

EUROTRA – GEGENWÄRTIGER STAND UND PLANUNG 1987/88

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

Das multinationale EUROTRA-Projekt zur maschinellen Übersetzung befindet sich gegenwärtig in der Phase II, die bis Juni 1988 abgeschlossen sein soll. In dieser Phase der linguistischen Forschungen sollen Defizite bei den kooperativen Aufgaben wie Entwicklung und formale Beschreibung linguistischer Strategien oder die Ermittlung, Definition und Spezifikation von universeller linguistischer Information durch verstärkte Anstrengungen ausgeglichen werden. Die auf die Einzelsprache bezogenen Arbeiten sollen Zuverlässigkeit und Coverage der Systemkomponenten erhöhen und die theoretische Ausarbeitung verbessern. Die Phase III soll dann bis Mitte 1990 der Systemstabilisierung und der Auswertung dienen. Der vorliegende Beitrag stellt neben dem Forschungsstand auch den Stand des Organisationsbereichs dar und unterstreicht dessen Bedeutung für den Erfolg des Forschungsprojektes.

Seit dem Beginn der Phase II des Projekts EUROTRA im Jahre 1985 sind weltweit die Aktivitäten auf dem Gebiet der Maschinellen Übersetzung und besonders auch in der sog. "Maschinengestützten Übersetzung" verstärkt gefördert worden. In Europa werden drei Systeme zur Maschinellen Übersetzung von der Industrie auf den Markt gebracht und in einigen Fällen auch in größerem Umfang angewendet; dazu kommt eine große Zahl von Software-Paketen mit integrierter Textbearbeitung und Terminologiepflege zur Unterstützung des professionellen Übersetzers.

Die kurze Übersicht in *Haller 1987a* zeigt, daß das Interesse an der Erstellung von MÜ-Systemen und der Forschung auf diesem Gebiet sehr stark gewachsen ist; allerdings haben auch die verstärkten Anstrengungen bisher noch nicht zu einem System geführt, das den bisher existierenden durch seine linguistische Konzeption, seine gründliche Ausarbeitung und benutzerfreundliche Software überlegen wäre.

Dabei gelingt es heute bereits, mit Einsatz geeigneter Software zur Sprachdatenverarbeitung den Linguisten von informatischen Fragestellungen weitgehend zu entlasten; die größeren Probleme liegen nach wie vor darin, eine für ein multilinguales System geeignete linguistische Beschreibungsebene zu finden, die auf die Belange der Übersetzung zugeschnitten ist und gleichzeitig ein Maximum an Effizienz erlaubt.

Auf dem Wege zu diesen Zielen ist EUROTRA in den letzten Jahren sowohl organisatorisch als auch technisch ein Stück vorwärtsgekommen. Die Mitgliedsländer der Europäischen Gemeinschaft haben (inklusive Spanien und Portugal) nunmehr alle Assoziationsverträge unterschrieben, in denen sie sich verpflichten, die nationalen Beiträge zur Durchführung dieses Forschungs- und Entwicklungsprogramms zu leisten; der Schwerpunkt der Organisatorischen Arbeit liegt darin, die durch die in den einzelnen Ländern unterschiedlichen Fortschritte erzielten Ergebnisse einander anzugleichen. Zu diesem Zweck hat auch das Projektteam der EG-Kommission eine erhebliche Verstärkung erfahren.

Durch den frühzeitigen vollen Einsatz einiger Mitgliedsländer (darunter auch die Bundesrepublik Deutschland) liegen nunmehr erste Erfahrungen mit den linguistischen Spezifikationen sowie der vorgesehenen Software vor, was erwarten läßt, daß die Entwicklung für die restlichen Länder beschleunigt erfolgen kann; die erste praktische Demonstration

einer Übersetzung zwischen drei Sprachen wurde am 9.2.1987 in Luxemburg durchgeführt. Die deutsche Forschungsgruppe konnte wegen ihrer günstigen Voraussetzungen dazu einen erheblichen Beitrag leisten; sie konnte auf den hier vorhandenen Ergebnissen – Material des SFB 100 in Saarbrücken sowie der EUROTRA-D-Begleitforschung – aufbauen.

Dabei war (und ist auch für die Zukunft) große Disziplin und eine intensive Kooperationsbereitschaft in der Erarbeitung und Ausführung gemeinsamer Konzepte nötig; in den mittlerweile vom Projektmanagement geschaffenen Verfahren zur konkurrierenden Forschung und zur Integration der erarbeiteten Ergebnisse kann nicht ein Staat völlig eigene Wege gehen. Es ist zu hoffen, daß sich neben dem Forschungsbereich auch der Organisationsbereich, der womöglich in gleicher Weise über Erfolg und Mißerfolg eines solchen multinationalen Forschungsprojekts entscheidet, weiter festigt.

In dem Beschluß des Rates der Europäischen Gemeinschaft vom 04.11.82 zur Schaffung eines Systems zur automatischen Übersetzung zwischen sieben Amtssprachen wurden am 26.11.86 aufgrund des Beitritts von Spanien und Portugal einige Änderungen vorgenommen. Dabei werden die Gesamtkosten des Projekts erhöht, das Projekt-Team der Kommission vergrößert sowie die Phasen II und III des Projekts um insgesamt 18 Monate verlängert. Nach diesem Zeitraum soll ein voroperationelles prototypisches System zur Verfügung stehen, das – obgleich fachspezifisch und lexikalisch eingeschränkt – als Grundlage einer industriellen Entwicklung dienen soll.

Das Ende der Phase der linguistischen Forschung ist nunmehr auf Juni 1988 terminiert, die Phase der Stabilisierung und Auswertung auf Mitte 1990.

Für die erweiterte Phase II bis Juni 1988 soll ein größerer Schwerpunkt als bisher auf die kooperativen Aufgaben gelegt werden, da sich bei diesen bisher umfangreiche Probleme gezeigt haben:

- Entwicklung moderner linguistischer Strategien
- Beschreibung der linguistischen Strategien im EUROTRA-Formalismus, Beurteilung der Adäquatheit der EUROTRA-Software sowie Vorschläge zu ihrer Erweiterung und Verbesserung (in enger Zusammenarbeit mit der ebenfalls in Saarbrücken angesiedelten Gruppe, die zur Zeit an Software-Tuning und -weiterentwicklung arbeitet, siehe *Schütz 1987*)

- Weitere Ausarbeitung von Vorschlägen für die linguistischen Spezifikationen und die allgemeine Methodologie
- Ermittlung und Definition universeller linguistischer Informationen (wie Zeitrelationen, Modalität, semantische Relationen und Kategorien)
- Weiterentwicklung der Transferkonzepte, auch in Zusammenarbeit mit der deutschen Begleitforschung zu EUROTRA-D
- Entwicklung und Implementierung von Transferregeln
- Erstellung von Transferlexika.

Bei den sprachspezifischen Aufgaben geht es um den Test und weiteren Ausbau der einzelnen Komponenten, die in ihrem Umfang mehr als bisher über das vorhandene Korpus hinausgehen und Repräsentativität erreichen sollen. Auch hierbei ist die theoretische Auseinandersetzung und Gegenüberstellung mit den grammatischen Modellen der Begleitforschungsprojekte wichtig. Im einzelnen handelt es sich um folgende Aufgaben:

- Test und Überarbeitung der linguistischen Spezifikationen für die deutsche Sprache

- Test und Überarbeitung der Analysemoduln, die es erlauben, bestimmte Repräsentationsebenen zu erreichen und miteinander zu verknüpfen (Morphologie, Oberflächen- und Tiefensyntax, Interface-Struktur)
- Test und Überarbeitung der Synthesemoduln, die den Übergang von einer höheren zu einer niedrigeren Repräsentationsebene ermöglichen (Interface-Struktur, Syntax, Morphologie, Stringdarstellung)
- Erstellung der Transferkomponenten zwischen den Interface-Repräsentationen der anderen Sprachen und der deutschen Interface-Repräsentation.

Über die spezifische Behandlung der Probleme des Deutschen sowie über die Vorschläge von EUROTRA-D für die semantischen Komponenten wird in den Beiträgen Schmidt 1987, Zelinsky-Wibbelt 1987a, Steiner 1987 berichtet.

Literatur

Die Literaturangaben für die Fachaufsätze zum Themenschwerpunkt Maschinelle Übersetzung finden sich zusammengefaßt auf Seite 23.



DAS "LDV-Forum" IM ABONNEMENT

Auch Nicht-Mitglieder der GLDV können jetzt das LDV-Forum regelmäßig beziehen. Dazu ist es notwendig, formlos ein Abonnement bei der Redaktion

Redaktion des LDV-Forum, Prof. Dr. G. Knorz, Fachhochschule Darmstadt, Fachbereich IuD, Schöfferstraße 8, D-6100 Darmstadt

zu bestellen und den Preis der Einzelexemplare jeweils für das laufende Jahr auf das Konto

Gesellschaft für Linguistische Datenverarbeitung / LDV-Forum, Kto.-Nr. 554 090, BLZ 508 501 50, Sparkasse Darmstadt

zu überweisen (Vorkasse!). Für diese DM 30,- pro Jahr gehen Ihnen dann die beiden Hefte unmittelbar nach Erscheinen (ohne weitere Versandkosten) per Post zu. An der Tatsache, daß Sie den aktuellen Newsletter (wird gerade vorbereitet, siehe dazu das Editorial), den Rabatt auf die Reihe Linguistische Datenverarbeitung (OLMS) und vieles andere versäumen, sind Sie dann allerdings selbst schuld! Dazu müßten Sie dann doch Mitglied der GLDV werden (...was ich Ihnen in jedem Falle empfehlen kann).

Gerhard Knorz