

## Computergestützte Fachlexikographie mit dem Programmsystem EWF

Gunter Neubert

Durch die Einführung informationstechnischer Mittel und Verfahren in die Praxis des Übersetzens können zahlreiche Routinearbeitsgänge rationeller ausgeführt werden, auch die Abfrage von Wortschätzen, sofern sie computergespeichert vorliegen. Es zeigt sich aber immer deutlicher, daß gerade hier die entscheidende Engstelle liegt – die Erarbeitung und die kontinuierliche Laufendhaltung der terminologischen Informationsbasen ist ein intellektueller Arbeitsprozeß, der hohes Fach- und Sprachwissen erfordert und dessen Aufwand seitens der Auftraggeber von Übersetzungen, ja sogar seitens der Leiter von Übersetzungsabteilungen und seitens der Übersetzer selbst häufig noch immer nicht angemessen eingeschätzt wird. Übersetzer sind zwar unter bestimmten Voraussetzungen – u.a. Vorhandensein von Fachliteratur in den betreffenden Sprachen, Möglichkeit der Konsultation von Fachleuten mit Fremdsprachenerfahrung, Möglichkeit der Konzentration auf ganz wenige fachliche Gebiete bzw. Gegenstände – zur Eigenversorgung mit Terminologie in der Lage, aber auf jeden Fall nur für einen verhältnismäßig engen fachlichen Bereich. Fremdversorgung ist notwendig, und zwar:

- (1) auf dem Wege des wechselseitigen Austauschs insbesondere des aktuellen und/oder speziellen Wortguts z.B. mittels vernetzter Terminologiebanken oder zumindest durch Anschluß an leistungsfähige Banken;
- (2) seitens der Fachlexikographie mittels Wörterbüchern in gedruckter oder computergespeicherter Form.

Das gedruckte Wörterbuch scheint in der näheren Zukunft seine Bedeutung noch keineswegs einzubüßen, nicht nur wegen seiner Portabilität, sondern auch, weil der Benutzer terminologisch nützliche Informationen auf assoziativ, empirisch oder sogar aleatorisch gesteuerten Wegen zu finden vermag, die schwer oder nicht algorithmierbar sind. Vor einigen Jahren – zu Beginn der Einführung der computergestützten Fachlexikographie – sind dazu in Dresden Untersuchungen durchgeführt worden, indem mehrere Übersetzer ein und denselben Text mit Hilfe ein und desselben Wörterbuchs (ein Chemiewörterbuch Englisch–Deutsch mit etwa 60.000 Wortstellen), jedoch einmal als gedrucktes Buch und das anderemal als Computerwortschatz übersetzten. Es ergaben sich *Antwortquoten* (eine Frage an das Wörterbuch galt als beantwortet, wenn für die Wahl eines zielsprachigen Terminus ausreichende Informationen gefunden worden waren, also nicht nur dann, wenn der gesuchte Terminus tatsächlich im Wörterbuch stand) von etwa 80% beim gedruckten und etwa 37% beim computergespeicherten Wörterbuch.

Obwohl die Untersuchungen u.a. wegen viel zu kleiner Stichprobe strengen wissenschaftlichen Maßstäben nicht entsprachen, scheint die erhaltene Aussage dennoch in erster Näherung richtig, wie jüngere Erfahrungswerte zeigen.

### Das Terminologiebank-Programmsystem EWF

Bemühungen um eine Rationalisierung terminographischer Tätigkeiten sollten neben Terminologiediensten, worunter wir wohl in erster Linie auf aktuelles und spezielles Wortgut orientierte Fundi zu verstehen haben, auch die traditionelle *Fachlexikographie* einbeziehen, die vorrangig auf die Erfassung des terminologischen Grundstocks breiter, durchaus aber auch spezieller Disziplinen gerichtet ist. Bei der Schaffung des Terminologiebank-Programmsystems EWF (das ist das Akronym von "Elektronisches Wörterbuch der Fachsprachen") an der Technischen Universität Dresden mußte dieser Aspekt von vornherein berücksichtigt werden, bestand doch (und besteht noch, inzwischen seit mehr als 30 Jahren) eine kleine Entwicklungsstelle für Fachwörterbücher, an die bei gleichbleibendem Personalbestand ständig steigende Forderungen gestellt wurden. Diese ergaben sich ganz zwangsläufig aus der Notwendigkeit, die geschaffenen Wörterbücher in bestimmten Abständen überarbeiten zu müssen. Betreut werden gegenwärtig drei Wörterbücher der CHEMIE UND CHEMISCHEN TECHNIK, nämlich ENGLISCH-DEUTSCH, DEUTSCH-ENGLISCH und RUSSISCH-DEUTSCH, ein Wörterbuch der ELEKTROTECHNIK/ELEKTRONIK RUSSISCH-DEUTSCH, jeweils 50–60.000 Wortstellen, ein Wörterbuch der SPANENDEN WERKZEUGMASCHINEN UND WERKZEUGE RUSSISCH-DEUTSCH sowie einige kleinere abgeleitete Wörterbücher, alle herausgegeben vom VEB Verlag Technik Berlin.

Das grundlegende Merkmal der *Speicherstruktur* des EWF-Systems ist – wie bei den meisten gleichartigen Projekten – die Adressierbarkeit aller Informationselemente. Als Informationselemente werden die Termini und die Begleitinformationen sowie die einzelnen Zeichen alphanumerisch eingetragener Informationen aufgefaßt. An Begleitinformationen sind u.a. möglich: Erklärung, Definition, chemische Formel, Nomenklaturbenennung, Fachgebietsangaben, Gebräuchlichkeitsangaben, Genus/Numerus, die Sprache. Neben diesen aus Wörterbüchern gewohnten Begleitinformationen sind außerdem Plätze vorhanden für Angaben wie Lexikograph, Redakteur, Erarbeitungsdatum, Quelle u.ä., die gewöhnlich nicht im Wörterbuch erscheinen, aber für die lexikographische Technologie wichtig sind. Ein Platz dient der Abfragezählung – schließlich weiß kaum ein Lexikograph, in welcher Häufigkeit die von ihm ins Wörterbuch geschriebenen Artikel nachgeschlagen werden. Einige weitere Begleitinformationen ergaben sich aus der ökonomisch diktierten Notwendigkeit, das System und die einmal erfaßten Wortschätze so vielseitig wie irgend möglich nutzen zu können. Berücksichtigt wurde insbesondere die Möglichkeit des Aufbaus mehrsprachiger Thesauren. Einzelheiten können hier nicht beschrieben werden; es sei insbesondere hinsichtlich der

inhaltlichen Konzeption der einzelnen Begleitinformationen auf Neubert 1981 verwiesen.

### **Das Begriffsstellenprinzip**

Ein Blick auf die Informationselemente eines Wörterbuchartikels läßt rasch erkennen, daß sie (mindestens) zwei Gruppen zugeordnet werden können: Informationen, die für den einzelnen Terminus gelten (z.B. eine Angabe "USA" oder das Genus), und Informationen, die für alle diejenigen Termini des Artikels gelten, die dieselbe Bedeutung haben (z.B. die Definition, die Fachgebietsangabe). Ein Lexikograph ist ferner bestrebt, den Wortschatz eines bestimmten Fachgebiets möglichst mehrfach auszunutzen, in unserem Falle also z.B. für alle drei Chemiewörterbücher. Aus beiden Gründen erscheint es zweckmäßig, die traditionelle lexikographische Einheit der "Wortstelle" zugunsten einer sog. "Begriffsstelle" aufzugeben. In einer *Begriffsstelle* sind alle die Termini gespeichert, die (in einem Festlegungen durchaus zugänglichem Grad) gleiche Bedeutung haben. Das bedeutet einerseits, daß die Termini einschließlich sämtlicher Synonyme aller bearbeiteten Sprachen in einer lexikographischen Einheit vereint sind, und andererseits, daß Polysemie bzw. Homonymie mehrere lexikographische Einheiten zur Folge hat. Die Unterscheidung Ausgangs-/Zielsprache wird gegenstandslos. Inwieweit dieses Prinzip allerdings durchsetzbar ist, hängt davon ab, wie streng die – am Ende doch nur über die Termini erfaßbaren – Begriffe definiert sind bzw. definiert werden können. Zwischen dem Lexem der Gemeinsprache und dem Terminus im engeren Sinne existiert weder in realisierter Sprache (im Fachtext!) noch im Inventar eine feste Grenze, und im Übergangsbereich sind Wörter zu finden, die zwar mit deutlichem Fachbezug gebraucht werden, jedoch keine im Sinne des betreffenden Fachgebiets streng definierten Begriffe benennen. Für solche Wörter lassen sich folgerichtig auch keine Begriffsstellen nach der oben entwickelten Vorstellung aufbauen. Das EWF-System berücksichtigt dies, indem für eine konkrete Datei Einträge gestattet sind, die nur in einer Sprachbeziehungrichtung benutzt werden dürfen; Ausgabe in der anderen Richtung wird blockiert. Übrigens sind auch durchweg einsprachige Einträge verarbeitbar, so daß beispielsweise der Wortschatz eines Lehrbuchs Deutsch für Ausländer während der Erarbeitung des Lehrwerks aufgenommen und zur Vorbereitung der Übungen nach den verschiedensten Aspekten klassifiziert ausgegeben werden konnte (s. Neubert 1981).

### **Computertechnologie der Fachlexikographie**

Unmittelbar zum Programmsystem gehören Programme für Einspeicherung und Änderung im Dialog- und Stapelbetrieb, für Abfrage ebenfalls im Dialog- und

Stapelbetrieb und für Klassifizierung der Begriffsstellen nach dem Inhalt der Begleitinformationen und nach den die Benennung konstituierenden Zeichen bzw. Zeichenketten.

Traditionell ist die lexikographische Arbeit in der Sicht unserer Fachlexikographen in *vier Arbeitsgänge* unterteilt: die Kompilation, die Redaktion, die Korrektur des Satzmanuskripts und die Korrektur der Druckfahnen. Die Kompilation umfaßt die Selektion der relevanten Termini aus der Literatur, die Klärung ihrer Bedeutung und die Zuordnung der Äquivalente. Es entstehen eine Anzahl von Einzelkarteien bzw. -dateien für die Untergebiete der Disziplin, die in der Redaktionsphase zusammengeführt werden. Die dabei anfallenden Routinearbeiten, wie alphabetisches Sortieren, Mischen, Aufbereiten für den Satz, Verweiskontrolle, werden vom Computer rasch und zuverlässig ausgeführt. Auf die Unterstützung der intellektuellen Prozesse der Terminusfindung und Bedeutungsklä rung etwa mit Hilfe von Konkordanzen wurde nach einigen Experimenten wieder völlig verzichtet. Die zahlreichen und oft umfangreichen Quellenwerke – Monographien, Zeitschriftenartikel, Vortragsdrucke, Nachschlagebücher u.a.m. – liegen weder maschinenlesbar vor, noch lohnt sich die Überführung in den Computer; denn das Erschließen der fachlexikographisch relevanten Informationen verläuft auf verschlungenen Pfaden mit vielen “Seitensprüngen”, für die derzeit Rechnerunterstützung nur ausnahmsweise denkbar ist.

Sehr bedeutsam aber sind die Möglichkeiten, Arbeitsschritte bei der *Überarbeitung* eines Wörterbuchs für eine Neuauflage durch den Computer ausführen zu lassen, auf die bisher oft ganz verzichtet werden mußte, weil keine ausreichende personelle Kapazität vorhanden war oder weil sie ökonomisch nicht zu rechtfertigen waren. Einige Beispiele:

- Die in einem größeren Fachwörterbuch vereinten Wissensgebiete entwickeln sich unterschiedlich rasch und müssen für eine Neuauflage unterschiedlich intensiv (einige stark, andere überhaupt nicht) überarbeitet werden. Per Computer lassen sich mit Hilfe der einschlägigen Begleitinformationen leicht fach- bzw. sachgebietsbezogene Subdateien exzerpieren, die gezielt durchgearbeitet werden können.
- Analog wird verfahren, wenn ein grundlegendes Quellenwerk fachlich aktualisiert erschienen ist. Wiederum wird, diesmal mit Hilfe der Begleitinformation Quelle, eine Subdatei exzerpiert, bearbeitet und wieder eingemischt.
- Bei der herkömmlichen lexikographischen Technologie war es sehr mühevoll, den einzelnen Autoren der früheren Wörterbuchaufgabe genau “ihre” Teilwortschätze zur Überarbeitung zurückzugeben.
- Die Wörterbuchartikel können jetzt auch abhängig von ihrem Erarbeitungsdatum durchgesehen werden.
- Es können sämtliche Vorkommen einer bestimmten Schriftzeichenkette nicht nur am Terminusanfang, sondern an beliebiger Stelle innerhalb des Terminus und sogar innerhalb der Erklärungen bzw. Definitionen aufgefunden und zur

Überarbeitung bereitgestellt werden. Das ist z.B. bei Nomenklaturumstellungen wie kürzlich für die Chemiewörterbücher aufgrund von IUPAC-Regelungen erforderlich (jetzt *cobalt* anstelle *kobalt*, *alcohol* anstelle *alkohol*, *eth* anstelle *äth*, *bismut* anstelle *wismut* usw.).

Mit dieser Aufzählung sollen weder sämtliche Möglichkeiten der Computerunterstützung des Fachlexikographen genannt sein, noch können wir zu behaupten wagen, sie überhaupt zu kennen. Die computertechnischen Mittel entwickeln sich rasch und werden immer kostengünstiger und leichter verfügbar. Sicher akkumuliert der sie nutzende Lexikograph Erfahrungen, die es ihm gestatten, gleichsam Aufträge an die Sprachwissenschaft zu formulieren, ihre Ergebnisse in Form programmfähiger Algorithmen zu seiner Unterstützung vorzulegen – ein neues, weites Feld der "sprachtechnischen" Verifizierung linguistischer Erkenntnisse.

### Erhöhung der Antwortquote

Übrigens ergeben sich aus den Nutzungserfahrungen schon jetzt Schlußfolgerungen für die Terminologierepräsentation im Wörterbuch selbst, insbesondere wenn dieses per Computer abgefragt wird. Zahlreiche Wortbildungsprodukte entstehen beim Abfassen des Texts durch den Fachmann. Der Übersetzer sucht sie im Wörterbuch oft vergeblich, weil sie dem Lexikographen entweder nicht begegnet sind oder von ihm nicht als aufnahmewürdig erachtet worden sind. Wie eingangs erwähnt, vermögen die Wörterbuchbenutzer mittels ihrer (allerdings recht unterschiedlich ausgeprägten) Kenntnis der Wortbildungsregeln dennoch verwertbare Informationen aus dem Wörterbuch zu gewinnen. Es erscheint sinnvoll, einerseits Wörterbüchern detailliertere, sprachwissenschaftlich basierte Benutzungshinweise voranzustellen und andererseits die Abfrageprogramme computergespeicherter Wörterbuchversionen um Prozeduren auf der Basis der Wortbildungsregeln zu ergänzen, damit eine von den subjektiven Nutzervoraussetzungen unabhängige, größtmögliche Antwortquote erreicht wird. Die formalen Regeln scheinen in dieser Hinsicht bereits zumindest teilweise algorithmierbar zu sein (z.B. die Aufspaltung von Komposita, die Rückführung von Derivativa auf ihre Basen, die Wandlung zwischen Partizipien und Infinitiven); für die semantischen Wortbildungsregeln sind Ansätze vorhanden (z.B. unter Ausnutzung des Prinzips der morphematisch-semantischen Motivation, von Analogieprinzipien) (Neubert 1985 a und 1985 b, Reinhardt 1984).

**Bibliographie***Zitierte Wörterbücher*

TECHNIK-WÖRTERBUCH CHEMIE UND CHEMISCHE TECHNIK ENGLISCH-DEUTSCH

H. Gross u.a., Berlin: Technik, 1984, 3. Aufl.

TECHNIK-WÖRTERBUCH CHEMIE UND CHEMISCHE TECHNIK DEUTSCH-ENGLISCH

H. Gross u.a., Berlin: Technik, 1983, 2. Aufl.

TECHNIK-WÖRTERBUCH CHEMIE UND CHEMISCHE TECHNIK RUSSISCH-DEUTSCH

H. Gross u.a., Berlin: Technik, 1987, 5. Aufl.

TECHNIK-WÖRTERBUCH ELEKTROTECHNIK/ELEKTRONIK RUSSISCH-DEUTSCH

H. Gross u.a., Berlin: Technik, 1982, 4. Aufl.

TECHNIK-WÖRTERBUCH SPANENDE WERKZEUGMASCHINEN UND WERKZEUGE RUSSISCH-DEUTSCH

H. Gross u.a., Berlin: Technik, im Druck.

*Sonstige Literatur*

Neubert, G. (Hrsg.) (1981), *Rechnerunterstützung bei der Bearbeitung fachlexikalischer Probleme*, Leipzig: Enzyklopädie.

Neubert, G. (1985 a), *Kenntnisse über terminologische Bildungselemente und -regeln – eine Voraussetzung für erfolgreiches Übersetzen wissenschaftlich-technischer Literatur*, in: Proc. X. World Congress of F.I.T., Wien: Braumüller, 281–285.

Neubert, G. (1985 b), *Motivation in Terminology Facilitating Networking*, in: Infoterm Series 8. Proc. 2nd Infoterm Symposium Wien 1985-04-14/17, München, New York, London, Paris: Saur, 542–545.

Reinhardt, W. (1984) und Neubert, G., (Hrsg.), *Das deutsche Fachwort der Technik – Bildungselemente und Muster*, Leipzig: Enzyklopädie.