

EUROPÄISCHE HERAUSFORDERUNGEN

von Wolfgang Teubert

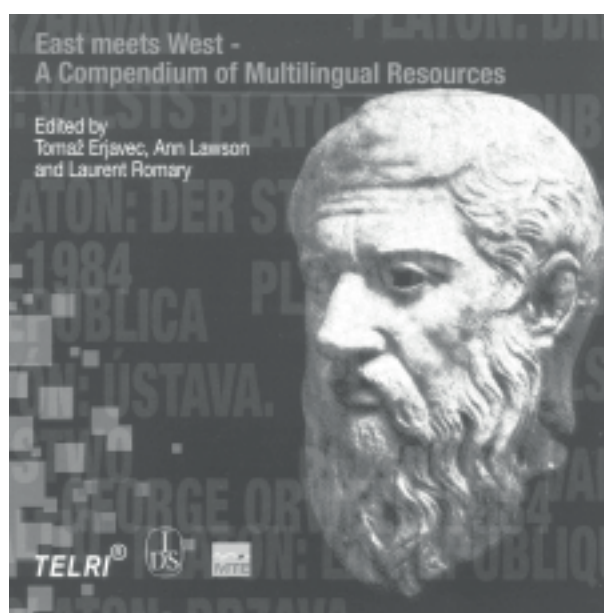
Deutsch ist keine isolierte Sprache. Seine heutige Gestalt ist von anderen europäischen Sprachen beeinflusst. Eine jahrhundertelange Auseinandersetzung mit antiken, italienischem, französischem und schließlich englischem Weltverständnis bescherten uns mit einem abendländischen Begriffsgefüge auch einen gemeinsamen Wortschatz, der sich vielfach von griechischen und lateinischen Wurzeln ableitet und willkommene Brücken zu anderen Sprachen baut. Diesem »lessico intellettuale europeo« (so der Titel eines europäischen Langzeitprojekts) verdankten frühere Bildungseliten die Leichtigkeit gegenseitigen Verständnisses in einem durchaus polyphonen europäischen Diskurs, in dem kulturelle und sprachliche Identitäten im nationalen und regionalen Rahmen gewahrt waren. So soll es auch in einem vereinigten Europa bleiben. Ein demokratisches Europa beruht auf der Akzeptanz unterschiedlicher Kulturräume bei gleichzeitiger Anerkennung allgemeinverbindlicher Diskursregeln. Funktionieren kann das nur, insoweit es gelingt, Mehrsprachigkeit zu verallgemeinern. Hier kommen auf die nationalen Sprachinstitute in Europa neue Aufgaben zu.

In einem Europa gleichberechtigter Bürger müssen die Sprachgrenzen für alle, und nicht nur für die Eliten, durchlässig werden. Das Ziel ist erreicht, wenn jeder Europäer sich unabhängig von seiner Muttersprache mit seinen Mitbürgern überall in Europa verständigen, am politisch-gesellschaftlichen Leben teilnehmen und sich mit Informationen versorgen kann. Mehr und besserer Sprachunterricht tut Not. Gleichzeitig hängt die Verwirklichung einer europäischen Informationsgesellschaft aber auch von der Entwicklung multilingualer Sprachtechnologie ab, die hilft, Wissen verfügbar zu machen, fremdsprachige Texte zu verstehen und Texte in fremden Sprachen zu entwerfen. Bei allem Fortschritt liegt noch ein weiter Weg vor uns, gewiss viel weiter, als uns die Gurus künstlicher Intelligenz und automatischer Übersetzung nicht müde werden zu versichern.

Die Computer haben inzwischen zwar die Regeln von Syntax und Grammatik gelernt, aber sie verstehen aus

prinzipiellen Gründen nicht, was Wörter bedeuten und was Sätze sagen. Doch Verstehen ist essenziell. Denn Menschen können keine Texte übersetzen, die sie nicht verstehen. Vorhandene zweisprachige Wörterbücher, wie dick sie auch sein mögen, enthalten nur einen Bruchteil des nötigen Wissens. Übersetzen kann mit ihnen nur, wer in der einen Sprache zu Hause ist und die andere wenigstens mäßig kennt. Computer sind indessen gewiss nicht klüger als wir. Um sie fit zu machen, müssen sehr viel mehr Sprachdaten gesammelt, d.h. aus allen verfügbaren Sprachressourcen herausgefiltert werden. Nur so

kann ihr mangelndes Sprachverständnis kompensiert werden. Für die Datengewinnung benötigen wir »intelligente« Computerprogramme – der menschliche Arbeitsaufwand wäre schlicht unbezahlbar. Monolinguale Analysen, wie sie heute vielfach noch das Bild prägen, sind nur der erste Schritt. Voller ungelöster Probleme ist der weitaus wesentlichere Teil der Aufgabe, nämlich die Daten einer Sprache mit den Daten anderer Sprachen zu vernetzen. Bei der gebotenen Größenordnung muss auch hier der Computer dem Menschen die Arbeit abnehmen.



Platons Staat: CD-ROM mit 20 Übersetzungen

Die Aufgabe, multilinguales Sprachwissen aus ein-, zwei-

und mehrsprachigen Textkorpora zu extrahieren und es für die künftige, bedeutungsverarbeitende Generation von sprachtechnologischer Software aufzubereiten, ist viel zu groß, als dass Sprachinstitute sie als Einzelkämpfer in Angriff nehmen könnten. Wir brauchen eine europäische Infrastruktur, die dem Wissensaustausch und der Zusammenarbeit zwischen Forschung und Industrie, zwischen Linguistik und Informatik, zwischen großen und kleinen Sprachen dient. Schon vor Jahren hat die Europäische Kommission die Dringlichkeit erkannt; sie fördert seither in mehreren Programmen einschlägige Maßnahmen. Das IDS ist an wichtigen Projekten maßgeblich beteiligt.

Die Grundlage für die heutigen Projekte wurde 1991-1993 in dem Projekt NERC (*Network of European Reference Corpora*) gelegt, in dem das IDS einer von sechs Partnern war. Hier wurden die Blaupausen für die derzeitigen Infrastrukturmaßnahmen erarbeitet. Erstes konkretes Projekt war PAROLE, in dem es um die Erstellung vergleichbarer textualer Ressourcen (je 20 Mil-

tionen Wörter) und lexikalischer Ressourcen (je 20.000 Wörter) für alle EG-Amtssprachen in strikt einheitlicher Darstellung geht. Dieses Projekt wurde im Frühjahr 1998 abgeschlossen. Im Anschlussprojekt SIMPLE, das soeben begonnen hat, geht es nun um die Vernetzung dieser Wörter auf der Bedeutungsebene. Das IDS bearbeitet die deutschen Sprachdaten nicht allein; es sucht vielmehr die Zusammenarbeit mit anderen Instituten in Deutschland, Österreich und der Schweiz. Gleichzeitig pflegt es Kontakte zu anderen einschlägigen Projekten in Europa und Übersee. Das IDS ist ferner einer der beiden Partner des neuen EU-Projekts ELAN (*European Language Activities Network*), das normierte Sprachressourcen mit plattform-unabhängigen Abfragesprachen weltweit zugänglich macht.

Seit 1995 wird am IDS das einzige gesamteuropäische Infrastrukturprojekt für Sprachressourcen koordiniert: TELRI, *Trans-European Language Resources Infrastructure*. Zur Zeit sind Institute in etwa dreißig Ländern (einschließlich der früheren Sowjetunion) aktiv an diesem Projekt beteiligt. TELRI bietet mit vielfältigen Aktivitäten ein Forum für den Austausch von Daten, Verfahren und Know-how, fördert den Aufbau von Sprachressourcen und deren Auswertung, bildet junge Wissenschaftler aus und entwirft innovative Projekte zur multilingualen Vernetzung von Sprachdaten. Seminare, Kurse, aktuelle Webseiten ebenso wie traditionelle Publikationen sind Teil eines wachsenden Angebots an Dienstleistungen für Partner in Forschung und Industrie. TRACTOR (*TELRI Research Archive for Computational Tools and Resources*) sammelt und verbreitet einzel- und mehrsprachige textuale und lexikalische Ressourcen sowie Computerprogramme, die aus diesen Ressourcen Sprachwissen extrahieren.

Die Finanzierung von TELRI ist bis ins Jahr 2001 mit Mitteln der EU gesichert. Ein viel gefragtes Ergebnis der bisherigen Arbeit ist eine CD-ROM mit über 20 Übersetzungen von Platos *Staat*, wobei diese Übersetzungen satzweise miteinander verknüpft sind. Wer sich dafür interessiert, womit das deutsche Wort *Arbeit* in anderen europäischen Sprachen gleichgesetzt wird, kann die Belege dieses Wortes zusammen mit den parallelen Stellen in anderen

Sprachen auf den Bildschirm holen (siehe Schaubild). In diesen Paralleltexten findet sich manche Entsprechung, die man in zweisprachigen Wörterbüchern vergeblich suchen würde. Parallelkorpora enthalten implizit das Sprachwissen ganzer Generationen von Übersetzern. Dieses Wissen wird die wesentliche Grundlage des Fortschritts in maschinellen Übersetzungshilfen sein. Es muss so aufbereitet werden, dass die Computerprogramme es verarbeiten können. Die Verfahren dafür liefert die Korpuslinguistik. Diese ist mehr als nur eine Methode sprachwissenschaftlichen Arbeitens. Sie ermöglicht anders als traditionelle Ansätze eine verstehensunabhängige

ARBEIT

DEUTSCH	ENGLISCH	FRANZÖSISCH
Wie die Dichter ihre Werke und die Väter ihre Kinder lieben, so ereifern sich die Geldeute für ihr Geld, weil es ja der Erfolg ihrer Arbeit ist; und dann wegen seines Nutzens, wie die anderen Leute.	For just as poets feel complacency about their own poems and fathers about their own sons, so men who have made money take this money seriously as their own creation and they also value it for its uses as other people do.	En effet, de même que les poètes chérissent leurs poèmes, et les pères leurs enfants, de même aussi ceux qui se sont enrichis prennent au sérieux leurs richesses, comme étant leur oeuvre propre (ils y tiennent aussi en raison de leur utilité, comme les autres).
Ich danke dir, mein Bester! Sag mir noch das: Wenn das Werk der Ungerechtigkeit darin liegt, überall, wo sie ist, Zwietracht zu säen, dann wird sie bei Freien und Sklaven gegenseitigen Streit und Haß bewirken und sie zu gemeinsamer Arbeit unfähig machen, nicht?	That is good of you, my friend; but tell me this: if it is the business of injustice to engender hatred wherever it is found, will it not, when it springs up either among freemen or slaves, cause them to hate and be at strife with one another, and make them incapable of effective action in common?	Mais c'est bien bon à toi de procéder ainsi, homme excellent, dis-moi alors ceci : si tel est donc l'effet de l'injustice, d'introduire la haine partout où elle se trouve, est-ce que, quand elle s'introduit aussi bien chez les hommes libres que chez les esclaves, elle ne les fera pas se haïr les uns les autres, entrer en dissension, et être incapables d'agir en commun les uns avec les autres?
Aber mit nichts so gut wie mit einem Winzermesser, das für diese Arbeit gefertigt ist?	But nothing so well, I take it, as a pruning knife fashioned for this purpose.	Mais avec aucun, je crois, aussi bien qu'avec la serpette fabriquée à cet effet.
Wenn nun ein Bauer seine Erzeugnisse zum Markt schafft oder sonst ein Handwerker und sie kommen nicht zur gleichen Zeit wie die Käufer ihrer Ware, dann versäumen sie ihre eigene Arbeit, während sie auf dem Markte sitzen, nicht?	If, then, the farmer or any other craftsman taking his products to the market place does not arrive at the same time with those who desire to exchange with him, is he to sit idle in the market place and lose time from his own work?	Alors le cultivateur, ou encore l'un des artisans, qui a apporté sur la place publique une partie de ce qu'il produit, s'il n'y vient pas au même moment que ceux qui ont besoin d'échanger contre ce qu'il fournit, restera-t-il assis sur l'agora, laissant en sommeil son activité d'homme au service du public?
Denn andere Leute wieder, die dies sehen, bieten sich zur Hilfe an; in richtig geleiteten Staaten zumeist die Schwächlichen, die zu keiner anderen Arbeit taugen.	But there are men who see this need and appoint themselves for this service—in well conducted cities they are generally those who are weakest in body and those who are useless for any other task.	Il y a des hommes qui, voyant cela, se fixent à eux-mêmes cette charge ; dans les cités correctement administrées ce sont en général les hommes aux corps les plus faibles, impropres à toute autre fonction.

Darstellung und Verknüpfung sprachlicher Daten und Phänomene, wie sie Voraussetzung für jegliche Sprachtechnologie ist. Die am IDS herausgegebene Zeitschrift *International Journal of Corpus Linguistics* widmet sich der Weiterentwicklung dieses relativ jungen theoretischen Programms.

Dem IDS kommen diese Projekte auch für seine eigene Forschung zugute. Die Extraktion von Sprachwissen aus Korpora und seine Aufbereitung in lexikalischen Datenbanken profitiert von der Zusammenarbeit mit Partnerinstituten, die darin bereits Erfahrungen gesammelt haben. Zur Zeit wird ein deutsch-französisches Arbeitsvorhaben vorbereitet, bei dem aus einem Parallelkorpus automatisch Übersetzungseinheiten und ihre Entsprechungen in der anderen Sprache extrahiert und Über-

setzungshilfen aufbereitet werden. Solche Übersetzungseinheiten sind, wie jeder, der schon einmal übersetzt hat, weiß, meist erheblich größer als das Einzelwort, das die Grundeinheit für traditionelle Wörterbücher bildet. Die moderne Lexikographie, am IDS und anderswo, trägt dieser neuen Sichtweise Rechnung.

Der Autor ist wissenschaftlicher Mitarbeiter am Institut für deutsche Sprache in Mannheim.