreiben.
4. Jagd Treibjagd.
5. zusammengetriebene Tiere.
6. *Psychologie* a) Antrieb, Motiv, Beweggrund, b) Neigung, Tendenz.
7. (*Tennis, Golf etc.*) Drive, Treibschlag.
8. *Militär* Vorstoß, heftiger Angriff, kraftvolle Offensive.
9. Vorstoß, energische Unternehmung.
10. *figürlich* Kampagne, Feldzug, bes. Werbefeldzug, großangelegte (Werbe)Aktion.
11. *Volkswirtschaft, amerikanisches Englisch, umgangssprachlich* große Verkaufsaktion zu herabgesetzten Preisen, Verbilligungskampagne.
12. Hochdruckbetrieb, auf Hochtouren laufender (Geschäfts) Betrieb.
13. lebhafte Bewegung.
14. Trieb-, Stoßkraft.
15. Schwung, Tempo, Energie.
16. Tendenz, Strömung, Richtung, Neigung.
17. Fahrstraße.
18. *brit. Englisch (private)* Auffahrt (*zu einer Villa etc.*).
19. *Technik* Antrieb.
20. *Technik* Antriebs-, Betriebsart.
21. Floß, geflößte Baumstämme.

drive, Adjektiv (Langenscheidts enzykl. Wörterbuch 1992)

22. *Technik* Antriebs..., Trieb..., Treib...

drive, trans. Verb (Langenscheidts enzykl. Wörterbuch 1992)

23. (vorwärts-, an)treiben, mit sich treiben.
24. *figürlich* treiben.
25. (ein)treiben, (ein)rammen, (ein)schlagen.
26. (zur Arbeit) antreiben, überanstrengen, -lasten, hetzen, jagen.
27. veranlassen (to, into zu; to do zu tun), bringen (to, into zu), dazu bringen (to do zu tun).
28. nötigen, zwingen (to, into zu; to do zu tun).
29. zusammentreiben, vor sich hertreiben.
30. forttreiben, vertreiben, verjagen (from von).
31. Jagd hetzen, jagen.
32. Jagd durchstöbern.
33. (*Auto etc.*) lenken, steuern, fahren.
34. (*Zugtiere*) lenken, (*Pflug*) führen.
35. (im Auto etc.) befördern, fahren, bringen (to nach).
36. *Technik* antreiben.
37. (*Feder etc.*) führen.
38. zielbewußt durchführen, zum Abschluß bringen.
39. (*Gewerbe*) (zielbewußt) (be)treiben.
40. (*Stollen, Tunnel etc.*) bohren, vortreiben.
41. *umgangssprachlich* hinausschieben, -zögern.
42. *Sport (Ball)* mit einem Treibschlag ab- od. zurückspielen, kräftig schlagen.
43. (*Golf (Ball)*) vom Abschlagmal kräftig abspielen.

drive, intrans. Verb (Langenscheidts enzykl. Wörterbuch 1992)

44. (dahin)treiben, (dahin)getrieben od. getragen werden.
45. rasen, eilen, brausen, jagen, stürmen, stürzen, rennen.
46. a) (*Auto*) fahren, chauffieren, ein od. das Auto len-

ken,
b) kutschieren.

47. (spazieren)fahren, eine (Spazier)Fahrt unternehmen.
48. *sport* einen Treibschlag ausführen.
49. (*Golf*) den Ball mit einem Treibschlag vom Abschlagmal abspielen.
50. zielen (at auf acc).
51. ab-, hinzielen, hinauswollen (at auf acc).
52. schwer arbeiten (at an dat).
53. einen Stollen vortreiben.
54. *britisches Englisch* (das Vieh) auftreiben.

drive, trans. Verb (Ernst 1985)

- antreiben, treiben
- (Kraftfahrzeuge) steuern, fahren, lenken
- *sink* (Bergbau) abteufen, niederbringen, absenken (Schacht)
- (Elektronik, Akustik, Computer) treiben
- [at] (allgemein) zielen [nach, auf]
- *advance* (allgemein) vortreiben
- *wedge in* (Bergbau) hereinkeilen, -treiben

drive, intrans. Verb (Ernst 1985)

- (Kraftfahrzeuge) autofahren
- [*against*], *collide* auffahren [auf], anfahren, kollidieren, zusammenprallen

drive, Substantiv (Ernst 1985)

- Wagenfahrt, Fahrt
- *gear* Antrieb, Antriebsorgan, Getriebe, Triebwerk
- *impact force* Stoßkraft, Vorstoßkraft
- (Raumfahrttechnik) Triebwerk
- *approach* Auffahrt, Rampe
- (Elektronik, Röhren) Gitteransteuerung
- (Computer) magnetische Erregung von Kernen in Amperewindungen, Treiben
- (Elektronik) Ansteuerung
- (Straßenbau, Verkehr) Privateinfahrt, -anfahrstraße

condensation, Substantiv (Langenscheidts enzyklopädisches Wörterbuch 1992)

1. Kondensieren, Verdichten, Kondensation, Verdichtung, Eindickung, Verflüssigung
2. Kondensat, Kondensationsprodukt, Niederschlag, Schwitzwasser
3. *Chemie* Kondensation
4. *Physik* a) Kondensation, Verdichtung (*Gase etc.*), b) Konzentration des Lichtes
5. (*Psychoanalyse*) Wiedergabe (zweier od. mehrerer Gedanken, Erinnerungen, Gefühle od. Impulse) durch ein Wort od. Wortbild (*in Allegorien, Träumen etc.*)
6. Zusammendrängung, Anhäufung
7. *figürlich* gedrängte Kürze, Abkürzung, Zusammenfassung
8. gekürzte Fassung (*eines Romans etc.*)

condensation, Substantiv (Ernst 1985)

- (allgemein, Chemie) Kondensation, Kondensierung, Verdichtung, Verflüssigung

Die lexikale Übertragung ist eine Auswahl Aufgabe. Hierfür ist ein Entscheidungsmechanismus notwendig, dem wiederum Wissen und Entscheidungskriterien zur Verfügung stehen müssen. Der lexikalen Übertragung können folgende Arten von Wissen dienen, wobei beim heutigen Stand der Entwicklung nicht alle Arten bereits in markteingeführten Systemen angewendet werden:

Syntax

Die syntaktische Analyse stellt fest, ob das Wort *drive* im aktuellen Satz Substantiv, Adjektiv oder Verb ist. Ist es Substantiv, kommen nur die Übersetzungsäquivalente in Frage, die der Substantivart von *drive* zugeordnet sind. (Wenn das Übersetzungsäquivalent ein deutsches Substantiv ist, liegt der "Normalfall" vor, in dem die syntaktische Form in Ausgangs- und Zielsprache "dieselbe" ist. Wenn nicht, bewirkt die lexikale Übertragung gleichzeitig auch eine Strukturveränderung. Vgl. Schubert 1987: 158.)

Syntaktisches Wissen kann die Zahl der Alternativen verringern. Es bietet keine Garantie dafür, daß in jedem Fall eindeutig eine einzige Lösung ausgewählt wird.

Vertikale Aufteilung des Wortschatzes in Fachgebiete

Viele der Übersetzungsäquivalente in den gezeigten Beispielen sind mit einer Fachgebietsangabe versehen. Dies ist offenbar ein brauchbares Instrument, das den Menschen, für die die Wörterbücher gemacht sind, die Auswahl erleichtert.

Auch maschinelle Übersetzungssysteme verwenden eine derartige vertikale Aufteilung des Wortschatzes sehr häufig. Bei einfacheren Systemen ist dies so realisiert, daß der Anwender beim Starten eines Übersetzungsauftrags angibt, ob das System ein bestimmtes Fachwörterbuch verwenden soll. Bei komplexeren Systemen kann der Anwender einen Suchpfad durch eine Hierarchie über- und untergeordneter Fachwörterbücher angeben, so daß das System seine Fachwortschätze in der so vorgegebenen Reihenfolge durchsucht. Beiden Systemtypen liegt die Methode zugrunde, daß das System das zuerst gefundene Übersetzungsäquivalent für das aktuelle Wort verwendet, wobei der Anwender durch die Steuerung der Suchreihenfolge Einfluß darauf hat, welches das erst gefundene Äquivalent ist.

Die Methode ist mit Einschränkungen effektiv. Sie funktioniert gut, solange der zu übersetzende Text sich innerhalb eines einzigen Fachgebiets hält (Lukas 1994: 82). Berührt

ein Text mehrere Fachgebiete (oder ein Fachgebiet, das sich nicht mit der für das aktuelle maschinelle Übersetzungssystem gewählten Fachgebietseinteilung deckt), wäre es notwendig, daß der Anwender nicht einmal für den gesamten Text eine Fachgebietsangabe vorausschickt, sondern jeweils kapitelweise, abschnittsweise, satzweise oder syntagmenweise präzisere Fachgebietsangaben macht. Das Arbeiten mit Fachgebietsangaben ist aufwendig und arbeitsintensiv. Es setzt unter anderem voraus, daß jedes Paar aus einem ausgangssprachlichen Wort und einem zielsprachlichen Übersetzungsäquivalent im Wörterbuch des Systems einem Fachgebiet zugeordnet ist. Ist der Extremfall, die Befragung des Anwenders für jedes Syntagma, gegeben, dann kann man auf den Aufwand der Auszeichnung der Wörterbucheinträge verzichten. Statt den Anwender nach Fachgebieten zu fragen, legt das System dem Anwender dann einfach direkt die Übersetzungsäquivalente für das aktuelle Wort zur Auswahl vor. Es ist offensichtlich, daß bei einem solchen Verfahren der Automatisierungsgrad niedrig ist: der Anwender muß während des gesamten Übersetzungsvorgangs anwesend sein und mitarbeiten, und er muß Übersetzerkompetenz für Ausgangs- und Zielsprache besitzen. Das Translation Support System der Firma ALPS (heute ALPNET) war auf dieser Form des interaktiven Dialogs aufgebaut. Es ist nicht mehr auf dem Markt (vgl. Hutchins 1988: 27, Newman 1988, Alan Weaver 1988, Schmid 1992). Fachgebietseinteilung ist ein gutes, aber kein hinreichendes Instrument der lexikalen Übertragung. Es funktioniert, solange die vom Anwender getroffene Fachgebietsauswahl für den gesamten Text gilt und solange die Einteilung in Fachgebiete immer gerade so präzise vorgenommen werden kann, daß innerhalb eines Fachgebiets nur noch ein Übersetzungsäquivalent übrigbleibt.

Horizontale Aufteilung des Wortschatzes in Stilebenen

Die horizontale Aufteilung des Wortschatzes in Stilebenen funktioniert im Prinzip wie die Fachgebietseinteilung. Sie ist jedoch weit weniger ergiebig. Sie ist zwar in der Lage, einige stilistisch markierte Übersetzungsäquivalente in Texten auszuschließen, die die entsprechende Stilmarkierung nicht tragen. Jene Textsorten, für die maschinelle Übersetzung sich lohnt, liegen jedoch durchweg auf einer neutralen Stilebene, in die auch die übergroße Mehrzahl der fachsprachlichen

Übersetzungsäquivalente fällt, so daß die Unterscheidungskraft dieses Instruments gering bleibt.

Semantik

Alle bisher dargestellten Entscheidungskriterien lassen sich leicht formalisieren: der Anwender gibt die Fachgebietsangabe xyz vor, im Wörterbuch steht bei einigen Übersetzungsäquivalenten die Fachgebietsangabe xyz und bei anderen ABC, DEF oder GHI. Ein einfacher Algorithmus kann diese Angaben vergleichen und das Äquivalent mit der Angabe xyz auswählen. Das System braucht nicht zu "wissen", was ein Fachgebiet ist oder warum das eine Äquivalent dazugehört und das andere nicht. Es vergleicht einfach Zeichenketten wie xyz und ABC.

Wenn die beschriebenen Kriterien ausgeschöpft sind, ohne daß die Auswahl Aufgabe der lexikalischen Übertragung eindeutig entschieden ist, können die Entwickler maschineller Übersetzungssysteme als nächstes versuchen, dasjenige Wissen nutzbar zu machen, das – soweit man weiß – Menschen einsetzen, wenn sie übersetzen, nämlich Wissen über die Bedeutung des aktuellen Textes.

Wenn der Auswahlentscheidung der lexikalischen Übertragung Wissen über Bedeutung zugrundegelegt werden soll, muß es in ähnlich formaler Weise notiert und instrumentalisiert werden können, wie die Fachgebietsangaben im angeführten Beispiel. Ein Algorithmus, der selbst nicht denkt und nichts versteht, muß in der Lage sein, mit rein formalen Mitteln einen vorgegebenen Bestand semantischen Wissens zu konsultieren und anhand dessen die vorliegende Auswahl Aufgabe zu entscheiden.

Im Gegensatz zu syntaktischen und morphologischen Regeln ist semantisches Wissen jedoch nicht präzise, sondern vage und unscharf. Die Bearbeitung unscharfen Wissens ist sehr viel komplexer als die syntaktischer Regeln. Die maschinelle Übersetzung steht im Grunde seit Anfang der 50er Jahre vor der damals erkannten semantischen Barriere. Melby (1995: 151) hat kürzlich in einem leidenschaftlichen Plädoyer darauf hingewiesen, daß es in der maschinellen Übersetzung keinen Fortschritt gibt, solange die semantische Barriere nicht überwunden wird, und dies gilt ebenso für alle anderen Bereiche der Sprachtechnologie, in denen bedeutungsgestützte oder intelligente Arbeitsgänge erforderlich sind wie etwa in der Spracherkennung, dem bedeutungs- oder begriffsgesteuerten Information Retrieval, der Relevanzerkennung und Informationswege-

lenkung, der Textzusammenfassung und anderen innovativen Entwicklungsgebieten.

Lösungen sind im wesentlichen in zwei Richtungen zu suchen. Entweder abstrahiert man die Bedeutung so lange, bis sie sich in exakte Regeln fassen läßt, oder man entwickelt Methoden, das Unschärfe mit exakten Mechanismen zu erfassen. Vor allem letzterer Ansatz macht das aus, was heute Künstliche Intelligenz genannt wird.

Der erste Ansatz, das Exaktmachen der Bedeutung, liegt in der Verlängerung der bisher in der maschinellen Übersetzung üblichen Methoden, mit denen etwa syntaktische Regelmäßigkeiten erfaßt werden. Hierher gehörten Methoden, die mit semantischen Merkmalen, semantischen Rollen, Tiefenkasus, Theta-Rollen oder ähnlichem arbeiten. Hierbei wird im allgemeinen versucht, die Kombinierbarkeit von Wörtern über das hinaus, was die Syntax zuläßt, semantisch zu spezifizieren. Dabei werden beispielsweise semantische Klassen angegeben, die bei einem Verb wie *drive* die syntaktische Funktion des Subjekts, des direkten Objekts usw. ausüben können, wobei dann Wörter, die für diese Funktionen in Frage kommen (im Beispiel v.a. Substantive) im Wörterbuch mit einer entsprechenden Klassenzuordnung versehen sein müssen. Technisch funktioniert ein solcher Mechanismus genauso, wie der oben für die Fachgebietsmarkierung xyz beschriebene. Inhaltlich zeigt sich jedoch bald, daß die Zuordnung zu semantischen Klassen wie BELEBT, MENSCHLICH, ABSTRAKT, KONKRET USW. einerseits nicht immer objektiv eindeutig ist (d.h. von verschiedenen Lexikografen nicht immer intuitiv gleich entschieden wird) und andererseits oft nicht präzise genug ist, um die erforderlichen Entscheidungen so genau fällen zu können, daß am Ende genau ein Übersetzungsäquivalent ausgewählt und nicht lediglich die Zahl der möglichen Alternativen reduziert wird.

Der zweite Ansatz, der in die Künstliche Intelligenz führt, geht den umgekehrten Weg, indem er versucht, der Bedeutung ihren unscharfen Charakter zu belassen. Auch dieser Ansatz ist jedoch durch die Natur des Computers gezwungen, das Unschärfe mit exakten Algorithmen zu bearbeiten.

Wesentliches Merkmal der meisten Varianten dieses Ansatzes ist es, daß kein Versuch unternommen wird, Bedeutung exakter wiederzugeben als dies in der Sprache normalerweise der Fall ist. Die Bedeutung wird implizit gehalten (Schubert 1988b). Dies hat unter anderem zur Folge, daß es nicht wie in der Syntax möglich ist, mit Regeln zu arbeiten, die den Sachverhalt, für den sie gelten sol-

len, vollständig beschreiben. Man bemüht sich beispielsweise nicht mehr, vollständig zu beschreiben, welche semantischen Eigenschaften Subjekte des Verbs *drive* haben müssen. Man gibt vielmehr Beispiele für solche Subjekte an. Dies öffnet den Weg in die beispieldgestützte Übersetzung. Sie ist für die Systementwickler auch dadurch attraktiv, daß Beispiele ein leicht und billig verfügbares Rohmaterial sind. Beispiele für die Verwendung und Kombinierbarkeit von Wörtern sind Texte, die in Korpora gesammelt und als Wissensquelle für einen semantischen Entscheidungsmechanismus aufbereitet werden können.

Es gibt eine Reihe von Modellen, wie entsprechende Korpora beschaffen sein müssen (einsprachig, zweisprachig in parallelen Übersetzungen, zweisprachig ohne direkte Übersetzungsentsprechung...) und wie sie aufbereitet werden (linearer Text, getaggt, d.h. mit Wortartangaben ausgezeichnet, geparster, d.h. mit syntaktischen Funktionsangaben ausgezeichnet, semantisches Netz...). Es ist überdies möglich, den Entscheidungsmechanismus in unterschiedlicher Art und Weise auf diese Wissensquelle zuzugreifen zu lassen, z.B. durch statistischen Frequenzabgleich oder durch linguistischen Abgleich.

Das gemeinsame Kennzeichen aller beispieldgestützten Techniken ist dabei die Tatsache, daß Beispiele (im Gegensatz zu Regeln) zufällig zusammengetragen sind und dadurch keine Gewähr dafür bieten, daß der im aktuellen Satz zu entscheidende Fall wörtlich in den Beispielen vorkommt. Die Auswahl Aufgabe muß daher anhand von Beispielen entschieden werden können, die der Aufgabe nicht genau entsprechen, sondern ihr ähneln. Aus diesem Grund ist ein Mechanismus zur Messung semantischer Ähnlichkeit erforderlich. Hierfür gibt es sehr unterschiedliche

Ansätze. Victor Sadler hat einen sehr eleganten Mechanismus entwickelt und beschrieben, der diese Funktion erfüllt. Seine Arbeiten umfassen sowohl einen einsprachigen, ausschließlich im Esperanto arbeitenden Mechanismus (Sadler 1989: 53-67), als auch einen verbesserten, der sich auf parallele Korpora in einer Ausgangs- oder Zielsprache und Esperanto stützt (Sadler 1989: Kap. 6-8).

Ein schwerwiegendes Problem aller Formen des Arbeitens mit unscharfem Wissen ist die Tatsache, daß die Bearbeitung unscharfer Information mit (notgedrungen) exakten Algorithmen außerordentlich aufwendig ist. Für jede einzelne Entscheidung der lexikalischen Übertragung müssen enorm viele Datenbankzugriffe und Vergleichsoperationen ausgeführt werden.

Weltwissen

Hauptproblem der experimentellen semantischen Entscheidungssysteme ist neben ihrer rechenintensiven Arbeitsweise die Tatsache, daß beispieldgestützten Ansätzen nur solches Wissen zugänglich ist, das implizit in den verfügbaren Beispielen enthalten ist. Einem so ausgerüsteten System mit einem geeigneten Korpus fehlt jedoch möglicherweise der Zugang zu dem, was für Übersetzer ihr fachliches Hintergrundwissen ist. Übersetzerische Auswahlentscheidungen, zu denen die der lexikalischen Übertragung gehören, gründen sich oft auf inhaltliches Verständnis des in einem Text dargestellten Sachverhalts, das über das sprachliche Verständnis des zu übersetzenden Textes hinausgeht.

In der maschinellen Übersetzung und anderen Bereichen, die sich der Künstlichen Intelligenz bedienen, ist es daher sinnvoll, auch Hintergrundwissen, das oft als Weltwissen bezeichnet wird, miteinzubeziehen.

4. Praktische Ergebnisse

Die bisherigen Abschnitte geben einen Einblick in den maschinellen Übersetzungsvorgang und die im Verlauf dieses Vorgangs zu lösenden Aufgaben. Bevor ich vor diesem Hintergrund in Abschnitt 5. der Frage nachgehe, inwieweit intelligente Bearbei-

tungsschritte überhaupt automatisierbar sind, sollen hier praktische Übersetzungsergebnisse anhand eines Beispieltextes gezeigt werden.

Zuvor ist jedoch ein warnendes Wort zum Thema Systemevaluierung erforderlich.

4.1. Sprachtechnologische Systemevaluierung

Qualitätsprüfung bei sprachtechnologischen Systemen ist ein spezialisiertes Fach, mit dem man sich nicht ohne die erforderliche Sachkenntnis beschäftigen sollte. Gegenüberstellungen maschinell übersetzter Texte, wie ich sie hier gebe, verleiten oft zu

pauschalen Aburteilungen (z.B. durch Zimmer 1996), die einer genaueren fachlichen Analyse nicht standhalten und denen auch die praktischen Erfahrungen industrieller Anwender entgegenstehen (vgl. Kingscott 1996). Es sei daher hier angemerkt, daß

der nachfolgende Vergleich vom Umfang und von der Anlage her keinen Anspruch darauf erhebt, als Systemevaluierung zu gelten. Es ist weder ein Übersetzungstest mit systematisch zusammengestellten Testsätzen für ein Black-Box-Verfahren, noch ist der verwendete Beispieltext geeignet, als repräsentatives Zufallskorpus zu gelten. Ausführlichere Systemevaluierungen geben zwei Marktstudien der Londoner Firma Ovum Ltd. (Lockwood/Leston/Lachal 1995, Mason/Rinsche 1995). Was die Beurteilung maschinell übersetzter Texte so schwierig macht, ist unter anderem die Tatsache, daß maschinelle Übersetzung nicht vollautomatisch verläuft. In jedem Fall sind Arbeitsgänge erforderlich, die der Mensch auszuführen hat, wodurch die "reine" Leistung des Systems nicht immer erkennbar ist. Der wesentlichste Eingriff des Anwenders liegt in der Erstellung fehlender Wörterbucheinträge. Wegen des praktisch unendlichen Umfangs des Wortschatzes gehört zu jedem Übersetzungsauftrag an ein maschinelles Übersetzungssystem auch ein Wörterbuchergänzungsarbeitsgang. Hierbei werden nicht nur Wörter und ihre Übersetzungsäquivalente eingetragen, sondern auch grammatische und andere Merkmale gesetzt,

die dem System die Verwendung des neuen Eintrags in der lexikalen und der Strukturübertragung ermöglichen. Diese lexikografische Tätigkeit läßt dem Anwender Raum für unterschiedliche Entscheidungen und für Fehler. Wenn ein System wie Metal, das für professionellen Mehrbenutzerbetrieb gedacht ist, eine Zeitlang in Betrieb gewesen ist, ist praktisch ein individuelles System entstanden, das sich in seinen Wörterbüchern von der gelieferten Version unterscheidet. Diese individuelle, fachgebietspezifische Wörterbucharbeit trägt ganz wesentlich zur Qualität der Übersetzungsergebnisse bei. Hinzu kommt, daß kein seriöser Hersteller vollautomatische Qualitätsübersetzung verspricht. Die Systeme liefern Rohübersetzungen, die manuell korrigiert werden müssen. Es sei erwähnt, daß es durchaus auch einen von der Öffentlichkeit oft unbemerkten Markt für Rohübersetzungen gibt, bei dem es dem Empfänger nicht um sprachlich korrekte, druckreife Übersetzungen, sondern lediglich um Verständnishilfen beim Querlesen umfangreichen fremdsprachlichen Informationsmaterials geht, das eventuell online empfangen wird (vgl. Lockwood/Leston/Lachal 1995: 143).

4.2. Übersetzungsvergleich

Ich vergleiche im folgenden das Sprachenpaar Englisch-Deutsch zweier Systeme, die hier für zwei Typen maschineller Übersetzungssysteme stehen sollen. Das eine ist das System Power Translator Professional der Firma Globalink Inc. (USA). Es ist seit etwa 1989 auf dem Markt (früher unter dem Namen Globalink) und verkauft sich sehr gut (Mason/Rinsche 1995: 192). Das System läuft auf PC unter dem Betriebssystem DOS mit der grafischen Betriebssystemerweiterung Windows. Der Preis liegt bei 700 DM für ein Sprachenpaar mit Grundwörterbüchern (vgl. auch Jessen 1996). Das System steht auch im Internet zur entfernten Auftragserteilung zur Verfügung (Berberich 1996).

Das andere System ist Metal, Version 3.0.3. Es ist eine Entwicklung des Siemens-Konzerns, die auf Forschungsarbeiten der Universität Texas in Austin (USA) zurückgeht (vgl. Hutchins 1988: 26, Schneider 1995, Whitelock/Kilby 1995: 171-200). Es ist seit 1989 auf dem Markt; der Vertrieb wurde 1995 zu Gunsten einer Neuentwicklung eingestellt. Metal läuft auf Arbeitsplatzrechnern des Herstellers Sun unter dem Betriebssystem Unix und kann im Rechnernetz Client-Arbeitsplätze unter Unix oder Windows bedienen. Der Preis für die Kernsoftware und ein Sprachenpaar lag bei 56.000 DM. Der Siemens-Konzern hat jetzt der Firma Gesellschaft für multilinguale Systeme mbH (GMS) in Berlin und München Entwicklungsrechte übertragen, die das System für Personal Computer neu entwickelt. Eine

Einzelplatzfassung für Laien wird zur Zeit (Mai 1996) vom Verlag Langenscheidt unter dem Produktnamen T1 für ca. 300 DM auf den Markt gebracht. Die GMS plant darüber hinaus eine professionelle Einzelplatzversion und eine netzfähige Version. Der Beispieltext entstammt der Einbau- und Gebrauchsanleitung eines Compact-Disk-Laufwerks, das sowohl zu einer Stereoanlage als auch zu einem Computer gehören kann. Es ist ein technischer Text, der viele typische Merkmale dieser Textsorte aufweist: kurze Sätze, z. T. unvollständige Sätze in einer Aufzählung, Syntaxfehler. Der Text ist möglicherweise selbst eine Übersetzung aus einer nicht genannten Ausgangssprache.

Die beiden Übersetzungssysteme liefern die unten gezeigten Übersetzungen. Die Systeme befanden sich bei dem Experiment im Lieferzustand und enthielten also über die beim Erwerb des Systems mitgelieferten Wörterbücher hinaus keine vom Anwender erstellten Wörterbücher. Allerdings wurden die für diesen Text erforderlichen Wörterbuchergänzungen vorgenommen, wie dies zu jedem maschinellen Übersetzungsarbeitsgang dazugehört. Wie in 4.1. erwähnt, ist zu bedenken, daß es sich um Rohübersetzungen handelt, die noch der manuellen Nachkorrektur bedürfen. Zum Vergleich stelle ich die der Gebrauchsanleitung beigegebene deutsche Fassung daneben, die ebenfalls nicht den Anspruch einer perfekten Übersetzung erheben kann.

	Ausgangstext	Power Translator Professional ³	Metal	Deutsche Fassung des Originals
1	Precautions	Vorsichten	Vorsorgen	Zur besonderen Beachtung
2	Installation	Einbau	Installation	Zum Einbau
3	Avoid placing the drive in a location subject to: - high humidity - high temperature - excessive dust - mechanical vibration - direct sunlight	Meiden Sie Stellen des Laufwerkes in einem Ort abhängig von: hohe Feuchtigkeit hohe Temperatur übermäßiger Staub mechanische Vibration direkte Sonne	Vermeiden stellen das Laufwerk in einer Stelle abhängig von hoher Luftfeuchtigkeit hohe Temperatur übermäßig abstauben mechanische Schwingung richtet licht.	Folgende Umgebungsbedingungen sind zu vermeiden: - hohe Feuchtigkeit - hohe Temperatur - übermäßiges Staubvorkommen - Erschütterungen - direkte Sonneneinstrahlung
4	The drive should only be used in the horizontal position.	Das Laufwerk soll in der horizontalen Position nur benutzt werden.	Das Laufwerk sollte in der horizontalen Stellung nur benutzt werden.	Das Laufwerk ist nur für den waagerechten Einbau geeignet.
5	Do not force the power cable.	Zwingen das Stromkabel nicht.	Erzwingen Sie das Netzanschlußkabel nicht.	Beim Anschließen und Abtrennen des Netzkabels keine Gewalt anwenden.
6	It is keyed to protect the drive.	Es wird abgestimmt, um das Laufwerk zu schützen.	Es ist abgestimmt, um das Laufwerk zu schützen.	Die abgeschrägten Kanten im Netzstecker dienen zum Schutz vor falschem Anschluß.
7	Operation	Operation	Betrieb	Zum Betrieb
8	Do not move the drive when it is in use.	Bewegen das Laufwerk nicht, wenn es in Gebrauch ist.	Bewegen Sie das Laufwerk nicht, wenn es in Anspruch ist.	Das Laufwerk während des Betriebs nicht bewegen.
9	Doing so may cause data error and damage the optical pick-up.	Tun so mag Daten Fehler verursachen und mag @+der optisch Anziehen schädigen.	So zu machen kann den Datenfehler verursachen und den optischen Lesekopf beschädigen.	Andernfalls sind Datenfehler die Folge, und die optische Abtasteinheit kann beschädigt werden.
10	Avoid exposing the drive to sudden changes in temperature as condensation may form on the lens inside the drive.	Meiden Sie Aussetzen des Laufwerkes zu plötzlichen Veränderungen in Temperatur, als Kondensation auf der Linse Innerhalb @+der Laufwerk bilden mag.	Vermeiden Sie, plötzlichen Änderungen in Temperatur das Laufwerk auszusetzen, wie sich Kondensation auf der Linse innerhalb des Laufwerks bilden kann.	Setzen Sie das Laufwerk keinen plötzlichen Temperaturwechseln aus, weil andernfalls mit Kondensatbildung an der Linse der Abtasteinheit zu rechnen ist.
11	Should the surrounding temperature suddenly rise while the drive is turned on, stop using the drive and leave the power on at least one hour before operating it or turning it off.	Sollen die umgebende Temperatur steigt plötzlich auf, während das Laufwerk angeschaltet ist, sollen Benutzen des Laufwerkes halten und sollen den Strom auf mindestens einer Stunde verlassen vor Operieren von ihm oder Drehen von ihm aus.	Wenn die umgebende Temperatur plötzlich steigen sollte, während das Laufwerk eingeschaltet wird, hören Sie auf, das Laufwerk zu benutzen und hinterlassen Sie der Erstein-schaltung mindestens eine Stunde, bevor Sie es bedienen oder es ausschalten.	Wenn die Umgebungstemperatur bei eingeschaltetem Laufwerk plötzlich ansteigt, lassen Sie das Laufwerk mindestens eine Stunde lang unbe-nutzt eingeschaltet, bevor Sie damit arbeiten.
12	Operating the drive immediately after a sudden increase in temperature may result in a malfunction.	Operieren des Laufwerkes sofort nach einer plötzlichen Erhöhung in Temperatur mag in einem Versagen resultieren.	Das Laufwerk sofort bedienen, nachdem eine plötzliche Zunahme um Temperatur in einer Störung resultieren kann.	[Satz nicht übersetzt.]
13	Transportation	Beförderungsmittel	Transport	Zum Transport
14	Close the disc drawer before moving the drive.	Schliessen Sie die Platte-Schublade vor Bewegen des Laufwerkes.	Schließen Sie die CD-Schublade, bevor Sie das Laufwerk bewegen.	Schließen Sie die Disc-Schublade, bevor Sie das Laufwerk transportieren.
15	Keep the original packing materials.	Behalten Sie die ursprünglichen Packen-Materialien.	Führen Sie die ursprünglichen Dichtungsmaterialien.	Bewahren Sie die Originalverpackung zum Transport des Laufwerks auf.
16	When you need to ship the drive to another location, re-packing it in its original container will help you transport it safely.	Wenn Sie das Laufwerk zu einem anderen Ort versenden müssen, neu wird Packen von ihm in seinem ursprünglichen Behälter Ihnen helfen, es heil zu transportieren.	Wann Sie das Laufwerk zu einer anderen Stelle versenden müssen, wird sein ursprünglicher Behälter Ihnen es das Wiederverpacken herein helfen, es sich zu transportieren.	

Zur Auswertung dieser Rohübersetzungen wäre mehr zu bemerken, als in den Rahmen dieses Beitrags paßt. Es mag auf den ersten Blick befremden, daß das professionelle maschinelle Übersetzungssystem Metal dem (trotz seines Namens) eher für Laien geeigneten PC-Produkt Power Translator Professional nicht deutlicher überlegen ist. Selbstverständlich kann kein Urteil anhand eines so kurzen Testtextes abgegeben werden. Es gilt zu bedenken, daß maschinelle Übersetzung sich vor allem für große Mengen gleichartiger und stilistisch konsistenter Texte lohnt, wie sie beispielsweise in der technischen Dokumentation vorkommen. Das Übersetzungssystem muß auf eine neue Textsorte eingestellt werden, indem gründliche Wörterbucharbeit zur Einbindung der für das betreffende Gebiet spezifischen Terminologie geleistet wird (Lukas 1994: 84-96). Dies ist intensive Vorarbeit, die die Bearbeitung der ersten Übersetzungsaufträge aus einer neuen Textsorte langwierig und aufwendig macht, die sich dann aber durch die große Menge zu übersetzender Texte auszahlt.

Da hier beide Systeme im Lieferzustand getestet wurden, war die Textsorte für beide Systeme neu. Neben der Rohübersetzungsqualität, die das System auf Anhieb liefert, ist die entscheidende Qualitätsfrage die nach den Eingriffsmöglichkeiten für den Anwender. Anders gesagt geht es um die Frage, welche der oben gezeigten Übersetzungsfehler in welchem der beiden System behebbar sind. Einige wenige Beispiele mögen hier genügen:

Rechtschreibfehler

Power Translator hat kein *ß* und wandelt jedes vom Anwender eingegebene *ß* in *ss* um. Der Anwender hat hierauf keinen Einfluß.

Parsingfehler

Power Translator übersetzt (5) *Do not force* als *Zwingen ... nicht*. Hierzu kommt es wahrscheinlich dadurch, daß das System *do* als finites Verb parst, obwohl kein Subjekt vorhanden ist und das Verb am Satzanfang steht. Der Anwender hat hierauf keinen Einfluß.

Fehler in der Strukturübertragung

Metal übersetzt (12) *a sudden increase in temperature* als *eine plötzliche Zunahme um Temperatur* statt *eine plötzliche Zunahme der Temperatur* (oder *ein plötzlicher Temperaturanstieg*). Ursache ist der vorgegebene Wörterbucheintrag für *increase* *Zunahme*, der die Präposition *um* festlegt, was in vielen

Fällen, jedoch nicht in diesem, richtig ist. Der Anwender kann hierauf in zweierlei Weise Einfluß nehmen. Einerseits kann er den Mehrwortausdruck *increase in temperature* *Temperaturanstieg* direkt ins Wörterbuch aufnehmen. Andererseits kann er sich bemühen, einen möglichst allgemeinen Eintrag zu schreiben, der nicht nur das aktuelle Übertragungsproblem, sondern gleichzeitig auch eine Reihe ähnlicher Probleme löst, die in künftigen Übersetzungen auftreten könnten. Zu diesem Zweck kann der Anwender parallel zu dem vorhandenen Eintrag mit *um* festlegen, daß bei *increase* eine präpositionale Ergänzung engl. *in* + SUBSTANTIV als dt. SUBSTANTIV + [ÜBERSETZUNG VON *increase*] übersetzt wird, was *Temperatur + Anstieg = Temperaturanstieg* ergibt.

Hier zeigt sich, daß für den beschriebenen Test nur *fehlende* Wörterbucheinträge nachgetragen wurden. Einträge, die zwar vorhanden waren, aber präzisierende Varianten erfordert hätten, wurden zunächst nicht bearbeitet. Der Anwender hat bei Metal in diesem Bereich recht umfangreiche Eingriffsmöglichkeiten.

Fehler in der lexikalen Übertragung

Power Translator vermeidet die Auswahlentscheidungen der lexikalen Übertragung mit Ausnahme der einfachsten syntaktischen Fälle, indem es dem Anwender nicht gestattet, für ein Ausgangssprachliches Wort mehr als ein Übersetzungsäquivalent einzutragen, es sei denn, die Äquivalente gehören zu unterschiedlichen Lesarten des Ausgangswortes, die verschiedenen Wortarten angehören. Für *table* kann z.B. *Tisch* und [*einen Vorschlag*] *einbringen* eingetragen werden, da *Tisch* zur substantivischen Lesart von *table* gehört und [*einen Vorschlag*] *einbringen* zum Verb *to table*; es ist jedoch nicht möglich, neben *Tisch* auch *Tabelle* ins Wörterbuch aufzunehmen. Der Schluß ist zulässig, daß dem Anwender diese Beschränkung auferlegt ist, weil das System Power Translator nicht in der Lage ist, eine Auswahlentscheidung zu treffen, die semantische oder auch nur präzisere syntaktische Kriterien als die Wortartunterscheidung erfordern würde.

Metal hat hier umfangreiche Möglichkeiten, die alle Instrumente der Fachgebietenaufteilung und der Analyse des syntaktischen Kontextes heranziehen. Der Bereich der Semantik bleibt (mit Ausnahme einfacher semantischer Merkmale) jedoch ausgeschlossen. Der Anwender hat bei Metal weitgehenden, bei Power Translator fast keinen Einfluß.

Korrektur unklarer Textstellen

Übersetzer sind sehr häufig gezwungen, unklare oder mißverständliche Textstellen korrigierend und erklärend zu übersetzen, wodurch im Grunde oft die Übersetzung besser ist als das Original. Satz (6) ist ein solcher Fall. Maschinelle Übersetzungs-

systeme sind hier überfordert. Auch zusammenfassende Übersetzungen umständlich formulierter Vorlagen, wie sie die Originalübersetzung in Satz (15) und (16) liefert, sind maschinell nicht möglich. Andererseits vergessen maschinelle Systeme aber auch nicht so leicht einen Satz (vgl. 12).

5. Zur Automatisierbarkeit intelligenter Bearbeitungsschritte

Die bisherige Geschichte der maschinellen Übersetzung zeigt ein Auf und Ab der Hoffnungen und Erwartungen. Zunächst glaubte man, das Ziel der vollautomatischen Qualitätsübersetzung schnell erreichen zu können. Dann mußte man erfahren, daß Übersetzen eine sehr viel komplexere intellektuelle Tätigkeit ist, als man geglaubt hatte, und sich der Automatisierung sehr viel erfolgreicher entzieht als viele hochkomplexe Berechnungs-, Steuerungs- und Regelungsaufgaben der industriellen Fertigung, der Konstruktionsberechnung oder der Hochenergiephysik. Seit einigen Jahren geht der Trend dahin, daß einerseits große und erfolgreiche Entwicklungsprojekte der maschinellen Übersetzung aus Mangel an langfristiger Investitionsbereitschaft abgebrochen werden, während gleichzeitig beträchtliche Millionenbeträge in Vorhaben fließen, die sich nicht erst mit der maschinellen Übersetzung geschriebener Texte aufhalten, sondern es sich zum Ziel setzen, gesprochene direkte und telefonische Gespräche in Echtzeit zu übersetzen (Verbmobil 1991, Kay/Gawron/Norvig 1991, Wahlster/Engelkamp [Hg.] 1992, Kurematsu 1993, Agnäs u.a. 1994*).

Vor diesem Hintergrund stellt sich die Frage, ob Übersetzen überhaupt automatisierbar ist, oder, genauer gesagt, ob sich der heute erreichte Grad der Teilautomatisierung weiter steigern läßt. Es fragt sich weiter, ob diese Entwicklung irgendwann an einer Obergrenze der Automatisierbarkeit Halt machen wird oder ob die vollautomatische Qualitätsübersetzung vielleicht doch nicht grundsätzlich unmöglich, sondern nur bisher nicht erreicht ist.

Melby (1995: 151) warnt dringend davor, sich von der Fortentwicklung der maschinellen Übersetzung irgendeine wesentliche Qualitätssteigerung zu erhoffen, solange nicht die semantische Barriere überwunden ist. Ich habe den Stellenwert von Semantik und Weltwissen oben (Abschnitt 3.2.) im wesentlichen anhand der lexikalischen Übertragung gezeigt. Es wäre hinzuzufügen, daß jene Bearbeitungsgänge, die bisher schon befriedigend funktionieren (Parsing, Strukturübertragung, syntaktische Synthese usw.), zwar noch für gewisse linguistische und technische Verbesserungen offen sind, daß sie eine wesentliche Qualitätserhöhung jedoch nur dann erfahren können, wenn auch ihnen effektive semantische Mechanismen zur Verfügung stehen,

die weitgehend denen entsprechen, die ich am Beispiel der lexikalischen Übertragung besprochen habe.

Wenn es gelingt, von den in der Forschung bereits erarbeiteten Vorschlägen und Experimentalimplementierungen zu technisch realisierbaren Mechanismen mit der für die anstehenden sprachlichen Auswahlentscheidungen erforderlichen Trennschärfe zu gelangen, dann wird der immer wieder (z.B. von Bernhard 1994) angemahnte spürbare Fortschritt in der Übersetzungsqualität möglich. Eine solche semantische Maschine könnte ein treffsichereres Parsing und damit eine auf sichererer Grundlage stehende Strukturübertragung ermöglichen. Sie könnte die noch sehr in den Anfängen steckende Analyse und Übertragung satzübergreifender Textkohärenzphänomene voranbringen, und sie könnte vor allem eine kontextsichere lexikale Übertragung ermöglichen. Die vollautomatische Qualitätsübersetzung ist jedoch auch mit einer solchen semantischen Maschine nicht zu erreichen. Auch für stilistisch monotone technische Texte ohne literarische Anspielungen, Wortspiele oder poetische Höhenflüge gilt die grundlegende Einsicht, daß die Sprache sich dynamisch weiterentwickelt, daß neue Wörter entstehen und – für die maschinelle Übersetzung fataler – alte in leicht veränderter Weise verwendet werden. Es gilt weiter, daß jeder Text, auch Sachtexte, Fachtexte und gründlichst auf Laiensicherheit hin redigierte technische Dokumentation, in jedem Augenblick auf Hintergrundwissen Bezug nehmen kann, kulturspezifische Selbstverständlichkeiten voraussetzen oder Wörter aus der Fachin die Gemeinsprache verschieben kann und umgekehrt. Hier hat die Automatisierbarkeit eine Grenze – auch bei dynamischer Wissensakquisition für die Wissensquellen der semantischen Maschine. Die Ursache für diese inhärente Unschärfe und die dynamische Veränderlichkeit der Sprache ist in der Grundgegebenheit zu suchen, daß ein prinzipiell überschaubares Repertoire lexikaler und struktureller sprachlicher Instrumente einem unbegrenzten und in sich unstrukturierten Kontinuum des Ausdrückbaren gegenübersteht. Auch diese Einsicht steckt in Wilhelm von Humboldts berühmter Aussage, die Sprache müsse "von endlichen Mitteln unendlichen Gebrauch machen" (Humboldt 1836: 106).

6. Esperanto und die maschinelle Übersetzung

Als Anfang der 50er Jahre klar wurde, daß die ersten Versuche mit der maschinellen Übersetzung nicht zu jenem schnellen und überwältigenden Erfolg wurden, den man erhofft hatte, hat Yehoshua Bar-Hillel, der als der erste Wissenschaftler bekannt geworden ist, der eigens für Forschung an der maschinellen Übersetzung angestellt wurde, den Gedanken geäußert, wenn es mit Volkssprachen

nicht funktioniere, müsse man es eben zunächst mit etwas einfacherem versuchen. Bar-Hillel (1951: 235) nannte hier neben vereinfachten Sprachen wie Basic English explizit die Plansprachen Esperanto und Interlingua. Es ist interessant zu verfolgen, welchen Weg das Esperanto seither in der Sprachtechnologie genommen hat.

6.1. Ideen und Projekte

Der Weg des Esperanto in die Sprachtechnologie verläuft in drei Etappen. Zunächst kommt die Idee, dann ein gleichberechtigter Einsatz und zuletzt eine bewußte Entscheidung für Esperanto wegen seiner Besonderheiten (vgl. Schubert 1989: 26-28).

In der ersten Etappe ist zu beobachten, daß, wie bei Bar-Hillel, hier und da die Idee auftaucht, eine Plansprache könne in der maschinellen Übersetzung eine sinnvolle Verwendung finden (seltener in anderen sprachtechnologischen Anwendungen, da diese sich erst nach der maschinellen Übersetzung entwickelt haben).

Als Vorläufer gehören in diesen Zusammenhang auch die Arbeiten des sowjetischen Erfinders Petr Trojanskij (später Smimov-Trojanskij), der in den 30er Jahren, bevor es Computer gab, eine Übersetzungsmaschine beschrieben hat, für die er 1933 ein Patent erhielt. Kernstück war ein Algorithmus mit einer "logische Syntax" genannten Repräsentation, die Trojanskij mit Hilfe von Funktionsmorphemen des Esperanto ausdrückte (Beschreibung 1933: Trojanskij 1959). In diesem Punkt ähnelt Trojanskij's System der ebenfalls in den 30er Jahren entstandenen Dependenzsyntax Lucien Tesnière's (1959/1982: 64), der Esperanto-Wortartendungen als universelle syntaktische Funktionsetiketten verwendet. Trojanskij soll 1941 eine experimentelle mechanische Übersetzungsmaschine in Betrieb gehabt haben, und 1948 ist von einem Projekt für eine elektrische Maschine die Rede (Beschreibung 1947: Smimov-Trojanskij 1959; vgl. Hutchins 1986: 23). Als in der Sowjetunion die maschinelle Übersetzung im eigentlichen Sinne in Gang kam, wurden Smimov-Trojanskij's Arbeiten gründlich studiert (Denisov 1965: 80-83).

Die zweite Etappe bilden konkreter ausgearbeitete Algorithmen und experimentelle Implementierungen. Hier wird Esperanto im Prinzip genauso behandelt wie alle anderen Sprachen, mit denen die betreffenden Wissenschaftler arbeiten. Es wird zwar natürlich darauf hingewiesen, daß Esperanto sich durch seine regelmäßige Grammatik für viele der Bearbeitungsschritte besser eignet als Volks-

sprachen, aber in seiner Funktion steht Esperanto hier neben Sprachen wie Deutsch, Englisch und Russisch. (Eine praktische Anwendung einer anderen Plansprache als Esperanto in der Sprachtechnologie ist nicht bekannt.) Soweit sich heute feststellen läßt, sind manche Veröffentlichungen reine Planstudien (Sjögren 1970, Kelly 1978, Dietze 1986), während andere kleinere Programme enthalten (Ben-Avi 1977). Etwas später folgen größere Projekte, bei denen Esperanto neben Volkssprachen bearbeitet wird, so im maschinellen Übersetzungsprojekt SUSY der Universität des Saarlandes (Maas 1985), in der Sprachverarbeitung, z.T. mit Anwendungen im rechnergestützten Unterricht (Sherwood/Sherwood 1982, Janot-Giorgetti 1985, Sherwood 1985a,b, Koutny/Olaszy/Kisfaludy 1988, Koutny/Olaszy 1994), bei Projekten, die maschinelle Übersetzung und Sprachverarbeitung miteinander verbinden (Sato/Kasuya 1987, Sato 1989), sowie bei einer Reihe anderer sprachtechnologischer Anwendungen (z.B. Stopoloni 1982). Seit Personal Computer Gemeingut geworden sind, wird die Zahl der privaten und halbprivaten Entwicklungen von maschinellen Übersetzungssystemen und anderen sprachtechnologischen Anwendungen immer schwerer überschaubar. Auch hierunter findet sich eine wachsende Anzahl von Systemen, die Esperanto als Ausgangs- oder Zielsprache einbeziehen, zum Beispiel de Kat (1985), Mohai (1986), Li (1986). Es gibt zahlreiche neuere, die nicht veröffentlicht sind.

Die dritte Etappe der Verwendung des Esperanto in der Sprachtechnologie beginnt, als es für eine Funktion ausgewählt wird, die nur das Esperanto wegen seiner besonderen Eigenschaften als Plansprache erfüllen kann. Eine derartige besondere Rolle ist die einer Zwischensprache in mehrsprachigen maschinellen Übersetzungssystemen. Unabhängig voneinander wählen in den 80er Jahren der Physiker Hiroshi Katsumori und der Softwareingenieur Toon Witkam Esperanto als Zwischensprache.

Katsumori's Projekt ist eine experimentelle universitäre Entwicklung eines Übersetzungs-

systems für die Sprachen Englisch und Japanisch (Katsumori/Fukuda 1984, Makino/Hirata/ Katsumori 1986).

Das Projekt Distributed Language Translation (DLT) des niederländischen Softwarehauses Buro voor Systeemontwikkeling (BSO) in Utrecht ist größer angelegt. Nach einer Vorstudie 1982-1983 (Witkam 1983) führt die Firma von 1984 bis 1990 ein Forschungs- und Entwicklungsprojekt im Umfang von ca. 50 Mitarbeiterjahren durch. Ziel ist ein Prototyp eines mehrsprachigen maschinellen Übersetzungssystems. Die Mitte der 80er Jahre sehr innovativen Merkmale des Systems sind verteilter Betrieb im weltweiten Rechnernetz, problemlose Erweiterbarkeit um neue Ausgangs- und Zielsprachen, Verzicht auf Nachkorrektur (aber nicht auf Mitarbeit des Menschen während des Analysevorgangs) und Ausrichtung auf Anwender ohne Kenntnis der Zielsprache.

Das Projekt DLT hatte das Ziel, einen Prototypen eines maschinellen Übersetzungssystems mit diesen Eigenschaften zu entwickeln. Es hatte nicht primär das Ziel, die Einsetzbarkeit des Esperanto in der maschinellen Übersetzung zu testen. Es verdient Beachtung, daß sich die Entscheidung für Esperanto als Zwischensprache als konsequente Schlußfolgerung aus den aufgestellten Systemanforderungen ergab. Erst nachdem dieser Schluß gezogen worden war, wurde Esperanto eingesetzt – wodurch es dann zwangsläufig einem sehr aufschlußreichen Praxistest unterzogen wurde.

Die Entscheidung für Esperanto als Zwischensprache ergibt sich aus folgenden Überlegungen. Um der Anforderung des verteilten Betriebs im weltweiten Rechnernetz⁵ zu genügen, ist DLT so angelegt, daß der Übersetzungsvorgang in zwei Phasen erfolgt, die an getrennten Orten und zu unterschiedlichen Zeitpunkten ablaufen können. In der ersten Phase wird aus der Ausgangssprache in die Zwischensprache übersetzt. Die zwischensprachliche Fassung des Textes ist die Transportform, in der der Text über das Rechnernetz übermittelt wird. Die zweite Phase erfolgt dann beim Empfänger des Textes, wo aus der Zwischensprache in die Zielsprache weiterübersetzt wird. Der Systementwurf sieht vor, daß bei Absendung des Textes nicht bekannt zu sein braucht, wer den Text wann und in welcher Zielsprache lesen wird. Die Zwischensprachfassung muß also die Weiterübersetzung in jede beliebige Zielsprache erlauben. Hieraus (und aus dem Gebot der Vermeidung kombinatorischer Vielfalt, vgl. 2.3. oben) ergibt sich die Forderung nach einer autonomen, von Ausgangs- und Zielsprachen unabhängigen Zwischensprache.

Da vollautomatische Qualitätsübersetzung unmöglich ist, sieht der Entwurf von DLT einen interaktiven Disambiguierungsdialo g zwischen Anwender und System vor. Der räumlich und zeitlich verteilte Betrieb im weltweiten Netz schließt aus,

daß diese Form des Eingreifens in der zweiten Phase des Übersetzungsvorgangs erfolgt. Der interaktive Dialog ist bei DLT daher in die erste Phase, die Übersetzung aus der Ausgangssprache in die Zwischensprache, verlegt. Hierdurch wird überdies noch ein weiteres Problem gelöst: Indem die Interaktion bei der Analyse der Ausgangssprache erfolgt, wird es möglich, das System für einsprachige Anwender ohne Kenntnis der Zielsprachen auszuliegen.

Aus der Verlagerung der Mitarbeit des Anwenders in die erste Phase des Übersetzungsvorgangs ergibt sich für den Entwurf des Systems DLT eine entscheidende Konsequenz: Bei der zweiten Phase, der Weiterübersetzung aus der Zwischensprache in die Zielsprache, muß eine vollautomatische Qualitätsübersetzung erfolgen. Die Erfahrung der 50er Jahre, daß vollautomatische Qualitätsübersetzung unmöglich ist, gilt zwar nach wie vor uneingeschränkt, doch ist daneben eine Reihe beachtenswerter Erfolge getreten, die immer dann erzielt werden konnten, wenn die Möglichkeit bestand, die Ausgangstexte übersetzungsgünstiger zu gestalten, als normale, frei formulierte Texte dies sind. Eine relativ korrekturfreie, vollautomatisch erstellte Übersetzung gelingt bisher nur bei planmäßig vereinfachten Dokumentationssprachen (Arnold u.a. 1994: 156) und sehr eng begrenzten Fachsprachen (Arnold u.a. 1994: 159). Für die Zwischensprache von DLT mußte eine ähnliche übersetzungsfördernde Lösung gefunden werden, wobei das Problem hier jedoch komplizierter war als bei den genannten Fällen erfolgreicher vollautomatischer Qualitätsübersetzung. Während dort die Ausgangstexte in ihrer sprachlichen Komplexität reduziert waren, hatten die Entwickler von DLT lediglich Zugriff auf die Zwischensprache. Der Systementwurf ging davon aus, daß als Ausgangstexte zwar keine Texte in Frage kommen sollten, bei denen sprachliche Stilfragen eine Rolle spielen, also keine schöne Literatur, keine Werbesprache u.ä., daß aber Sachtexte aller Art übersetzbar sein sollten.

DLT brauchte daher eine Zwischensprache, die (zumindest auf der Ebene von Sachtexten) über die volle Ausdruckskraft einer menschlichen Sprache verfügt, die gleichzeitig aber in einer solchen Weise klarer ist, daß vollautomatische Weiterübersetzung bei hoher Übersetzungsqualität möglich ist. Die Bedingung der übersetzungsfördernden Klarheit würde eine künstliche Zwischenrepräsentation nahelegen, die Bedingung der vollen Ausdruckskraft verlangt dagegen eine menschliche Sprache. Eine künstliche Zwischenrepräsentation bringt überdies das Problem der intersubjektiven Bedeutungsfestlegung mit sich, auf das wiederholt hingewiesen wurde (Schubert 1988c: 143, Wilks nach Melby 1995: 143). Für DLT lautete die Lösung Esperanto (vgl. Schubert 1988c, 1992b).

Zwei Fragen, die in diesem Zusammenhang vielfach

gestellt werden, sind hier zu beantworten: Esperanto wird oft als künstliche Sprache bezeichnet – gelten dann nicht dieselben Vorbehalte wie gegen künstliche Zwischenrepräsentationen? Esperanto wird oft als menschliche Sprache wie jede andere bezeichnet – könnte also ebensogut Deutsch oder

Englisch Zwischensprache eines Systems wie DLT sein?

Diese beiden Fragen führen zu grundsätzlichen Überlegungen über die Natur des Esperanto, auf die ich im folgenden kurz eingehen möchte.

6.2. Esperanto: Plansprache im Übergang

Die Antworten auf die aus der Erörterung der Eigenschaften der im maschinellen Übersetzungssystem DLT erforderlichen Zwischensprache heraus aufgeworfenen Fragen – deren Voraussetzungen einander widersprechen, so daß höchstens eine bejaht werden könnte – sind für die Funktionsfähigkeit der Sprache als Zwischensprache der maschinellen Übersetzung von direktem Belang. Die Antworten lauten nach meiner Überzeugung wie folgt.

Als Esperanto 1887 veröffentlicht wurde, war es in der Tat keine Sprache im eigentlichen Sinne, sondern ein Sprachprojekt. Es stand künstlichen Symbolsystemen nahe. Damals lagen wie bei künstlichen Zwischenrepräsentationen die Bedeutungen der strukturellen und lexikalischen Elemente des Esperanto nicht intersubjektiv fest, sondern mußten mit Hilfe von Wörterbüchern, also über Referenzsprachen, ermittelt werden. Im Laufe von mehr als einem Jahrhundert praktischer Anwendung hat sich dies geändert: Die Bedeutung von Konstruktionen, Wörtern und Morphemen wird jetzt, wie in Volkssprachen, durch Konvention, also durch den Gebrauch in der Sprachgemeinschaft festgelegt. Daß es sich um eine Zweitsprachgemeinschaft handelt, ändert hieran nichts. Willkürliche Festlegungen durch Lexikografen, die Akademie oder andere Sprachnormungsinstanzen haben nur einen geringen Einfluß – wie in Volkssprachen auch. Gleichzeitig hat sich Esperanto jedoch eine große strukturelle Klarheit erhalten, die der Funktion als Zwischensprache dienlich ist. Um die Zwischensprache in DLT im strengen Sinne des Parsing in Wortbildung und Syntax eindeutig zu machen, brauchten gegenüber dem von Menschen gesprochenen und geschriebenen Esperanto nur zwei Änderungen eingeführt zu werden: ein Morphemtrennzeichen (das es in der Frühzeit des Esperanto schon einmal gegeben hatte) und ein Syntagmenendzeichen für bestimmte Mehrdeutigkeitsfälle wie Präpositions- und Relativsatzdependenz.

Für eine Zwischensprache wird allerdings bisweilen weitergehend gefordert, sie müsse nicht nur strukturell (morphologisch, syntaktisch usw.), sondern auch semantisch eindeutig sein. Hier ist zu bedenken, daß diese Forderung prinzipiell unerfüllbar ist. Die Bedeutung, das sprachlich Ausdrückbare, ist ein ungliedertes Kontinuum, das nicht aus sprachübergreifend feststellbaren

kleinsten, nicht weiter aufteilbaren Bedeutungseinheiten oder "semantischen Atomen" besteht. Es gibt daher keine semantische Eindeutigkeit, da sich jeder als Bezugsgröße herangezogene Begriff konzeptuell weiter aufspalten läßt. Auch vollkommene semantische Kompositionalität, also die automatische Herleitbarkeit der Bedeutung komplexer sprachlicher Ausdrücke aus der Summe der Bedeutungen ihrer Elemente, ist unerreichbar. Zwar ist Esperanto sehr weitgehend kompositional, doch muß die Sprache zwangsläufig an eine Grenze semantischer Unschärfe stoßen, an der die Idiomatisierung einsetzt. Dies gilt sowohl auf der Satz- und Syntagmenebene (Dasgupta 1993) als auch auf der Wortebene (Schubert 1993).

Was aus der Sicht der maschinellen Übersetzung in der Summe bleibt, ist das Esperanto als eine Zwischensprache, die erst durch den unreflektierten Gebrauch in einer Sprachgemeinschaft über einen längeren Zeitraum hinweg jenen Grad an Autonomie erlangt hat, der ihr die volle, für die maschinelle Übersetzung erforderliche Ausdruckskraft verleiht, die aber gleichzeitig durch ihre planvolle Gestaltung in Syntax, Morphologie und Wortbildung einen Grad an übersetzungsfördernder Klarheit besitzt, die sie in der Funktion der Zwischensprache Volkssprachen überlegen macht.

Esperanto ist eine Sprache, die den Übergang vom Stadium eines künstlichen Symbolsystems zu dem einer menschlichen Sprache erst sehr kürzlich durchlaufen hat. Ihre besonderen, der Sprachtechnologie so willkommenen Eigenschaften besitzt sie aus diesem Grunde. Esperanto ist keine Volkssprache, sondern eine Sprache, deren Sprechergemeinschaft in anarchischer, aber sehr konservativer Weise die Merkmale pflegt und bewahrt, die der internationalen und interkulturellen Verständlichkeit der Sprache dienen – wodurch das Esperanto unwillkürlich auch in sehr besonderer Weise übersetzungsgünstig ist. Gleichzeitig ist Esperanto jedoch auch kein künstliches Symbolsystem mehr, sondern eine Sprache, die autonom geworden ist, in der also die Bedeutung der strukturellen und lexikalischen Elemente durch Konvention festgelegt und aufrechterhalten wird, was nur durch jahrzehntelangen unreflektierten und spontanen Gebrauch der Sprache in einer Sprachgemeinschaft erreicht werden kann.

6.3. Ausblick

Die zuletzt besprochene Rolle des Esperanto als Zwischensprache in einem maschinellen Übersetzungssystem ist auf den ersten Blick die eines unsichtbaren Instruments, das für den Anwender unsichtbar bleibt. Sie kann in der maschinellen Übersetzung (und bei Vorliegen entsprechender Entwurfsbedingungen auch in anderen sprachtechnologischen Systemen, vgl. Schubert 1995) nützlich sein, muß aber nicht schon deswegen auch auf das aktive Interesse der Sprachgemeinschaft treffen. Eine solche Einstellung wäre jedoch kurzsichtig.

Zum einen ist darauf hingewiesen worden (Danzin u.a. 1992), daß den Sprachen der industrialisierten Länder in bezug auf die Sprachtechnologie eine ähnlich schicksalhafte Entwicklung droht, wie sie sich nach der Erfindung der Druckkunst mit beweglichen Lettern vollzogen hat: die Sprachen, die sich das neue Medium zu eigen machen konnten, nahmen an Bedeutung zu, die anderen glitten in die Zweitrangigkeit ab. Im Hinblick auf diesen sogenannten Gutenberg-Effekt sollte jede Sprachgemeinschaft darauf achten, sprachtech-

nologisch Schritt zu halten. Dies gilt auch für die Esperanto-Gemeinschaft.

Zum anderen ergibt sich aus einer Funktion wie der einer Zwischensprache in der maschinellen Übersetzung eine ausbaufähige Position, die sich nutzen läßt. Wenn sprachtechnologische Systeme mit einer Esperanto-Komponente ausgerüstet werden müssen, warum soll man dann nicht auch Esperanto für andere, sichtbare Funktionen wählen? Die EU braucht eine einheitliche Dokumentations-sprache für legislative und administrative Vorschriften und Vorlagen, für technische und patentrechtliche Dokumentation und viele andere Anwendungen, bei denen dann aus einer zentralen Form heraus je nach Bedarf maschinelle Übersetzungen in jede gewünschte Zielsprache angefertigt werden können.

Wenn die Esperanto-Gemeinschaft daran interessiert ist, ihrer Sprache eine internationale Mittler- und Vermittlerposition zu verschaffen, ist die Sprachtechnologie ein berücksichtigenswerter Weg, der neben anderen beschritten werden sollte.

Anmerkungen

- 1 Die deutschen Glossen stehen hier zur Erklärung für den Leser. Der Analysevorgang selbst erfolgt rein einsprachig, im Beispiel also im Englischen. Daher ist es nicht nötig, für verschiedene Lesarten oder Übersetzungen eines Wortes getrennte Einträge in den einsprachigen Wörterbüchern und Regelsystemen vorzusehen, solange sie sich nicht in ihren syntaktischen und morphologischen Merkmalen unterscheiden.
- 2 Ich übernahm die Einträge aus den Wörterbüchern wörtlich, lasse jedoch Informationen weg, die zur lexikalischen Übertragung nicht direkt beitragen, so insbesondere Genusangaben, Querverweise, Mehrwortausdrücke und Komposita, Kollokationen und Redewendungen usw. Angaben des Fachgebiets und der Stilebene, die in den Quellen abgekürzt erscheinen, gebe ich im Klartext wieder.
- 3 Das System Power Translator Professional kennzeichnet mit "@" Pronomen und andere Funktionswörter, deren regierendes Satzglied im Analysevorgang nicht gefunden worden ist, so daß die morphologische Form nicht bestimmt werden konnte.
- 4 Aus diesen und vielen anderen Projekten des maschinellen Dolmetschens gibt es neuere Veröffentlichungen. Oft handelt es sich jedoch, wie im Falle des deutschen Verbundprojekts *Verbomobil*, um so große Vorhaben, daß eine Fülle technischer und sprachtechnologischer Veröffentlichungen erscheint, die jeweils nur einen winzigen Teilaspekt des Gesamtsystems behandeln. Ich versuche hier, auf Veröffentlichungen zu verweisen, die einen Gesamteindruck des angestrebten Systems bieten. Allerdings scheint sich die Zielsetzung des Projekts *Verbomobil* nach den hier genannten Durchführbarkeitsstudien eingreifend gewandelt zu haben, ohne daß mir aktuelle Übersichtsdarstellungen bekannt wären. Die drei Studien *Verbomobil* (1991), *Kay/Gawron/Norvig* (1991) und *Wahlster/Engelkamp* (Hg.) (1992) sind Teil der Projektausschreibungsunterlagen des Bundesministeriums für Forschung und Technologie. Ob sie in zugänglicherer Form veröffentlicht sind, ist mir nicht bekannt.
- 5 Vieles im Entwurf des Systems DLT paßt ausgezeichnet auf die Arbeitsbedingungen, die man heute im Internet vorfindet, das damals jedoch noch nicht als weltweites, öffentliches Netz bestand. Der DLT-Ansatz erweist sich hier im Rückblick als realistisch und vorausschauend.

Literatur

Agnäs, Marie-Susanne u.a. (1994): Spoken Language Translator: First-Year Report. SICS Research Report R94:03 / SRI Cambridge Report CRC-043. Kista: Swedish Institute of Computer Science / Cambridge: SRI International

Arnold, D.; L. Balkan, R. Lee Humphreys, S. Meijer, L. Sadler (1994): Machine Translation. Manchester/ Oxford: NCC Blackwell

Bar-Hillel, Yehoshua (1951): The Present State of Research on Mechanical Translation. *American Documentation* 2: 229-237

Ben-Avi, Stephen (1977): An Investigation into the Use of Esperanto as an Intermediate Language in a Machine Translation Project. Manchester: Victoria University (Diss.)

- Berberich, Frank (1996): Die digitalen Dolmetscher bieten ihre Dienste mittlerweile in Echtzeit an. Computer-Zeitung 18. Januar 1996: 8
- Bernhard, Ursula (1994): Machine Translation: Ten Years On. Where Are the Users? In Clarke/Vella (Hg.) 1994
- Clarke, Douglas; Alfred Vella (Hg.): International Conference: Machine Translation Ten Years On. Cranfield: Cranfield University (vervielfältigte Teilnehmerfassung des Tagungsbandes, Veröffentlichung geplant)
- Danzin, A. u.a. (1992): Vers une infrastructure linguistique européenne. Rapport présenté par A. Danzin et le Groupe de réflexion stratégique pour La Commission des Communautés Européennes (DG XIII). EU-Kommission Dokument 5210/92 FR
- Dasgupta, Probal (1993): Idiomaticity and Esperanto Texts. Linguistics 31: 367-386
- Denisov, Petr Nikitič (1965): Principy modelirovanija jazyka. Moskva: Izdatel'stvo Moskovskogo Universiteta
- Dietze, Joachim (1986): Projekt der rechnergestützten Erarbeitung eines Wörterbuches von Esperantowurzeln. Wissenschaftliche Zeitschrift der Universität Halle 35[G 5]: 90-91
- Ernst, Richard (1985). Wörterbuch der industriellen Technik. Bd. 2: Englisch-Deutsch. 5. Aufl. Wiesbaden: Brandstetter
- Hodges, Andrew (1983): Alan Turing: The Enigma. Neue Aufl. 1992. London: Vintage
- Humboldt, Wilhelm von (1836): Über die Verschiedenheit des menschlichen Sprachbaues. Berlin: Lambert Schneider/Königliche Akademie der Wissenschaften
- Hutchins, W. John (1986): Machine Translation: Past, Present, Future Chichester: Horwood / New York usw.: Wiley
- Hutchins, W. John (1988): Recent Developments in Machine Translation. In Maxwell/Schubert/Witkam (Hg.) 1988: 7-63
- Janot-Giorgetti, Marie-Thérèse (1985): Parol-rekono kun limigita vortaro, ĝia apliko en la lernado de parolataj lingvoj. In Koutny (Hg.) 1985: 57-68
- Jessen, Sönke (1996): Steuerung der Übersetzungsqualität in der rechnergestützten Übersetzung. Untersuchung am System Globalink Power Translator. Flensburg: Fachhochschule Flensburg, Studiengang Technikübersetzen (unveröff. Diplomarbeit)
- Kat, J. O. de (1985): Traduko el la internacia lingvo al naciaj. In Koutny (Hg.) 1985: 259-266
- Katsumori, Hiroshi; M. Fukuda (1984): Esuperanto o tyūkai-gengo to suru kikai-hon'yaku no kokoromi. WGNL Meeting 1984: 7-26. o.O.
- Kay, Martin; Jean Mark Gawron, Peter Norvig (1991): Verbmobil: A Translation System for Face-to-Face Dialog. o.O.
- Kelly, Ian D. K. (1978): How Esperanto Can Aid Machine Translation. Computer Weekly 19. Januar 1978: 14
- Kingscott, Geoffrey (1996): Machine Translation in Action: Metal at Boehringer Ingelheim. Language International 8[2]: 5-8
- Klubkov, Pavel' A. (1984): Iz istorii interlingvističeskoj mysli v Rossii v konce XVIII – načale XIX vv. In A. D. Duličenko (Hg.): Istorija i sovremennoe sostojanie interlingvistiki. Interlinguistica Tartuensis 3. Tartu Riikliku Ülikooli Toimetised // Učenyje zapiski Tartuskogo Gosudarstvennogo Universiteta // Acta et Commentationes Universitatis Tartuensis 671: 115-125
- Koutny, Ilona (Hg.) (1982): Homa lingvo kaj komputilo. Budapest: Neumann János Számítógéptudományi Társaság
- Koutny, Ilona (Hg.) (1985): Perkomputila tekstoprilaboro. Budapest: Scientia Eldona Centro
- Koutny, Ilona; Gábor Olaszy (1994): Fonetika fono kaj realigo de esperantlingva elektronika parolgenero. In Hans Michael Maitzen, Herbert Mayer, Zlatko Tišljar (Hg.): Aktoj. Internacia Scienca Simpozio "Esperanto 100-jara". Vieno: Esperanto-Laborkomitato Aŭstria/Pro Esperanto / Maribor: Inter-Kulturo, 187-202
- Koutny, Ilona; Gábor Olaszy, Katalin Kisfaludy (1988): Esperanto Speech Synthesis and Its Application in Language Learning. In Tamás Szende (Hg.): From Phonology to Applied Phonetics. Magyar Fonetikai Füzetek // Hungarian Papers in Phonetics 19: 47-54
- Kurematsu, Akira (1993): Overview of Speech Translation at ATR. In Sergei Nirenburg (Hg.): Progress in Machine Translation. Amsterdam/Oxford/Washington: IOS Press / Tokyo/Osaka/Kyoto: Ohmsha, 33-41 + 297-318
- Landsbergen, S. P. J. (1987): Isomorphic Grammars and Their Use in the Rosetta Translation System. In Margaret King (Hg.): Machine Translation Today. Edinburgh: Edinburgh University Press, 351-372
- Landsbergen, Jan [S. P. J.] (1994): A Compositional Definition of the Translation Relation. In: Lisette Appelo, Theo Janssen, Franciska de Jong, Jan Landsbergen (Hg.): *Compositional Translation by M. T. Rosetta*, Dordrecht/Boston/London: Kluwer, 15 - 33
- Langenscheidts enzyklopädisches Wörterbuch der englischen und deutschen Sprache (1992). Teil 1. 10. Aufl. Berlin usw.: Langenscheidt
- Li Wei (1986): Aŭtomata tradukado el la internacia en la ĉinan kaj anglan lingvojn: E-Ĉ/A maŝintraduka sistemo. Grundlagenstudien aus Kybernetik und Geisteswissenschaft/Humankybernetik 27: 147-152
- Lockwood, Rose; Jean Leston, Laurent Lachal (1995): Globalisation. Creating New Markets with Translation Technology. London: Ovum
- Lukas, Mirjam (1994): Merkmale der maschinellen Übersetzbarkeit von Fachtexten. Arbeitspapiere der GMD 855. Sankt Augustin: Gesellschaft für Mathematik und Datenverarbeitung / Flensburg: Fachhochschule Flensburg, Studiengang Technikübersetzen (Diplomarbeit)
- Maas, Heinz Dieter (1985). La traduksistemo SUSY kaj ĝiaj perspektivoj. In Koutny (Hg.) 1985: 163-174
- Makino, Satoru; Mitsuru Hirata, Hiroshi Katsumori (1986): Esuperanto o tyūkai-gengo to suru kikai-hon'yaku. WGNL Meeting 1986: 3-28 o.O.
- Mason, Jane; Adriane Rinsche (1995): Translation Technology Products. London: Ovum
- Maxwell, Dan; Klaus Schubert, Toon Witkam (Hg.) (1988): New Directions in Machine Translation. Dordrecht/Providence: Foris
- Melby, Alan K. (mit C. Terry Warner) (1995): The Possibility of Language. Amsterdam/ Philadelphia: Benjamins
- Mohai, Lajos (1986): Komputila tradukprogramo el Esperanto hungaren. In Péter Broczkó, Ilona Koutny, András Lukács (Hg.): Interkybernetik. Budapest: John von Neumann Society for Computing Sciences. Bd. 1: 47-51
- Newman, Patricia E. (1988): Information-only Machine Translation. In Vasconcellos (Hg.) 1988: 178-189
- Ouyang Wendao (1982): Komputila modelo por ĉin-lingva kompreniĝo uzante Esperanton kiel inter-lingvon. In Koutny (Hg.)

1982: 31-38

Perevodnaja mašina P. P. Trojanskogo. (1959). Moskva: Izdatel'stvo Akademii nauk SSSR

Sadler, Victor (1989): Working with Analogical Semantics. Dordrecht/Providence: Foris

Sato, Shigeru (1989): Phonetic Form Generation in a Semantics-to-Speech System of Japanese. Sendai: Tohoku University (Diss.)

Sato, Shigeru; Hideki Kasuya (1987): Automatic Translation / Speech Synthesis of Japanese from Written Esperanto Incorporating a Linguistic Knowledge Base Editor. In J. Laver, M. A. Jack (Hg.): European Conference on Speech Technology. Edinburgh: CEP, Bd. 2: 414-417

Schmid, Annemarie (1992): MÜ: Übersetzungsergebnisse im Vergleich. *Lebende Sprachen* 37[2]: 51-55

Schneider, Thomas (1995): The METAL System. Status 1995. In *MT Summit V. Proceedings* [Luxemburg]: EU-Kommission (vervielfältigte Teilnehmersausgabe)

Schubert, Klaus (1987): *Metataxis*. Dordrecht/Providence: Foris

Schubert, Klaus (1988a): The Architecture of DLT – Interlingual or Double Direct? In Maxwell/Schubert/Witkam (Hg.) 1988: 131-144

Schubert, Klaus (1988b): Implicitness as a Guiding Principle in Machine Translation. In Denés Vargha (Hg.): *Coling Budapest*. Budapest: John von Neumann Society for Computing Sciences, 599-601

Schubert, Klaus (1988c): Ausdruckskraft und Regelmäßigkeit: Was Esperanto für automatische Übersetzung geeignet macht. *Language Problems and Language Planning* 12: 130-147

Schubert, Klaus (1989): Interlinguistics – Its Aims, its Achievements, and Its Place in Language Science. In Klaus Schubert (mit Dan Maxwell) (Hg.): *Interlinguistics*. Berlin/New York: Mouton de Gruyter, 7-44

Schubert, Klaus (1992a): Diderichsen, Tesnière och datorn. Några textgrammatiska idéer om maskinöversättning. In Jonna Louis-Jensen, Jóhan Hendrik W. Poulsen (Hg.): *The Nordic Languages and Modern Linguistics 7*. Tórshavn: Fróðskaparsetur Føroya, Bd. 2: 565-577

Schubert, Klaus (1992b): Esperanto as an Intermediate Language for Machine Translation. In John Newton (Hg.): *Computers in Translation*. London/New York: Routledge, 78-95

Schubert, Klaus (1993): Semantic Compositionality: Esperanto Word Formation for Language Technology. *Linguistics* 31: 311-365

Schubert, Klaus (1995): Parameters for the Design of an Intermediate Language for Multilingual Thesauri. *Knowledge Organization* 22: 136-140

Schubert, Klaus (1996): Melby on MT Semantics [Rezension von Melby 1995]. *Machine Translation News International* [14]: 13-14

Schubert, Klaus (demn.): Rezension von Melby 1995. *Language Problems and Language Planning*

Sherwood, Bruce Arne (1985a): Sintezo de Esperanto kaj de diversaj naturaj lingvoj. In Koutny (Hg.) 1985: 49-56 + Bibl. 158-159

Sherwood, Bruce Arne (1985b): Komputila traktado de esperanta teksto. In Koutny (Hg.) 1985: 153-160

Sherwood, Judith N.; Bruce Arne Sherwood (1982): Computer Voices and Ears Furnish Novel Teaching Options. *Speech Technology* [1982]: 46-51

Sjögren, Siv (1970): En syntax för datamaskinell analys av esperanto. FOA P rapport C 8264-11(64). Stockholm: Försvarets forskningsanstalt

Smirnov-Trojanskij, P. P. (1959): O perevodnoj mašine, postroennoj na osnove odnojazyčnoj perevodno-jazykovoj metodologii. In *Perevodnaja mašina 1959*: 5-27

Stoppoloni, Silvio (1982): Plansprachliches Rechner-Dialogsystem (Predis): prítakso kaj perspektivoj. In Koutny (Hg.) 1982: 88-101

Tesnière, Lucien (1959): *Éléments de syntaxe structurale*. Neue Aufl. 1982. Paris: Klincksieck

Trojanskij, P. (1959): Mašina dlja avtomatičeskogo proizvodstva nuždajuščichsja tol'ko v literaturnoj obrabotke gotovyh pečatnyh perevodov s odnogo jazyka odnovenno na rjad drugih jazykov. In *Perevodnaja mašina 1959*: 35-39

Turing, A. M. (1969): Intelligent Machinery. In Bernard Meltzer, Donald Michie (Hg.): *Machine Intelligence 5*. Edinburgh: University Press

Vasconcellos, Muriel (Hg.) (1988): *Technology as Translation Strategy*. Binghamton: State University of New York

Verbmobil. Mobiles Dolmetschgerät (1991). München: DFKI usw.

Wachter, Lieve de (1995): The Computational Interpretation of Germanic Noun-Noun Compounds. In Paulo Alberto, Paul Bennett (Hg.): *Lexical Issues in Machine Translation*. Luxembourg: Office for Official Publications of the Commission of the European Community, 19-36 + Bibl. 133-138

Wahlster, Wolfgang; Judith Engelkamp (Hg.) (1992): Wissenschaftliche Ziele und Netzpläne für das Verbmobil-Projekt. Saarbrücken: Deutsches Forschungszentrum für Künstliche Intelligenz

Weaver, Alan (1988): Two Aspects of Interactive Machine Translation. In Vasconcellos (Hg.) 1988: 116-123

Weaver, Warren (1955): Translation. In William N. Locke, A. Donald Booth (Hg.): *Machine Translation of Languages*. New York: Technology Press/MIT / London: Chapman & Hall, 15-23

Whitelock, Peter; Kieran Kilby (1995): *Linguistic and Computational Issues in Machine Translation System Design*. 2. Aufl. London: UCL Press

Witkam, A. P. M. (1983): *Distributed Language Translation*. Utrecht: BSO

Zimmer, Dieter E. (1996): Computer können dem Menschen das Übersetzen von Texten nicht abnehmen. Bloß ein wenig erleichtern. Texte, frisch geschreddert. *Die Zeit* 5. April 1996: 80