

Harald Zimmermann

Zur Entwicklung eines Europäischen Übersetzungssystems (EUROTRA)

Regensburg, November 1978

GLIEDERUNG:

0 Vorbemerkung

1 Zielsetzung

2 Organisation und Verantwortung

3 Gegenwärtiger Stand der Planungen, weitere Vorgehensweise

4 Einige ergänzende Bemerkungen zum Verständnis des King-Entwurfs

5 Vorschlag für einen Stufenplan zur Entwicklung von EUROTRA

ANMERKUNGEN

0 Vorbemerkung

Adressat der hier vorgelegten Stellungnahme ist in erster Linie der Ad-hoc-Ausschuss "Informationslinguistik", der den Bundesminister für Forschung und Technologie (BMFT) u.a. in dieser Frage beraten soll. Daher stehen Aspekte des Systems und der Projektkonzeption, die für die Bundesrepublik Deutschland von besonderem Interesse erscheinen, im Vordergrund.

Diese Studie stellt zugleich eine Ergänzung und Kommentierung des EUROTRA-Projektentwurfs dar, der von Frau M. King unter dem Titel "EUROTRA: Project Description" (Fassung vom 25.09.1978) vorgelegt wurde. Auf diese Vorgeschichte der Projektkonzeption wird in Abschnitt 3 eingegangen. Zunächst soll in Ergänzung zu Kap. 3 die Zielsetzung des Systems erläutert werden.

1 Zielsetzung

Mit Wahrscheinlichkeit wird man in Europa auf lange, unabsehbare Zeit mit dem Problem der Mehrsprachigkeit leben müssen. Es gilt daher, nach Wegen zu suchen, die daraus resultierenden Kommunikationsprobleme zu bewältigen. Eine Reihe von Möglichkeiten müssen dabei neben- und miteinander erprobt werden, so z.B. eine erweiterte und verbesserte Ausbildung in Fremdsprachen.

Ein weiterer Weg erscheint zu einer Zeit, da die computertechnologische Entwicklung zu dem einzelnen Arbeitsplatz im Betrieb und Büro vordringt, besonders erwägenswert. Es handelt sich um die Einbringung einer computergestützten Übersetzung in den Gesamtrahmen der textuellen Kommunikation. Der Begriff 'computergestützt' steht dabei anstelle von 'vollautomatisch' in der Erkenntnis, dass eine vollkommen automatisch ablaufende Übersetzung beliebiger Texte auf lange Sicht ein Wunschtraum bleibt, daran ändern auch tiefergehende Einsichten in das Funktio-

nieren von Sprache kaum etwas. Dieses bedeutet zugleich, dass überall dort, wo gute und zuverlässige Übersetzungen benötigt werden, der "Humanübersetzer" nicht ersetzt werden kann, sondern allenfalls in seiner Arbeit durch den Computer unterstützt wird. Diese Unterstützung kann auf vielfältige Weise geschehen: beim Erlernen der Fremdsprache, bei der Terminologiarbeit, beim "Nachschlagen" in einem (maschinellen) Lexikon, beim unmittelbaren (interaktiven) Übersetzen; allerdings muss der Übersetzer auf diese Unterstützung hin (um-)geschult werden. In einigen Bereichen, bei denen eine hoch qualifizierte ("gute") Übersetzung aus verschiedenen Gründen (zu hohe Kosten im Verhältnis zum Effekt, mangelndes qualifiziertes Personal) nicht sinnvoll oder möglich erscheint, eine maschinelle Rohübersetzung sich aber zur (Grob-)Information als nützlich und kostengünstig erweist, wird man u.U. auch mit automatisierten Verfahren auskommen. (Vielleicht ist dies beim Übersetzen von Massendaten wie Abstracts in der Chemie oder Medizin der Fall). In einigen wenigen, in Sprachumfang und Weltwissen sehr eingeschränkten Bereichen (sog. "Miniwelten") wird auch eine rein maschinelle Übersetzung höherer Qualität erreicht werden können.

Aus diesen Prämissen ergibt sich für das geplante "Europäische Übersetzungssystem" Arbeitstitel (EUROTRA), dass es bei dem Anspruch, in weiten Bereichen eingesetzt werden zu können (und damit einen merklichen Beitrag zur Überwindung von Sprachbarrieren zu leisten), Beschränkungen im Hinblick auf das gegenwärtig Machbare unterliegt. EUROTRA soll daher Komponenten enthalten für

- eine maschinelle Übersetzungshilfe für die Humanübersetzung
- eine maschinelle Rohübersetzung, die allerdings in Qualität und Leistung über den derzeit vorhandenen kommerziellen Systemen liegen dürfte
- die Einbringung maschineller Sprachverarbeitungsverfahren in allgemeine Methoden der Textbe- und -verarbeitung, der Lexikographie sowie der Terminologie- und Thesaurusarbeit
- die (Fremd-)Sprachenausbildung (Materialauswertung und -bereitstellung, v.a. zur fachsprachlichen Übersetzung, Test- und Drillhilfen, Simulationsverfahren).

2 Organisation und Verantwortung

Ohne Zweifel ist ein "Europäisches Übersetzungssystem" - sowohl in der Entwicklungsphase wie später in der Anwendung - eine europäische Angelegenheit. Andererseits hören die Sprachprobleme nicht an den Grenzen Europas (oder gar der Europäischen Gemeinschaften) auf, auch die Interessen der europäischen Staaten reichen darüber hinaus. Jedenfalls ist ein derartiges System mehr als eine nationale Angelegenheit (ganz abgesehen von den Entwicklungskosten). Insbesondere sind m.E. jedoch die Mitgliedsstaaten der Europäischen Gemeinschaft zu einer gemeinsamen Anstrengung verpflichtet, soweit diese zum jetzigen Zeitpunkt als in der Sache gerechtfertigt, im Umfang vertretbar und in der Verfahrensweise Erfolg versprechend erscheint. Man wird dabei davon ausgehen müssen, dass jedes Mitgliedsland sowohl Gebender als auch Nehmender ist: dies gilt für den "Input" (den Beitrag zur Systementwicklung, etwa im Hinblick auf die Analyse der jeweiligen Landessprache) wie auch für den "Output" (z.B. Übersetzung in die Landes-

sprache). Vor allem in der Phase der Systementwicklung können einem Mitgliedsland jedoch - bei aller Anerkennung der prinzipiellen Gleichverteilung von Aufgaben und Beteiligungen - Mehrleistungen zukommen, etwa aufgrund des bereits vorhandenen Know-Hows, vorhandener (Software-)Technologie oder besonderer Interessenschwerpunkte (z.B. im Bereich der Fachinformation). Trotz teilweise unterschiedlicher Interessen sollte die aktive Kooperation europäischer Wissenschaftler und Ingenieure auch in diesem Vorhaben gefördert werden. Dies wird zu einigen organisatorischen Problemen führen, v.a. dann, wenn die Aktivitäten nicht an einer Stelle lokalisiert, sondern (neben Koordinations- und Integrationszentrum) dezentrale F&E-Einheiten geschaffen werden.

Die Frage einer dezentralen Entwicklung (mit einem Koordinations- und Implementationszentrum) gegenüber einer Zentralisation aller Entwicklungsaufgaben ist dabei m.E. nicht zu entscheiden, ohne jeweils Nachteile in Kauf zu nehmen. Die absolute Zentralisation erleichtert zweifelsohne die Koordination von Teilaufgaben und bietet die besten Voraussetzungen für eine Kommunikation zwischen den Arbeitsgruppen. Allerdings würde die in Europa bereits vorhandene Infrastruktur (mit Zentren z.B. in Saarbrücken, Grenoble, Manchester, Pisa, Kopenhagen) dabei kaum ausgenutzt werden können. Es ist auch anzunehmen, dass die Mitgliedsstaaten dann entsprechende Aktivitäten in diesen Zentren abbauen müssten oder dass - bis zu einem gewissen Grad - Doppelarbeit geleistet würde.

Eine dezentrale Organisationsform des Projekts erschwert sicherlich die Kommunikation und Koordination, auch kann selbst bei präziser Aufgabenbeschreibung und -verteilung aufgrund spezifischer Interessen und Methoden der jeweiligen Entwicklungsstellen eine völlig konforme Entwicklung aller Systemteile - sofern überhaupt erwünscht - kaum erreicht werden.

Aufgrund der besonderen Situation in Europa, aber auch des Standes von Forschung und Entwicklung im Bereich der maschinellen Übersetzung, ist m.E. die folgende "dynamische" Organisationsform erwägenswert:

Von vornherein wird ein Zentrum für Forschung und Entwicklung installiert (in Luxemburg). Es wird mit der erforderlichen Hardware ausgestattet; hier wird die nötige Basissoftware bereitgestellt, entwickelt und gewartet, das Zentrum verantwortet die Entwicklung und Pflege der Arbeitswörterbücher (abgesehen von Testwörterbüchern); es stellt die technischen Anschlüsse zu Benutzern und dezentralen Entwicklungsstellen her; es sorgt für den Aufbau und die Integration der Rahmensoftware zur Textverarbeitung; es koordiniert und kontrolliert dezentrale Entwicklungen; es implementiert alle Systembausteine, führt die notwendigen Tests durch und betreut etwa erforderliche unabhängige Qualitäts- oder Effizienzuntersuchungen, wozu auch die Einbeziehung und Bereitstellung der notwendigen Test-Software zu rechnen ist.

Zu der Entwicklungsphase (u.a. in der Ausgangsphase, also etwa in den ersten 2 - 3 Projektjahren) werden die sprachspezifischen Teile des Systems dezentral in dafür geeigneten europäischen Zentren auf der Basis eines gemeinsamen Konzepts entwickelt. Dabei sollten - in angemessenem Rahmen - auch alternative Verfahren zugelassen sein; dies lässt zumindest im Hinblick auf Fragen der Effizienz und Qualität von Übersetzungen einen Freiraum, der im Zusammenhang mit Pilotanwendungen (durch unterschiedliche Benutzer) ggf. auch eine Diversifikation des Übersetzungsangebots erlaubt (z.B. kostengünstige, qualitativ ärmere Rohübersetzung gegenüber einer teureren, aber höherwertigen Übersetzung). Das Zentrum selbst hätte dabei die Möglichkeit, in

Entwicklungslücken selbst tätig zu werden (z.B. bei Sprachpaaren, in denen kein Mitgliedsland das erforderliche Know-How aufweist, oder in der Anwendung eines Alternativ-Verfahrens.

Softwaretechnologisch ist eine derartige Mischform heute problemlos zu realisieren. Erforderlich wäre in den dezentralen Entwicklungs- und Erprobungsstellen eine gewisse stand-alone-Kapazität und ein DFÜ-Anschluss an das Zentrum, organisatorisch könnte eine Arbeitsgruppe der Leiter der Entwicklungszentren an der Koordination beteiligt werden, daneben müssten aufgaben- und fachspezifische Kontakte, Kolloquien u.ä. auch den erforderlichen Wissens- und Informationsaustausch sicherstellen. Da die meisten der in Frage kommenden dezentralen Stellen sehr enge Beziehungen zu Hochschulen haben, könnten sie auch langfristig als Mittler zwischen Forschung und Entwicklung fungieren; d.h. nach Erfüllung der ersten Entwicklungsaufgaben im Rahmen von EUROTRA und Übernahme dieser Systemkomponenten durch das Zentrum könnte den dezentralen Stellen die Aufgabe zufallen, neue Erkenntnisse der Forschung auf dem Gebiet der maschinellen Sprachverarbeitung in die Anwendung hinüberzutragen. Zugleich bildet das hier jeweils vorliegende Potential an Wissenschaftlern, die sowohl mit Forschung als auch mit Entwicklung vertraut sind, eine wichtige Quelle für den Auf- und Ausbau des EUROTRA-Zentrums.

3. Gegenwärtiger Stand der Planungen, weitere Vorgehensweise

Überlegungen zu einem Europäischen Übersetzungssystem sind im Rahmen der Bemühungen der Kommission der Europäischen Gemeinschaften um die Überwindung von Sprachbarrieren entstanden. Erste Erfahrungen mit der Implementation eines auf dem Markt angebotenen Systems (SYSTRAN) haben dabei gezeigt, dass derartige Anstrengungen nicht ad-hoc, sondern auf solider Grundlage unternommen werden müssen, wobei die gegenwärtig in die Praxis umsetzbaren (sprach-)wissenschaftlichen Erkenntnisse mit zu berücksichtigen sind. Der den Aktionsplan "Mehrsprachigkeit" begleitende Beirat CETIL hat inzwischen eine Arbeitsgruppe eingesetzt, der Wissenschaftler der einzelnen Mitgliedsstaaten angehören, um die Möglichkeiten eines entsprechenden Entwicklungsprojekts vorzuklären.

In dieser Projektgruppe, die von Fall zu Fall durch weitere Experten und Fachgremien (z.B. zum Bereich Software und Semantik) ergänzt wurde, ist inzwischen eine erste gemeinsame Konzeption entwickelt worden, die in dem Entwurf von Frau King zusammengefasst ist. Diese Vorgehensweise hatte das Ziel, zu prüfen, inwieweit die Vorstellungen der Experten (insbesondere im Hinblick auf die zu entwickelnde bzw. zu integrierende Technologie und Software zur linguistischen Basis des Systems) auf eine gemeinsame Basis gebracht werden können. Wenn auch noch einige Präzisierungen ausstehen, ist dies - soweit solche Fragen überhaupt in einer Vorphase zu klären sind - sehr positiv verlaufen.

Die Auswahl dieser Experten (sie kommen für die Bundesrepublik Deutschland vorwiegend vom Sonderforschungsbereich "Elektronische Sprachforschung" aus Saarbrücken, für Frankreich aus Grenoble) erfolgte dabei weniger unter dem Gesichtspunkt der Repräsentativität, im Vordergrund stand die Frage der Erfahrung im Bereich der maschinellen Übersetzung. Da diese im allgemeinen in Europa bislang nur im Rahmen von Forschungsprojekten (an Hochschulen) vorliegt, wurde wiederum der Aspekt einer praxisnahen Zielsetzung solcher Projekte berücksichtigt. Insgesamt kann man davon ausgehen, dass die Expertengruppe weitgehend diesen Vorstellungen gerecht wurde. Mit der Vorlage der Konzeption in den einzelnen Mitgliedsstaaten zur Vorberatung und ersten Diskussion soll nun eine zweite Phase eingeleitet werden. Während dies für die Bundesrepublik Deutschland verhältnismäßig einfach erscheint (die Beratung erfolgt im Rahmen des

Ad-hoc-Ausschusses "Informationslinguistik", der seinerseits den Bundesminister für Forschung und Technologie berät), wird dieser Schritt in anderen Mitgliedsländern der Gemeinschaft nicht so leicht zu verwirklichen sein, da nicht überall derartige Verfahrensmöglichkeiten bestehen. Nach dieser Phase der länderseitigen Information, Diskussion und Abstimmung wird die Kommission einen Projektantrag an den Rat der Gemeinschaft stellen. Es ist gegenwärtig noch offen, ob dabei volle Projektförderung durch die EG beantragt oder nur eine Teilförderung (etwa 50 %) vorgesehen werden soll; im zweiten Fall könnten die Länder ggf. unmittelbar "ihre" Forschungszentren mitfinanzieren und über diesen Weg stärkeren Einfluss nehmen auf den Umfang und die Zielsetzung der Entwicklungsarbeiten im eigenen Land. Diese Frage bedarf weiterer Diskussion und sei hier offen gelassen.

4. Einige ergänzende Bemerkungen zum Verständnis des King-Entwurfs (Fassung vom 25.09.)

Dieser Entwurf ist als ein erstes Diskussionspapier zu verstehen, das von Frau King auf der Grundlage von Diskussionen in der sog. Koordinierungsgruppe (Projektgruppe) und der Expertengruppen erstellt wurde. Er kann im Rahmen der Diskussion in der Bundesrepublik Deutschland nur zur Grobinformation benutzt werden. Einige Teile (insbesondere die Teile 4 - 10) waren zum Entwurfszeitpunkt noch nicht diskutiert (mittlerweile ist dies geschehen zu den Teilen 4 und 5, eine revidierte Fassung wird an die Mitglieder des Ad-hoc-Ausschusses verteilt, sobald sie vorliegt).

Der Punkt 8 (Management) wurde mit mir vorab diskutiert, ich habe dabei die Alternative B in die Diskussion gebracht; gerade zu diesem Punkt wird z.Z. von der Kommission eine Studie erarbeitet.

Punkt 6 erscheint mir in dieser Form völlig indiskutabel, vergleiche dazu meinen Vorschlag in Abschnitt 5 dieses Papiers (dies gilt auch weitgehend für die Zeitplanung in Punkt 9). Insgesamt sollte der King-Entwurf als Einführung in die Problematik verstanden werden; hier wird sich noch einiges ändern.

5. Vorschlag für einen Stufenplan zur Entwicklung von EUROTRA

(Dieser Teil ist - in leicht abgewandelter Form - von mir inzwischen als EUROTRA-Papier Nr. 57/78 den Mitgliedern der Koordinierungsgruppe zur Verfügung gestellt worden).

Die Vorstudien und Vorüberlegungen zu EUROTRA, vornehmlich durch die von CETIL einberufene Koordinierungsgruppe und die Expertengruppen, haben gezeigt, dass die Entwicklung eines computergestützten Übersetzungssystems in EUROPA mit der Möglichkeit weiterer Anwendungen in Dokumentation und Information auf eine realistische Basis gestellt werden kann. Andererseits wurde auch deutlich, dass ein von der Gemeinschaft und den Mitgliedsländern gemeinsam getragenes Entwicklungsprojekt weiterer grundlegender und intensiver Planung und Betreuung bedarf, die auf der Ebene von Expertengesprächen und (ad-hoc-)Papieren allein nicht geleistet werden kann. Allenfalls kann auf diese Weise ein konzeptioneller Rahmen geschaffen werden, selbst dann, wenn man sich - wie geschehen - in einer Reihe von Grundsatz- und Detailfragen verständigen kann. Eine Anzahl weiterer Fragen muss jedoch - auch wegen des damit verbundenen Aufwands - in diesem Stadium der Überlegungen offen bleiben, Fragen, die für Erfolg oder Misserfolg des Projekts entscheidend sein können. Beispiele dafür sind:

- (a) die Wirksamkeit der Koordination und die Kommunikation der zu beteiligenden Institutionen
- (b) die Effizienz des Projektmanagements
- (c) das Erreichen einer einheitlichen Kompetenz (methodisch und fachlich) der zu beteiligten Stellen.
- (d) der Aufbau realistischer Anwendungssituationen
- (e) das Problem der Integration vorhandener Systeme oder Systemteile (i.S. einer Einbringung entsprechender Software bzw. auch vorhandenen Methodenwissens)
- (f) die Orientierung am Bedarf (z.B. Auswahl der Fachgebiete, der Sprachpaare, der Dokumentarten)
- (g) die (praktisch) notwendige Tiefe (Korrektheit) der Sprachanalyse /-übersetzung (z.B. rohe - aber billigere - Übersetzung; feine - aber teurere - Übersetzung)
- (h) die Gesamtentwicklungskosten des Projekts (sie hängen maßgeblich von den aktuellen Parametern ab, die in den Bereichen (d), (e), (f) und (g) definiert werden).

Aus diesen Gründen wird vorgeschlagen, die Entwicklung von EUROTRA in mehreren Stufen (Phasen) durchzuführen, die jeweils durch Meilensteine derart bestimmt sind, dass der Projektverlauf und die Projektkosten den (Zwischen-)Ergebnissen und zwischenzeitlich auftretenden Anforderungen und Erkenntnissen (z.B. Fachgebiete, Analysetiefe/Lexikoninformationen) angepasst werden können.

Es wird unterschieden zwischen einer Initialisierungsphase (EUROTRA-0, etwa 2 Jahre), einer Aufbauphase (EUROTRA-1, etwa 3 Jahre) und einer Ausbauphase (EUROTRA-2, ca. 2 Jahre).

Im Folgenden werden einige wesentliche Aufgaben der Initialisierungsphase - skizzenhaft - behandelt:

EUROTRA-0 (Initialisierungsphase)

4 Teilaufgaben (TA)

- TA 1: Formulierung bzw. Präzisierung von Pflichtenheften (1. Halbjahr)
- TA 1.1: Präzisierung der (durchaus unterschiedlichen) Anforderungen auf Benutzerseite im Hinblick auf die Integration des Systems in die konkrete Übersetzungsarbeit in Verbindung mit möglichen Pilotanwendern, z.B. DG IX; in der Bundesrepublik: Bundessprachenamt, FIZ(e), Siemens-Übersetzungsabteilung, Institut für Übersetzen und Dolmetschen der Universität des Saarlandes.
- TA 1.2: Softwarespezifikation
- TA 1.3: Textauswahl, Testbedingungen für Phase 0

TA 2: Integration (Verwertung) von vorhandenen MT-Systemen (z.B. SYSTRAN) im Rahmen (!) von EUROTRA als 1. zumindest provisorische, (Teil-)Aufgaben der Benutzer abdeckende Komponenten; zugleich Weiterentwicklung dieser Komponenten im Sinne dieser Integration, allerdings ohne allzu großen weiteren Kostenaufwand (d.h. keine Erweiterung vorhandener Systeme um andere Sprachpaare oder Sachgebiete, allenfalls Abrundung durch Lexikonausbau und Korrektur von Fehlern). Ergebnis: EUROTRA-0 bringt bereits benutzer-relevante Leistungen (Problem: wenn diese Leistung unbefriedigend ausfällt, sind u.U. Aversionen gegen MT zu befürchten). Bei der Realisierung der Neukonzeption kann zu Beginn eine Konzentration auf Sprachpaare und Sachgebiete erfolgen, die nicht durch vorhandene anwendungsfähige Systeme abgedeckt sind ((¹))
(1 Jahr)

TA 3: Realisierung von Teilkomponenten des "neuen" Systems (2./3. Halbjahr)

TA 3.1 Software (1. Stufe)

TA 3.2 Korpus-Lexika (Test; Aufwands- und Effizienzermittlung)

TA 3.3 Beginn der Realisierung der "neuen" Systemkomponenten (Modulen) in ausgewählten Bereichen

TA 4: Realisierung von Prototypen des Gesamtsystems (für alle Sprachen mit je 1-2 Sprachenpaaren in 1-2 Fachgebieten auf 1-2 "Niveaus" (d.h. mit unterschiedlichen Verarbeitungsgraden) (2./4. Halbjahr)

Balkenplan:

| | 1. Jahr | 2. Jahr |
|-----|------------------|------------------|
| TA1 | XXXXXXXXXXXXXXXX | |
| TA2 | XXXXXXXXXXXXXXXX | |
| TA3 | | XXXXXXXXXXXXXXXX |
| TA4 | | XXXXXXXXXXXXXXXX |

Zum Ende der Initialisierungsphase können somit u.a. folgende Probleme geklärt sein:

- Koordination der Arbeiten/Management
- voraussichtliche Kosten des Gesamtsystems
- Integrationsmöglichkeiten vorhandener Systeme
- Marktbedürfnisse + Kosten-Nutzen-Verhältnis
- Schwerpunkte der weiteren Entwicklung

Die Initialisierungsphase ist im Finanzierungsbedarf verhältnismäßig gut zu überblicken; verschiedene hier nicht näher erläuterte Komponenten (z.B. Einsatz von Hardware, Test- und Bewertungsverfahren, andere Anwendungsbereiche) können parallel dazu verfolgt werden.

ANMERKUNGEN:

(¹) Einen entsprechenden Vorschlag machte mir kürzlich Dr. Krollmann.