

Universität des Saarlandes  
Fachrichtung 5.5 Informationswissenschaft

Harald H. Zimmermann

EUROTRA – Ziele, Aufgaben, Stand des Projekts

F-517-04, 07.05.84

Einführung

Die Verfahren zur maschinellen Übersetzung in den 60-er und 70-er Jahren waren vorwiegend computerlinguistisch motiviert: Im Mittelpunkt stand dabei das Problem, zu zeigen, dass und ggf. bis zu welchem Grade natürliche Sprachen so formalisiert werden können, dass auch maschinelle Prozesse des Sprachverstehens oder zumindest der automatischen grammatischen Sprachbeschreibung damit möglich werden. Die relativ langsamen technischen Geräte (im Vergleich zu heute) führten vielfach zu Ad-hoc-Lösungen, v.a. im Bereich der Analyse-Strategie, um die Probleme der Disambiguierung grammatischer Strukturen oder von Wort- und Satzbedeutungen zu lösen.

Auch wenn v.a. in der Anfangszeit die Mittel für solche Forschungen aus dem militärischen Bereich kamen (Russisch-Englisch war ein beliebtes Übersetzungspaar), so darf dies nicht darüber hinwegtäuschen, dass die praktische Anwendung allenfalls als Fernziel zu sehen war.

Die neueren linguistischen Modelle (Dependenzgrammatik, generative Syntax, generative Semantik, Tiefenkasusgrammatik) wurden schließlich in der 2. Hälfte der 70-er Jahre bestimmend für "höherwertige" Systeme, d.h. systematischere Ansätze. So stellt SYSTRAN in seiner heutigen Form eher einen dependentiellen Strukturansatz im Syntax-Bereich dar, während das Saarbrücker System SUSY eher auf der Grundlage der generativen Syntax realisiert wurde. Der analysestrategische Bereich spannt sich von oberflächenorientierten stufigen Verfahrensweisen (SYSTRAN, SUSY) zu Verfahren, die Syntaxanalyse, Semantik u.a.m. integriert behandeln (vgl. v.a. die ATN-Parser).

Insgesamt kann heute festgehalten werden, dass alle bisherigen Anstrengungen, vollautomatische maschinelle Übersetzungen natürlichsprachlicher Texte mit sehr guter oder zumindest hoher Qualität für breite Fachgebiete und beliebige Textarten durchzuführen, nicht den gewünschten Erfolg gebracht haben.

Umgekehrt ist der Bedarf in Wirtschaft und Wissenschaft angesichts des wachsenden internationalen Wissensaustauschs an derartigen Hilfen stark gestiegen. Technologisch sind, besonders in Bezug auf Speicherplatz und Rechenzeit, große Fortschritte gemacht worden. Insofern tritt zunehmend der sprachwissenschaftlich-konzeptionelle Aspekt zurück hinter der Forderung, brauchbar-praktische Systeme zur maschinellen Übersetzung zu entwickeln, die i.W. nur noch folgenden Kriterien gerecht werden müssen:

- sie müssen den Übersetzungsprozess (bzw. den Wissenstransfer) beschleunigen
- sie müssen die Kosten der Übersetzung reduzieren

- sie müssen sich in eine moderne, micro-computergesteuerte Informationsumgebung einpassen.

Der letzte Aspekt hat dazu geführt, dass folgende zwei Wege beschritten wurden:

- Einerseits werden Verfahren wie SYSTRAN mit Schnittstellen zu Textsystemen versehen. Es ist z.B. heute möglich, einen englischen Text auf einem Textsystem zu erfassen und über eine Telekommunikationsschnittstelle an das Übersetzungssystem auf einem Großrechner zu überspielen. Nach der Übersetzung durch den MT-Host werden die Ergebnisse samt den Originalen wieder auf das Textsystem übertragen, wo der Übersetzer nun anhand der üblichen Editier-Funktionen eine Nachkorrektur vornehmen kann.
- Man kann natürlich auch versuchen, die Übersetzung auf dem Microcomputer/Textverarbeitungssystem selbst durchzuführen, wie es z.B. bei LOGOS geschieht. Hierbei muss man jedoch (heute) noch berücksichtigen, dass die Arbeitsgeschwindigkeiten und auch die Speichermöglichkeiten noch relativ bescheiden sind.

Für die Zukunft kann erwartet werden, dass Übersetzungssysteme nicht nur für den Übersetzerarbeitsplatz, sondern auch für den einer Ziel- oder Ausgangssprache nicht oder wenig Kundigen entwickelt werden. Dabei werden sog. "Informativ-Übersetzungen" aus "unbekannten" Fremdsprachen ebenso möglich sein (die dann nach einem Downloading in den Homecomputer evtl. leicht "stilistisch" korrigiert werden) wie Bürosysteme, die gute Übersetzungen in fremde Sprachen durch Rückfragen (Klärungsdialog) an den Autor möglich machen.

Die interessante Phase der Anwendung von MT hat also eigentlich erst begonnen. In diese Phase fallen nun die Arbeiten zu einem europäischen Übersetzungssystem.

Im folgenden werden die wichtigsten Ziele und Aufgaben sowie der derzeitige Stand (Mai 1984) des Projekts zur Entwicklung eines europäischen Systems zur computergestützten Übersetzung (EUROTRA) vorgestellt. Zum besseren Verständnis sollen hierzu einige Leitmotive für diese Forschungen und Entwicklungen vorangestellt werden:

- EUROTRA bedeutet eine gemeinsame europäische Anstrengung, das Problem der Sprachbarrieren in der schriftsprachlichen Kommunikation zwischen Institutionen, Betrieben, Gruppen und Personen in Teilbereichen zu reduzieren. Dies soll und kann nicht einzelstaatliche Aktivitäten (z.B. in Bezug auf andere Sprachrichtungen) in Mitgliedsstaaten der Europäischen Gemeinschaft ver- oder behindern. Es wird zudem auf dem europäischen Markt auch weiterhin andere Software geben (man denke an LOGOS oder TITUS), die EG wird z.B. auch weiterhin - soweit dies ökonomisch sinnvoll erscheint - das in USA entwickelte System SYSTRAN einsetzen. Das Projekt EUROTRA soll aus diesem Verständnis heraus deutlich machen, dass die Staaten der EG gewillt sind, in dieser Frage einen gemeinsamen eigenständigen Beitrag zu leisten, v.a. auch in planerischer und konzeptioneller Hinsicht, und nichts (oder wenig) dabei dem Zufall zu überlassen.

- EUROTRA bedeutet zugleich den Versuch, in den Mitgliedsstaaten der EG das Know-How zu Fragen der Sprachdatenverarbeitung systematisch zu verbessern. Während z.B. in der Bundesrepublik und Frankreich bereits größere Zentren mit diesem Schwerpunktthema bestehen, sind in anderen Mitgliedsländern bislang nur geringe oder gar überhaupt keine ernsthaften Forschungen und Entwicklungen hierzu vorhanden. Vor allem in diese Länder soll über das "Vehikel" eines gemeinsamen Projekts ein "Wissens- und Technologietransfer" erfolgen; umgekehrt sollen diese landesspezifischen F&E-Zentren ihr spezielles Sprachwissen mit einbringen. Man kann jedoch diesen Aspekt durchaus noch allgemeiner sehen: Die beteiligten nationalen Zentren sollen ihr Wissen und auch die Arbeiten zu EUROTRA im eigenen Lande nicht verstecken, vielmehr soll durch Kolloquien, Vorträge usf. dieses Wissen weitergetragen und auch das nationale wissenschaftliche Feedback gefördert werden. Dem wissenschaftlichen Wissens- und Erfahrungsaustausch kann und darf nicht entgegenstehen, dass aus verständlichen (rechtlich-finanziellen) Gründen softwaretechnologische Basisentwicklungen, die im Rahmen von EUROTRA entstehen, nicht ohne spezifische Erlaubnis weitergegeben werden können.
  
- EUROTRA ist (auch in der ersten Stufe) der angewandten Forschung zuzurechnen. Es war von vornherein ein erklärtes Ziel, keine bedeutende Grundlagenforschung zu betreiben, sondern soweit wie möglich auf vorhandenem Wissen bzw. modernsten existierenden Werkzeugen aufzubauen. Dabei sollten allerdings Ad-hoc-Lösungen weitgehend vermieden und zukunftssträchtige Konzepte eingebracht werden. In diesem Zusammenhang muss v.a. deutlich gemacht werden, dass die vorhandenen investiven Mittel derzeit zu begrenzt sind, als dass in die zu erarbeitende Wissensbasis des Systems (v.a. die "Lexika" i.w.S.) so viel investiert werden könnte, wie es die Forschungen zur Künstlichen Intelligenz derzeit als notwendig erachten. Umgekehrt haben bestehende Systeme wie SYSTRAN oder SUSY - sozusagen an der "Untergrenze" der EUROTRA-Entwicklungen liegend - gezeigt, dass unter bestimmten Restriktionen brauchbar-praktische maschinelle Übersetzungen möglich sind. Für ein modernes praxisorientiertes (d.h. auch in angemessener Kosten-Nutzen-Relation stehendes) System sind v.a. drei Verbesserungen gegenüber bestehenden Systemen wichtig:
  - Einsatz einer spezifischen Software zur Sprachdatenverarbeitung
  - Vertiefung der "grammatischen" Beschreibung etwa auf die Ebene der semantischen "Tiefenkasus" (unter Berücksichtigung von Morphologie und Syntax sowie von lexikalischer Semantik)
  - Verbesserung der Analysestrategien zur Erhöhung der Effizienz und Akzeptanz.
  
- EUROTRA ist als ein flexibles, offenes System konzipiert. Gerade weil die Notwendigkeit, in begrenzter Zeit ein praxiswirksames System zu entwickeln, den Eigenentwicklungen Grenzen setzt, muss es so modular gestaltet werden, dass neuere Ergebnisse in der Forschung möglichst verlustfrei eingebracht bzw. ergänzt werden können. Es wird daher zu geeigneter Zeit auch wohldefinierte Schnittstellen geben, die es Dritten erlauben, EUROTRA-Bausteine für eigene Weiterentwicklungen bzw. spezifische Anwendungen zugrunde zu legen und umgekehrt externe softwaretechnologische Entwicklungen (z.B. ge-

rade aus dem KI-Bereich) in EUROTRA einzubringen. Auf entsprechende Weise sollen im Übrigen vorhandene Ergebnisse - etwa solche des SFB 100 - ggf. über Brückensoftware eingebracht werden können.

- EUROTRA verlangt bei aller nationalen "Freiheit" größte Disziplin in der Ausführung gemeinsamer Konzepte und eine intensive Kooperationsbereitschaft. Dies ist v.a. eine Herausforderung an das Projekt-Management. Es kann nicht angehen, dass der eine oder andere Staat hier eigene Wege geht, ohne zumindest die Ergebnisse der gemeinsamen Entwicklung mit einzubeziehen; umgekehrt müssen gerade im Projekt durchaus auf allen Ebenen (dies zeigt z.B. gerade die Frage der Software-Entwicklung) Wettbewerbsmöglichkeiten geschaffen werden, v.a. um bestehende Verfahren und Entwicklungen mit einzubeziehen. Wie die "Vorgeschichte" von EUROTRA gezeigt hat (geradezu ein klassisches Beispiel des homo prae-europaeensis), sind im organisatorisch-wissenschaftspolitischen Bereich die größten Spannungen zu erwarten. Der Erfolg oder Misserfolg von EUROTRA wird womöglich weniger im Forschungs- als im Organisationsbereich entschieden.