

STATUSSEMINAR ZU EUROTRA-D UND BEGLEITFORSCHUNG IM FEBRUAR 1987
EUROTRA präsentiert erste Ergebnisse

Johann Haller
IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Uwe Reyle
Universität Stuttgart, Fachbereich Informatik
D-7000 Stuttgart

EUROTRA ist ein gemeinsames Projekt aller EG-Mitgliedsstaaten, in dem ein vorindustrieller Prototyp zur Maschinellen Übersetzung entwickelt werden soll. Alle 9 Sprachen der EG sollen sowohl Ausgangs- als auch Zielsprache sein können. In einem Statusseminar stellten die Projektgruppe EUROTRA-D am IAI, Saarbrücken sowie die von den Universitäten Stuttgart, Bielefeld und Berlin getragene Begleitforschung erste Forschungsergebnisse vor. In dem Bericht zu dieser Veranstaltung werden zunächst Konzept, Aufgaben und Abstimmung von EUROTRA-D-Forschungsgruppe und Begleitforschung skizziert. Es wird erklärt, für welche Teilaufgaben in dem EUROTRA-Stratifikationsansatz die Begleitforschung Resultate und Methoden der theoretischen Linguistik einbringen und hinsichtlich ihrer Anwendbarkeit überprüfen will. Der Verlauf des Statusseminars mit Vorträgen, Diskussion und Vorführungen wird kurz geschildert. Der Beitrag schließt mit Bemerkungen zum erreichten Stand und den Perspektiven der Maschinellen Übersetzung.

Vom 16. bis 18. Februar haben in einem gemeinsamen Statusseminar in Saarbrücken / Dudweiler die deutsche Forschungsgruppe EUROTRA-D (Universität des Saarlandes) und die Universitäten Stuttgart, Bielefeld und Berlin erste Ergebnisse ihrer Forschungen zur maschinellen Übersetzung vorgelegt und praktisch demonstriert.

EUROTRA ist ein gemeinsames Projekt aller 12 Mitgliedsstaaten, das zunächst die Computerlinguistik und den Wissenstransfer innerhalb der EG fördern soll. Jedoch ist auch die Erstellung eines vorindustriellen Prototyps vorgesehen, der bis 1990 mit einem begrenzten Wortschatz Übersetzungen zu Texten, die in der Verwaltungssprache der EG abgefaßt sind, liefern soll. Dabei sollen alle 9 Sprachen der Gemeinschaft analysiert und generiert sowie Transfer in allen 72 möglichen Richtungen realisiert werden. Die Kommission der EG organisiert dieses Projekt, finanziert zentrale Forschungsarbeiten zu Linguistik und Software und beteiligt sich an den Kosten der nationalen Teilkomponenten. An der Finanzierung der deutschen Teilkomponente EUROTRA-D am Institut der Gesellschaft zur Förderung der Angewandten Informationswissenschaft ist neben dem Saarland auch das Bundesministerium für Forschung und Technologie mit 52% beteiligt.

Die Aufgabe der deutschen Forschungsgruppe in Saarbrücken ist die Erstellung der deutschen Analyse- und Syntheseprogramme und der Wörterbücher, die Lexika für den Transfer aus den anderen Amtssprachen der EG ins Deutsche werden in Zusammenarbeit mit der Universität Bonn aufgebaut. Als Grammatikformalismus ist für die deutsche Gruppe (wie für alle anderen Sprachen) ein zentral entwickeltes Konzept vorgegeben, das auf Unifikationsgrammatiken basiert und den Übersetzungsvorgang in kleine Schritte zerlegt. Zwischen Ausgangs- und Zieltext werden eine Reihe von linguistisch motivierten Ebenen angenommen, zwischen denen wieder Übersetzungen stattfinden. Für das Deutsche werden Repräsentationen auf morphologischer, konfigurationaler und relationaler Ebene erarbeitet, die jeweils auf aktuellen Forschungsergebnissen in Syntax und Seman-

tik beruhen. Gleichzeitig arbeiten deutsche Forscher an der konzeptionellen Weiterentwicklung des EUROTRA-Formalismus und an Problemen der effizienten Implementierung mit. Zur Unterstützung der EUROTRA-D-Forschung fördert das BMFT gleichzeitig unabhängige Begleitforschungs-Projekte an den Universitäten Berlin, Bielefeld und Stuttgart. Die Aufgabe dieser Projekte ist vornehmlich, die Resultate und Methoden der theoretischen Linguistik für die Anwendung auf Maschinelle Übersetzung zu überprüfen und nutzbar zu machen.

So werden z.B. an der Universität Stuttgart von theoretischen Linguisten umfangreiche Grammatiken des Deutschen und Französischen im Rahmen einer Unifikationsgrammatik, der Lexikalisch Funktionale Grammatik, erstellt. Parallel dazu wurden Compiler entwickelt, die diese (und beliebige andere in diesem Formalismus entwickelten) Grammatiken in Parser und Generatoren überführen. Parser erzeugen aus Sätzen syntaktisch-semantische Repräsentationen. Generatoren überführen solche Repräsentationen wiederum in sprachliche Ausdrücke. Ähnliche Forschungen werden an der Universität Berlin auf der Basis eines verwandten Grammatikformalismus, der Generalisierten Phrasenstrukturgrammatik, durchgeführt. Das Berliner Projekt versucht darüberhinaus auch Ansätze und Methoden der Künstlichen Intelligenz zu inkorporieren. An der Universität Bielefeld wurden die einzelnen Komponenten auf ihre Effizienz untersucht und entsprechend optimiert.

Für eine zweite, geplante Projektphase ist vorgesehen, entsprechende Transferautomaten zu konstruieren. Diese Transferautomaten sollen unter Verwendung von strukturellen Transferregeln und eines Transferlexikons die von den Parsern erstellten abstrakten Repräsentationen einer Quellsprache in entsprechende Repräsentationen einer Zielsprache übersetzen, aus denen ein Generator dann sprachliche Strukturen erzeugen kann. Eines der Hauptmerkmale der Begleitforschung ist eine interdisziplinäre Zusammenarbeit und klare Arbeitsteilung zwischen theoretischen Linguisten und Informatikern. Eine solche Konzeption wird vor allem ermöglicht durch die Verwendung von mathematisch

klar definierten Grammatikformalismen, die es erlauben, die verschiedenen konzeptuellen Ebenen eines sprachverarbeitenden Systems (Grammatische Beschreibung vs. Implementierung in Form von Parsern, Generatoren, Transferautomaten) zu trennen. Durch die Verwendung von einheitlichen vorgegebenen Formalismen können dabei Grammatiken und Software-Komponenten völlig unabhängig voneinander erstellt, modifiziert und optimiert werden. Ein weiteres wichtiges Merkmal dieser Forschungskonzeption ist der modulare Aufbau der Systeme. Da einzelne Module der Übersetzungssysteme (Grammatiken, Parser und Generatoren) anwendungsunabhängig sind, können sie in beliebige Sprachverarbeitende Systeme inkorporiert werden, die eine Schnittstelle zu natürlichen Sprachen benötigen.

Auf dem Statusseminar stellten sich alle Forschungsinstitutionen mit Vorträgen und Vorführungen einer kleinen Menge von Sachverständigen aus Ministerium, Universitäten und Industrie vor. Der englische Sprachwissenschaftler D. Arnold von der Universität Essex hielt den Einführungsvortrag über den EUROTRA-Formalismus, der zu lebhaften Diskussionen Anlaß gab. Der Rest des ersten Tages war dann den verschiedenen Vorführungen gewidmet, wobei die EUROTRA-D-Gruppe ein kleines Modell einer mehrsprachigen Übersetzung mit einfachen Sätzen sowie verschiedenen Analysestufen des Deutschen präsentierte; die Universität Stuttgart und die Universität Berlin zeigten Analyse und Generierung für Deutsch und Französisch, Bielefeld stellte optimierte Parser vor.

Am zweiten Tag wurden die Projekte in einer Reihe von Kurzvorträgen auch theoretisch dargestellt; ergänzend dazu hielt H. Uskoreit (Stanford/Stuttgart) einen Überblicksvortrag über die "unifikationsbasierten Grammatiken". Einige Vorträge zeigten Möglichkeiten der Zusammenar-

beit zwischen der Begleitforschung und EUROTRA-D auf; besonders auffällig war dabei ein Experiment, bei dem mit dem Programm SEMSYN zur Generierung deutscher Sprache Sätze aus abstrakten EUROTRA-Tiefenstrukturen erzeugt wurden.

In der langen und wechselvollen Geschichte der maschinellen Übersetzung ist zur Zeit also wieder eine positive Tendenz sichtbar; für alle beteiligten Forscher ist klar, daß mittelfristig keine zu großen Erwartungen an die Güte der Übersetzungen gestellt werden dürfen, und daß es noch sehr vieler Vorarbeit in Sprachwissenschaft und Informatik bedarf, um auch nur einfache technische Texte so übersetzen zu lassen, daß eine Korrektur einer solchen "Rohübersetzung" für den menschlichen Bearbeiter einfacher ist als eine direkte Übersetzung. Dabei existiert auch noch die Konkurrenz der "maschinenunterstützten Übersetzung", bei der der Übersetzer durch Textsysteme und Terminologiedatenbanken unterstützt wird, was in größeren Firmen zunehmend praktiziert wird.

All dieser Probleme waren sich die deutschen Forscher, die in Dudweiler versammelt waren, deutlich bewußt, ebenso wie der Notwendigkeit, Ergebnisse auszutauschen und gemeinsam zu arbeiten. Sie wollen doch alle auf dem Wege zu dem faszinierenden Ziel eines übersetzenden Computers vorankommen: ein Stück sprachlicher Fähigkeit des Menschen nachahmen zu können, gibt das Gefühl, einen Teil seines Geistes besser zu verstehen. Für die Gesellschaft (und die Förderer) ist der Nutzen klar: EUROTRA kann ein entscheidender Beitrag zum Abbau der Informationshindernisse im vielsprachigen Europa sein.