Linguistische
Arbeiten 149

Herausgegeben von Hans Altmann, Herbert E. Brekle, Hans Jürgen Heringer,
Christian Rohrer, Heinz Vater und Otmar Werner

Hans-Ulrich Block

Maschinelle Übersetzung komplexer französischer Nominalsyntagmen ins Deutsche

Max Niemeyer Verlag
Tübingen 1984

meinen Eltern

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Block, Hans-Ulrich:
Maschinelle Übersetzung komplexer französischer Nominalsyntagmen
ins Deutsche / Hans-Ulrich Block. - Tübingen : Niemeyer, 1984.
(Linguistische Arbeiten ; 149)
NE: GT

ISBN 3-484-30149-X ISSN 0344-6727

VORWORT

Diese Arbeit ist unter dem Titel "Maschinelle Übersetzung französischer Nominalsyntagmen ins Deutsche" von der philosophischen Fakultät der Universität zu Köln als Dissertation angenommen worden. Referenten waren die Herren Professoren A. Greive und H. Vater. Das Rigorosum fand am 4. 2. 1984 statt.

Für die geduldige Unterstützung der Arbeit danke ich zunächst Herrn Professor A. Greive, der die Dissertation betreute. Für Förderung und Interesse danke ich Herrn Professor H. Vater, der mir in zahlreichen Gesprächen die Möglichkeit gab, Probleme der Arbeit zu diskutieren. Besonders bedanken möchte ich mich auch bei meinen Freunden und Kollegen Dr. B. Rieger, Dr. J. Rolshoven, Dr. C. Thiersch, Dr. P.-O. Samuelsdorff und Thérèse Torris, die mir in vielen Diskussionen über linguistische Datenverarbeitung wichtige Anregungen gaben. Frau Dominique Dumas danke ich dafür, daß sie mir rund um die Uhr ihre Grammatikalitätsurteile zur Verfügung stellte. Schließlich sei noch der Firma HDM/Bonn dafür gedankt, daß sie mir ihren Drucker zur Erstellung des reproduktionsreifen Manuskripts überließ.

0. INHALTSVERZEICHNIS

1. EINLEITUNG

Dieses Kapitel gibt einen Überblick über den derzeitigen Stand der maschinellen Übersetzung (M.Ü.) und skizziert eine alternative Sichtweise zu den bisher geläufigen Verfahren der M.Ü.

1.1 Kurzer Überblick über die Geschichte der maschinellen Übersetzung

Im Laufe der Geschichte der maschinellen Übersetzung läßt sich eine Entwicklung von zunächst sehr oberflächlichen Wort-für-Wort-Übersetzungen über komplexere Phrasenstrukturanalysen und -synthesen von Quellsprache (QS) und Zielsprache (ZS) über immer "tiefer" reichende syntaktische Analysen der QS bis hin zu sog. semantischen Repräsentationen der zu übersetzenden Sätze, oft an logische Notationen angelehnt, beobachten.

Die Wort-für-Wort-Übersetzungen wurden bereits Anfang der sechziger Jahre fallen gelassen, als immer klarer wurde, "daß dieser lexikalische Ansatz mit allen möglichen Flickverbesserungen (in syntaktischer Hinsicht) doch an der 'syntactic barrier' (ein Ausdruck von V. Yngve) scheitern mußte." (Dietrich u. Klein 1974:119)

Die "Phrasenstrukturmethode" hat sich relativ lange gehalten, bis schließlich 1966 eine von der National Academy of Science eingesetzte Kommission (ALPAC) zu dem Ergebnis kam, daß man "in absehbarer Zeit keine maschinellen Übersetzungen zu erwarten habe, die praktisch brauchbar, zugleich aber billiger als die menschlichen Übersetzer seien". (Dietrich u. Klein 1974:121)

Hiernach gelangte die Forschung der M.Ü zunächst zu einem Stillstand, bis mit dem Aufkommen der generativen Transformationsgrammatik die Möglichkeit gesehen wurde, wesentlich "semantiknähere" Übersetzungssysteme zu konstruieren. In jüngster Zeit (der letzten Phase) ging man dann noch einen Schritt weiter und entwarf Systeme mit einer logikähnlichen semantischen Repräsentationsebene, durch die der Transfer von der QS in die ZS ging.

Ich möchte die Wort-für-Wort-Übersetzungen nicht weiter behandeln, aber die drei anderen Methoden vorstellen, um die in dieser Arbeit angewandte Vorgehensweise vor den forschungsgeschichtlichen Hintergrund zu stellen. Ich beziehe mich dabei auf die drei Methoden mit den Begriffen "Oberflächenstruktur-Methode" (OS-Methode), "Tiefenstruktur-Methode" (TS-Methode) und "semantische Methode".

1.1.1 Die Oberflächenstruktur-Methode

Den Aufbau eines solchen Systems beschreiben Klein und Dietrich:

> (5)(a) Syntaktische Analyse des Quelltextes; das Ergebnis ist eine syntaktische Strukturbeschreibung, beispielsweise ein P-Marker;
> (b) Austausch lexikalischer Einheiten; dabei wird u.U. der Kontext mitberücksichtigt.
> (c) Umformung der Strukturbeschreibung von Q in eine entsprechende Strukturbeschreibung von Z.
>
> (Dietrich u. Klein 1974:119)

Zu beachten ist hier im Vergleich zu den späteren Verfahren, daß es keine Interlingua, also eine künstliche Zwischensprache zwischen QS und ZS gibt. Hieraus folgt sogleich, daß es sich um Methoden für jeweils nur zwei Sprachen handelt, die QS und die ZS. Ferner liegt immer ein zweisprachiges Wörterbuch zugrunde, in dem die Wortformen der QS und der ZS mit jeweiligen syntaktischen Informationen die Einträge bilden. Die Grammatik war von dem Computerprogramm nicht getrennt, sondern bildete eine Einheit. Den Regeln entsprachen "programmierte Algorithmen" (Huckert 1979:9). Die Oberflächenstruktur-Methode scheiterte schließlich an lexikalischen und syntaktischen Mehrdeutigkeiten sowie "an dem Problem der Übersetzung idiomatischer Wendungen" (Huckert 1979:9).

Ferner mag noch ein Grund für das Scheitern gewesen sein, daß man sich, ohne eine Theorie vor Augen gehabt zu haben, immer mehr in sog. ad-hoc-Lösungen verloren hat.

1.1.2 Die Tiefenstruktur-Methode

Mit dem Aufkommen der Transformationsgrammatik, insbesondere auch der generativen Semantik, deren Vertreter ja proklamierten, es gäbe eine universale, sprachunabhängige TS[1], wurden viele neue Forschungsarbeiten zur M.Ü. in Gang

---

1 So z.B. Bach (1968:114): "The actual rules of the base are the same for every language".

gesetzt. Den Aufbau solcher Verfahren skizzieren Dietrich und Klein. Die Verfahren enthalten

> (1)(a) eine explizite Grammatik G(Q) der Quellsprache, d.h. ein System von Regeln, das die Beziehung von Laut und Bedeutung in Q beschreibt, genauer gesagt, den Zusammenhang zwischen semantischen Repräsentationen und zulässigen Symbolfolgen in Q darstellt;
> (b) eine ebensolche Grammatik G(Z) für die Zielsprache;
> (c) ein System R von Regeln, das die semantischen Repräsentationen von G(Q) und von G(Z) einander zuordnet; dies ist nicht nötig, falls bei beiden Grammatiken die semantischen Repräsentationen dieselben sind [...].
>
> (Dietrich u. Klein 1974:119)

Die Vertreter der TS-Methode schränken obiges dahingehend ein, daß sie auf eigentliche semantische Repräsentationen verzichten und die Tiefenstruktur selbst als semantische Repräsentation betrachten[2].

Neben einzelsprachlichen Vorteilen, wie die Möglichkeit der Behandlung von diskontinuierlichen Konstituenten, wie in (1.1),

(1.1) a. Karl läuft weg.
b. ... weil Karl wegläuft.

Topikalisierungserscheinungen etc. hat diese Methode den Vorteil, daß leicht aus mehreren QS in mehrere ZS übersetzt werden kann. Angenommen, es sollen n Sprachen behandelt werden, und zwar so, daß aus jeder Sprache in jede andere Sprache übersetzt werden kann, dann benötigt man genau n Analysegrammatiken und n Synthesegrammatiken. Benutzt man die OS-Methode, so benötigt man n*(n-1) Übersetzungssysteme.

Obwohl theoretisch große Erfolge mit der TS-Methode erzielt werden konnten, gelangte man doch sehr rasch an ihre Grenzen. Dies lag hauptsächlich an der (immer noch) eher syntaktischen Natur der TS, die nicht erlaubte, semantische und pragmatische Mehrdeutigkeiten aufzulösen. So läßt sich etwa die Übersetzung von homme in (1.2) nicht eindeutig bestimmen, sodaß für (1.2) sowohl (1.3) a. als auch (1.3) b. in Frage kommen[3].

(1.2) Tous les hommes sont mortels.

(1.3) a. Alle Menschen sind sterblich.
b. Alle Männer sind sterblich.

---

2 "Die Hypothese lautet also, daß eine Übersetzung ohne Rückgriff auf semantische Repräsentationen, jedoch unter Einbeziehung der Tiefenstruktur möglich ist."

(Dietrich u. Klein 1974:122)

3 Vgl. Huckert 1979:18.

Ein weiterer Nachteil der TS-Methode besteht in der Beschränkung auf die Satzebene, sodaß z.B. die "Bedeutungsgleichheit von Einzelsätzen wie Er fährt einen Porsche. und Satzverknüpfungen wie Er fährt einen Wagen. Es ist ein Porsche." (Stachowitz 1973:9) nicht dargestellt werden kann[4], oder es z.B. keine Möglichkeit gibt, Pronomen zu beziehen.

1.1.3 Die semantische Methode

Bei dieser Methode ist die Interlingua keine Tiefenstruktur im Sinne der Transformationsgrammatik mehr, sondern entspricht eher einer "semantischen Repräsentation". Die semantische Repräsentation ist theoretisch satzübergreifend und erlaubt den Anschluß des Systems an eine Datenbasis, was für pragmatische Mehrdeutigkeiten (solche, für deren Auflösung Weltwissen notwendig ist) von Vorteil ist. Solche Anforderungen stellen z.B. die Mitarbeiter des SFB 99 an ihr System SALAT:

> Vielmehr soll die Konstruktion von SALAT parallel zur Entwicklung einer Sprachbeschreibungstheorie erfolgen, die ein semantisches Repräsentationssystem auf logischer Grundlage bereitstellt und sowohl die Einbeziehung pragmatischer Informationen als auch die kalkülmäßige Erfassung von Bedeutungszusammenhängen möglich macht. (SFB 99 I, 1976:13)

Hier wird die richtige Lesart eines ambigen Satzes herausgefunden "mit Hilfe der Deduktionskomponente unter Verwendung von aus dem umgebenden Text (kotext) und der Verwendungssituation (kontext) zu entnehmender Information sowie unter Rückgriff auf eine Datenbasis, die außerlinguistische, nicht dem Kotext zu entnehmende Information ('Weltwissen') bereitstellt [...]". (SFB 99 I, 1976:19). Es ist ein genereller Trend zu beobachten, der dahin geht, die M.Ü. nicht mehr vorwiegend als linguistisches Problem zu betrachten, sondern sie in den weiteren Rahmen der "Artificial Intelligence" (AI)-Forschung zu stellen[5].

1.2 Anforderungen an die maschinelle Übersetzung

Eine realistische Einschätzung der bisherigen und zukünftigen Entwicklung der M.Ü. scheint trotz aller positiven Teilergebnisse doch eher zu dem Ergebnis zu führen, daß eine vollautomatische Übersetzung mit hoher Quali-

---

4 Ich teile allerdings Stachowitz' Meinung, diese Sätze seien "bedeutungsgleich", nicht.

5 "Die AÜ (Automatische Übersetzung, H.-U. B.) wird als Teilgebiet der AI (Artificial Intelligence) begriffen." (Huckert 1979:10)

tät zu niedrigen, d.h. wirtschaftlich interessanten Preisen in den nächsten Jahrzehnten nicht erreicht werden wird. Daher bleibt die Frage: Warum betreibt man heute M.Ü.? Die Antwort darauf wird unterschiedlich ausfallen, je nachdem ob sie von einem "Sprachwissenschaftler" oder von einem "Wirtschaftsmanager" beantwortet wird.

Der Manager gibt sich mit einem System zufrieden, das zwar nicht völlig korrekt, wohl aber verständlich übersetzt[6], sofern es die Übersetzung nur schneller und billiger beschafft als ein menschlicher Übersetzer. Ihn interessiert die linguistische Theorie nur als Mittel zum Zweck, der hier die Verbesserung der Qualität und Effizienz des Übersetzungssystems ist.

Den Sprachwissenschaftler interessieren hingegen Kosten und Schnelligkeit des Systems, ja selbst das Resultat der Übersetzung wenig. Ihm geht es um die "Überprüfung linguistischer Theorien durch vollständige Algorithmisierung" (Huckert 1979:6). Huckert formuliert die unterschiedlichen Anforderungen an "ökonomisch orientierte" und "theoretisch orientierte Verfahren" wie folgt:

> -ökonomisch orientierte Verfahren müssen große Datenmengen in kurzer Zeit verarbeiten. Die Ausführungsgeschwindigkeit spielt dagegen in theoretischen Verfahren eine untergeordnete Rolle
>
> -Theoretische Verfahren konzentrieren sich auf interessante, relativ willkürlich gewählte Sprachausschnitte, die statistisch möglicherweise uninteressant sind. In ökonomischen Systemen wird der gewählte Sprachausschnitt im allgemeinen größer und homogen sein. Die ausgewählten Sprachausschnitte sind meist einem bestimmten Fachgebiet entnommen, z.B. der Meteorologie (System METEO), dem Flugzeugbau (TAUM - Aviation) oder der Textilindustrie (TITUS). Die in solchen Systemen bearbeiteten Sprachausschnitte sind linguistisch nicht sehr interessant.
>
> -ökonomisch orientierte Systeme können Sprachbeschreibungen verwenden, die linguistisch uninteressant (weil trivial) oder theoretisch überholt sind. Das Ziel solcher trivialer Verfahren ist die Korrektheit des Endresultats. Wie dieses Endresultat zustande kam, interessiert nicht. Dagegen ist die Korrektheit der linguistischen Theorie für theoretisch orientierte Verfahren ein zentrales Ziel und gleichzeitig Arbeitshypothese. Nicht das Endresultat, also die Übersetzung interessiert, sondern die Umsetzung der linguistischen Theorie in Regeln und Algorithmen.
>
> (Huckert 1979:6f.)

Huckerts sehr extreme Unterscheidung ist m.E. an einigen Punkten etwas abzuschwächen.

---

6 Ich habe bei verschiedenen Vorführungen kommerzieller Systeme feststellen müssen, daß sich der "Manager" meist mit Übersetzungen zufrieden gibt, die man nicht ohne Wohlwollen als verständlich erachten kann.

Erstens meint Huckert offensichtlich "Übersetzungstheorie", wenn er von "linguistischer Theorie" spricht, denn zur Überprüfung linguistischer Theorien benötigt man kein M.Ü.- System. Ein "grammar-tester" (Stachowitz 1973:3), also ein System, das lediglich Sätze einer Sprache gemäß einer eingegebenen Grammatik analysiert oder synthetisiert, ist hierzu besser geeignet.

Zweitens kann man durch die maschinelle Überprüfung von Grammatiken lediglich feststellen, ob sie "in sich" stimmen, d.h. widerspruchsfrei sind. Zur empirischen Evaluierung von Grammatiken benötigt man sehr große Datenmengen. Bisher sind aber gerade nur ökonomisch ausgerichtete M.Ü.-Systeme in der Lage, große Datenmengen in akzeptabler Zeit zu verarbeiten.

Daher stellt sich die Frage, ob es überhaupt möglich ist, mit Hilfe der M.Ü. die "Richtigkeit" einer linguistischen Theorie zu überprüfen. Ein realistischerer Anspruch an die theoretische M.Ü. ist der, eine linguistische Theorie daraufhin zu überprüfen, ob sie für die M.Ü. brauchbar ist, oder zu prüfen, welche von mehreren Theorien brauchbarer ist.

Unterscheidet man zwischen Erkenntniswissenschaften und Ingenieurswissenschaften, dann gehört die M.Ü. sicherlich in den zweiten Bereich. Ich verstehe daher die M.Ü. als Teil der angewandten Sprachwissenschaft. Unter diesem Aspekt spielt aber die Leistung eines Systems bezüglich benötigter Rechenzeit, Speicherplatz etc. keine untergeordnete Rolle mehr, sondern ist für die Evaluierung eines Systems relevant.

Was für das Verhältnis von M.Ü. zur linguistischen Theorie gilt, gilt im Prinzip auch für ihr Verhältnis zur Beschreibung der sprachlichen Daten in einer bestimmten Theorie: Die M.Ü. kann auf bereits von Linguisten erarbeitete Beschreibungen zurückgreifen und sie "anwenden". Leider gibt es zur Zeit noch keine explizite und vollständige linguistische Beschreibung einer Sprache:

> Das Widerstreben der Linguisten, einen Computer zu benützen, beruht natürlich auf der Tatsache, daß es keine umfassende Theorie der Grammatik gibt, die funktioniert. Schätzungen über die voraussichtliche Zeitdauer zur Konstruktion einer solchen Grammatik variieren beträchtlich. Wir haben sogar Meinungen gehört, daß dazu eine Zeit von etwa 500 Jahren erforderlich sei. Sicherlich ist diese Zahl eine Übertreibung. Dennoch hat eine Anzahl bekannter Linguisten ernsthaft versichert, daß es nach ihrer Ansicht etwa 150 Jahre grammatischer Untersuchungen bedürfe, bevor man eine komplette Grammatik für eine Sprache entwikkelt habe.
>
> (Stachowitz 1973:2)