

## Wortschatzarchitektur und elektronische Wörterbücher: Goethes Wortschatz und das Goethe-Wörterbuch

1	Einleitung	4.2.2	Vernetzung des elektronischen Goethe-Wörterbuchs
2	Zusammenhänge im Wortschatz und Aspekte der Wortschatzorganisation	4.2.3	Aktualisierbarkeit und Revidierbarkeit
3	Die Architektur des Goethe-Wortschatzes	4.2.4	Ein elektronisches Goethe-Wörterbuch als kostenlose Internet-Version
4	Ein elektronisches Goethe-Wörterbuch	4.2.5	Architektur und onomasiologische Erschließung
4.1	Mehrfachnutzung von SGML-Daten	4.3	Probleme und Hindernisse
4.2	Erweiterte Benutzungsperspektiven eines elektronischen Goethe-Wörterbuches	4.4	Prototypen
4.2.1	Dokumentationstiefen und Informationstypen	5	Zusammenfassung und Ausblick
		6	Literatur

### 1 Einleitung

Wer die alphabetische Ordnung des Wörterbuchs aufgibt, der versündigt sich an der Philologie, wetterte Jacob Grimm in die Richtung der ‚nach Wurzeln‘ geordneten Wörterbücher. Die alphabetische Ordnung eines Wörterbuchs geht an den internen Beziehungen im Wortschatz vorbei, sie ist nur praktisch, donnerte Hermann Paul zwei Generationen später und forderte, man möge auch begriffliche und andere Zusammenhänge bei der Wortschatzarbeit berücksichtigen. Zum Beispiel: *donnern* und *wettern* als rederekennzeichnende Verben stehen im Wörterbuch an völlig unterschiedlicher Stelle im Alphabet. Aber die beiden Verwendungsweisen gehören in mindestens zweierlei Hinsicht enger zusammen: zum einen sind die beiden Verben eng bedeutungsverwandt, zum anderen sind die beiden Verwendungsweisen zur Rederekennzeichnung wohl durch metaphorische Übertragung auf ähnliche Weise entstanden. Will man ein Wörterbuch drucken, dann muß man sich für einen einzigen Gesichtspunkt der Anordnung entscheiden, z.B. den Gesichtspunkt des Alphabets, den Gesichtspunkt der Bedeutungszusammenhänge oder den bisher nur in Registern erfaßten Gesichtspunkt der historisch-semanticen Charakteristik.

Daneben gibt es eine ganze Reihe von weiteren Gesichtspunkten der Wortschatzarchitektur und von Wortschatzzusammenhängen, die zu dokumentieren sich lohnt. Die Frage ist also nicht nur, ob man ein Wörterbuch alphabetisch oder onomasiologisch organisiert, sondern auch, ob und wie es gelingen könnte, eine Mehrzahl von lexikologischen Organisationsprinzipien bei der Dokumentation eines Wortschatzes gleichzeitig zu berücksichtigen. Im gedruckten Wörterbuch geht das nur eingeschränkt: Man muß sich für ein einziges Prinzip der Anordnung entscheiden und kann andere lexikologische Gesichtspunkte allenfalls durch erschließende Register einholen, ein Verfahren, das sich im ganzen als aufwendig, als wenig praktikabel und demzufolge auch als selten praktiziert erwiesen hat.<sup>1</sup>

Elektronische Datenbasen erlauben es nunmehr, auf einen Datenbestand, z.B. eine Wortschatzdokumentation, nach unterschiedlichen Kriterien zuzugreifen und dabei auch ganz unterschiedliche ‚Ansichten‘ des Datenbestandes je nach den Interessen und Fragestellungen von Benutzern hervorzubringen. Voraussetzung ist, daß die entsprechenden Informationen, z.B. zur lexikologischen Strukturierung, in expliziter Weise im Datenbestand enthalten sind. Ein Wörterbuch als elektronische Datenbasis mit mehrfacher Zugriffsmöglichkeit ändert damit auch seinen Charakter: Das gedruckte Wörterbuch ist im wesentlichen eine Ansammlung von Einzelwörtern bzw. ‚Einzelwortschicksalen‘ mit eingeschränkten Möglichkeiten, Zusammenhänge im Wortschatz aufzuweisen. Das elektronische Wörterbuch als komplex strukturierte lexikologische Datenbasis kann sehr viel mehr sein: ein Abbild der komplexen Zusammenhänge im Wortschatz und ein Informationssystem für sehr unterschiedliche Benutzerinteressen.

In den folgenden Abschnitten wird zunächst der Grundgedanke einer komplexen und mehrdimensionalen Wortschatzarchitektur erläutert, es folgt eine Fallstudie zum Wortschatz von Goethe und zum Goethe-Wörterbuch. Dabei werden neben einigen technisch-konzeptionellen Grundgedanken vor allem die erweiterten Nutzungsperspektiven eines elektronischen Wörterbuchs skizziert, das als elektronische Datenbasis erfaßt ist.

## 2 Zusammenhänge im Wortschatz und Aspekte der Wortschatzorganisation

Die grundlegende Einheit der Wortschatzstrukturierung ist die Verwendungsweise eines Wortes bzw. eines Lexems. Zwischen den einzelnen Lexemen bzw. ihren Verwendungsweisen bestehen vielfältige Zusammenhänge, und man kann die Architektur eines Wortschatzes als eine komplexe Ordnung bezeichnen. Die Komplexität der Wortschatzarchitektur beruht zum einen auf der Mehrzahl von lexikologischen Organisationsprinzipien und Strukturierungsgesichtspunkten, zum anderen auf der Tatsache, daß sich diese Organisationsprinzipien in zahlreichen Fällen auch kombinieren lassen. Und nicht zuletzt natürlich auch auf der großen Zahl der Elemente: der Wörter und Verwendungsweisen.

Ein lexikologisches Organisationsprinzip ist ein Gesichtspunkt, bei dessen Anwendung sich ganz bestimmte Zusammenhänge im Wortschatz ergeben. Nimmt man z.B. den Gesichtspunkt der Bedeutungsverwandtschaft, dann ergibt sich eine semantische Strukturierung des Wortschatzes als eine Art onomasiologisches Netz. Nimmt man den Gesichtspunkt der Zugehörigkeit von Wörtern bzw. Verwendungsweisen zu Fachgebieten oder Sachbereichen, dann ist das Ergebnis eine Fachgebiets- bzw. Sachbereichs-Strukturierung des betreffenden Wortschatzes. Wählt man unterschiedliche Gesichtspunkte der historischen Entwicklung, dann ergeben sich z.B. eine Herkunftsgliederung, ein Altersprofil oder eine historisch-semantische Strukturierung des Wortschatzes. Berücksichtigt man weiterhin die unterschiedlichen Gebrauchsaspekte von Wortschatzelementen, dann ergeben sich u. a. regionale oder gruppenbezogene Formen der Wortschatzstrukturierung, aber auch Wortschatzprofile, die sich auf kommunikative Aufgaben wie die Behandlung spezifischer Themen, die Realisierung funktionaler Textbausteine oder die Befolgung von Kommunikationsmaximen wie Präzision oder Originalität beziehen.

<sup>1</sup> Vgl. z.B. den Band VII des ‚Deutschen Fremdwörterbuchs‘ und die Register zu älteren Auflagen des ‚Kluge‘ (z.B. 1899, 1934), sodann aber auch die Ergebnisse in Goebel/Lemberg/Reichmann 1995.

Zwischen den einzelnen lexikologischen Organisationsprinzipien bestehen auch vielfältige Zusammenhänge und Kombinationsmöglichkeiten. Zum Beispiel: um Kommunikationsmaximen wie Präzision, Komprimierung oder Originalität zu befolgen, gebrauchen Sprecher oder Schreiber sehr oft bestimmte Wortbildungsmuster, etwa *-ung*-Bildungen für Formen der komprimierten Rede. Oder: die Gesichtspunkte der Herkunft und der thematischen Zugehörigkeit lassen sich fruchtbar kombinieren, weil viele Fremdwörter einem bestimmten Themenbereich, einem bestimmten Sach- oder Fachgebiet zugehören. Auch viele funktionale Teilwortschätze lassen sich weiter untergliedern, indem man den funktionalen Gesichtspunkt mit anderen Gesichtspunkten kombiniert (z.B. Personenbezeichnungen, die fremdsprachiger Herkunft sind; Ereignisbezeichnungen aus dem Themenbereich des Militärs; Querverweisausdrücke, die vor 1500 erstmals belegt sind usw.).

Vor dem Hintergrund der Tatsache, daß es zahlreiche Organisationsprinzipien für Wortschätze gibt, die teilweise auch untereinander zusammenhängen, erscheint der alte Gegensatz zwischen alphabetisch und onomasiologisch organisiertem Wörterbuch als zu einfach. Zu jedem lexikologischen Organisationsprinzip und zu jeder fruchtbaren Kombination von Organisationsprinzipien ist eine darauf bezogene Dokumentation in Form eines Wörterbuchs möglich. Daß viele dieser denkbaren Wörterbücher nie gedruckt wurden, hängt vor allem mit Geboten der Sparsamkeit zusammen. Es ist nun eine besondere lexikologische Herausforderung, die elektronische Datenbasis für ein Wörterbuch mit Markierungen derart anzureichern, daß die Auswertung der entsprechenden Markierungen bei der Benutzung unterschiedliche ‚Ansichten‘ und Profile des dokumentierten Wortschatzes erlaubt. Die alphabetische Zugriffsmöglichkeit bleibt als eine für das rasche Nachschlagen wichtige lexikographische Errungenschaft<sup>2</sup> natürlich erhalten, sie wird aber ergänzt durch zahlreiche weitere Zugriffsweisen, die auf den internen lexikologischen Markierungen beruhen. Die lexikalische Datenbasis bietet den Benutzern somit die Möglichkeit, unterschiedliche Ausschnitte und verschiedene Ansichten des Wortschatzes auszuwählen. Ist der Datenbestand eines Wörterbuchs als lexikologische Datenbasis vorhanden, dann kann ein Benutzer z.B. durch Ein- und Ausblenden der Belege zwischen zwei Wörterbuchtypen hin- und herschalten, die etwa dem kleinen und dem großen LEXER entsprechen. Man kann den Datenbestand nach Bedarf in einer onomasiologisch aufbereiteten Version nutzen, man kann aus dem Gesamtbestand interessante Teilwörterbücher abziehen, zum Beispiel zu einzelnen Wortarten, zu einzelnen Autoren, zu Textsorten, zu Regionalismen, zu den Fremdwörtern, zu einzelnen zeitlichen Schichten, aber auch zu Kombinationen wie z.B. den Regionalismen, die bei einzelnen Autoren belegt sind.

All diese Ansichten und Auswahlen produziert ein Computer nicht ohne weiteres, sondern aufgrund der Markierungen, die kompetente Lexikographinnen und Lexikographen im Datenbestand vorher verankert haben. Zu den Stärken der beteiligten Menschen gehören im besten Fall Intelligenz, Sprachbeherrschung, vielschichtige Bildung, Konsequenz, Sitzfleisch und philologischer Feinsinn, die Stärke des Computers ist es, in kurzer Zeit riesige Datenbestände im Hinblick auf die markierten Zusammenhänge zwischen den Elementen zu durchforsten und die Ergebnisse in geordneter Form auszugeben, so daß die

<sup>2</sup> Vgl. zur strengen Alphabetisierung als einer lexikographischen Errungenschaft u. a. Powitz 1988, 211 f. mit Fußnoten 3 und 11. Hier heißt es über Johannes de Janua (Johannes Balbus) und sein ‚Catholicon‘ (abgeschlossen 1286; erstmals gedruckt 1460): „Zum Erfolg des ‚Catholicon‘ trug ein weiterer Vorzug bei: die neuartige Organisation des lexikalischen Stoffes. Die Entdeckung und Anwendung des streng alphabetischen Prinzips ist der Stolz des Autors und sein tatsächliches historisches Verdienst“.

Menschen wieder intelligent mit den Ergebnissen weiterarbeiten können. So oder so ähnlich kann man die Verteilung der Arbeit Mensch/Computer vielleicht sehen.

Wir wollen die Idee einer mehrdimensionalen Strukturierung und Erfassung von Wortschätzen in elektronischen Wörterbüchern nun anhand von Goethes Wortschatz und anhand des gedruckten Goethe-Wörterbuchs erläutern und dabei auf einige Nutzungsperspektiven und zumindest kurz auf einige Gesichtspunkte der konzeptionellen und technischen Realisierung eingehen.

### 3 Die Architektur des Goethe-Wortschatzes

Goethes Wortschatz umfaßt über 90000 Stichwörter, von denen viele ein fein differenziertes Spektrum von Verwendungsweisen haben. Dieser Personal-Wortschatz ist in vielfacher Hinsicht aufschlußreich, und für jede der im folgenden genannten ‚Hinsichten‘ der Wortschatzstrukturierung ist es wünschenswert, Zugriff auf eine Dokumentation des entsprechenden Teilwortschatzes zu haben.

Ein bekanntes Strukturierungsprinzip sind die semantischen Zusammenhänge im Wortschatz. Der Zusammenhang zwischen Quasi-Synonymen und bedeutungsverwandten Wörtern bzw. Verwendungsweisen wird im Goethe-Wörterbuch in der Artikelposition **Syn** verbucht. Zu den semantischen Zusammenhängen im Wortschatz gehören aber auch die im GWb nicht in einer eigenen Rubrik aufgeführten Gegensatzrelationen, die in Goethes Darstellungen vor allem bei Formen der Kontrastierung eine wichtige Rolle spielen. So wird z.B. im Feld der Wahrnehmung ein Kontrast aufgebaut zwischen *scheinbar* und *eigentlich* bzw. *wahr*.

Goethes Wortschatz umfaßt zahlreiche wichtige Fachgebiete und thematisch-sachliche Bereiche der Zeit um 1800. Zu diesen Themenbereichen und Sachgebieten gehören – hier jeweils mit einigen Unterrubriken und Beispielen – unter anderem:

- Anatomie und Morphologie, z.B. mit Bezeichnungen für Körperteile und Körperbestandteile (*Gelenkapophyse*, *Kinnlade*, *Nasenknochen*, *Zwischenknochen*; *Retina*, im Schnittbereich zwischen Farbenlehre und Anatomie, *Hinterhaupt* im Schnittbereich zur Physiognomik) oder vielfältige Bezeichnungen für Aspekte der Gestalt (*Bildung*, *Fortsatz*, *Hervorragung*, *Sutur*, *Wölbung*);
- Dichtungs- und Literaturtheorie, z.B. mit Verben zur Bezeichnung von möglichen Wirkungen der Dichtung (*bewegen*, *rühren*), Bezeichnungen für literarische Gattungen (*Epos*, *Novelle*) oder literarische Darstellungsformen und Prinzipien (*Abstraktion*, *Anschauung*, *Anmut*, *Bild*, *Genauigkeit*, *Nutzen*);
- Farbenlehre, z.B. mit Bezeichnungen für Farben und Farbeindrücke, deren zugrundeliegende Wortbildungsmuster in besonders eindrucksvoller Weise auch die Formulierunsnöte in diesem Bereich erkennen lassen (*blaßgelb*, *bläulich*, *blutrot*, *dunkelorange*, *gelb*, *gelblich*, *gelbrot/rotgelb*, *graulich*, *hellgelblich*, *hochrot*, *rosenfarb*, *rötlich*, *weiß*), Bezeichnungen für die Manipulation von Farben und Farbsubstanzen (*lasieren*, *mischen*, *diluiieren*), für Geräte (*Chromatoskop*) und allerhand optische Erscheinungen und Verhältnisse (*Gegenlicht*);
- Geologie, Bergbau und Mineralogie, z.B. mit Bezeichnungen für unterschiedliche Gesteinsarten (*Alabaster*, *Bufonit*, *Feldspat*, *Frauenglas*, *Glimmer*, *Gneis*, *Gneisgranit*, *Granit*, *Granitgneis*, *Quarz*, *Sandstein*, *Schörl*, *Tuffstein*), Bezeichnungen für Personen

(*Bergbeflissene, Bergverständige*), Bezeichnungen für geologische Formationen (*Altar, Bank, Flözkluft, Gang, Gebürge, Hügel, Kluft, Steinbruch, Versteinerung*), Bezeichnungen für Eigenschaften von Gesteinen (*abgerundet, breccienartig, porphyrtartig, sechseckig, tafelförmig, quarzhaltig*), auch dieser Wortschatzsektor deutet teilweise auf Formulierungsnotwendigkeiten im Umgang mit der Vielfalt der Phänomene hin;

- Pflanzenkunde, z.B. mit Bezeichnungen für Arten von Pflanzen, für Teile von Pflanzen, für Eigenschaften und Formen der Entwicklung;
- Theater und Theatermanagement, z.B. mit Bezeichnungen wie *Gruppierung* oder *Entreegeld*;
- verschiedene Bereiche der verwaltungssprachlichen Praxis, z.B. mit Ausdrücken, die vor allem als Mittel der Höflichkeit und der Kennzeichnung von amtlichen Rollenkonstellationen dienen (als *ohnzielsetzend* oder *ohnmaßgeblich* sind in der Regel die Voten der Räte gekennzeichnet, um dem Fürsten die Entscheidungsfreiheit zu belassen).

Die strukturierte Beschreibung all dieser thematischen Wortschatze mitsamt ihren teilweise fachlichen Bestandteilen, mit ihren Neubildungen und ihrer differenzierten inneren Struktur ist ein wichtiger Beitrag zum Verständnis, wie das Werkzeug ‚Sprache‘ im Hinblick auf die Bewältigung thematischer Redeaufgaben funktioniert.

Der Goethe-Wortschatz enthält weiterhin sprachliche Mittel, die wesentliche Denkweisen und Auffassungen Goethes und seiner Zeitgenossen widerspiegeln, also vor allem diejenigen Wörter, die der Begründer des Goethe-Wörterbuchs, Wolfgang Schädewaldt, als Goethesche Grund- und Wesenswörter ansah. Kandidaten für diese Rubrik sind z.B. *dämonisch, Dasein, Erscheinung, ganz, Gegenwart, Geheimnis, Genie, Geselligkeit, handeln, heiter, Mitleben, Teilnahme, Zusammenwirken* und dergleichen. Das Studium des Gebrauchs dieser Wörter führt oft schnell und zielgerecht zu wesentlichen Grundauffassungen Goethes oder seiner Zeitgenossen, zu Wandlungen und Änderungen dieser Auffassungen (z.B. bei *Genie*) oder auch zu kultur- oder wissenschaftsgeschichtlich interessanten Entwicklungen (z.B. bei *Elektrizität* und seinem Wortfeld, das sich auf ein neues Phänomen und einen neuen Wissenschaftszweig bezieht, der Mitte des 18. Jahrhunderts erst neu beschrieben bzw. entwickelt wurde, oder bei den *Dampf*-Komposita).

Sodann enthält Goethes Wortschatz eine Reihe von Euphemismen, Neubildungen und Anwendungen von Wortbildungsmustern. Ein von Goethe verwendeter Euphemismus ist z.B. *den Vaterlandsboden verlassen* für ‚sterben‘. Eine Neubildung ist offenbar das Wort *feminisieren* im Sinne von ‚sich bei der sprachlichen Darstellung einem weiblichen Publikum anpassen‘, das Goethe in Anlehnung an *popularisieren* verwendet hat. Die Wortbildungen wurden oben im Zusammenhang mit den Formulierungsnotwendigkeiten schon erwähnt, in die man bei der Beschreibung naturkundlicher Phänomene geraten kann. Viele allgemeinsprachliche Wörterbücher würden sich für solche ‚ephemere‘ und nur kurzfristige Erscheinungen im Wortschatz gar nicht interessieren. Die Neubildungen sind aber zum einen wichtig, um das kreative Potential des Deutschen um 1800 beurteilen zu können. Wenn man weiterhin der Auffassung ist, daß die Anwendung solcher innovativer Verfahren der Ausgangspunkt ist für historisch-semantiche Neuerungen, die sich später als Bedeutungswandel etablieren, dann erscheint es als eine wichtige Aufgabe, den vorhandenen Pool dieser Neuerungen zu erfassen, zu beschreiben und zu kennzeichnen, um auf diese Weise beizutragen zu einem besseren Verständnis der historisch-semantiche Prozesse der Innovation, der Selektion und der Verbreitung von sprachlichen Neuerungen (vgl. hierzu Fritz 1998).

In einer ganz anderen Perspektive auf den Wortschatz Goethes und seiner Zeit kann man die Frage stellen, welche der damals geläufigen Verwendungsweisen den heutigen Lesern nicht mehr geläufig sind, deren Beherrschung aber auch für das Verständnis vieler anderer Texte des 18. und 19. Jahrhunderts zentral ist. Fritz Tschirch hat in der ‚Zeitschrift für deutsche Wortforschung‘ (1960) einen sehr erhellenden Rundgang zu einigen typischen Fußangeln und Fallgruben bei der Lektüre von Texten um 1800 veröffentlicht, der Pflichtlektüre vor jeder Beschäftigung mit älteren Texten der Goethe-Zeit sein sollte. So sind etwa Verwendungsweisen von *gemein*, *geradezu* oder *umständlich* ‚ausführlich‘ heute nicht mehr gebräuchlich oder zumindest nicht mehr ohne weiteres in ihrer richtigen Lesart verständlich. Eine Kennzeichnung derjenigen Verwendungsweisen aus der Zeit um 1800, die heute nicht mehr ohne weiteres verständlich sind, würde es erlauben, aus dem Datenbestand ein ‚Idiotikon der Goethezeit‘ herauszuziehen, ein kontrastives Wörterbuch der Art, wie es Huguet zum französischen Sprachgebrauch des 17. Jahrhunderts angelegt hat, insofern dieser Sprachgebrauch vom modernen Sprachgebrauch abweicht. Ein solches Wörterbuch wäre wertvoll, weil es die wesentlichen sprachlichen Gefährdungen und ‚falschen Freunde‘ enthielte, die bei der Lektüre der Klassiker auf den Leser lauern.

Eine weitere Dimension der Wortschatzstrukturierung ist die Gliederung in funktionaler Hinsicht, also im Hinblick auf die Frage, welchen Beitrag einzelne lexikalische Einheiten leisten, um bestimmte kommunikative Aufgaben zu realisieren. Goethes Wortschatz ist reichhaltig in funktionaler Hinsicht: Er enthält ein breites Repertoire von lexikalischen Mitteln für einzelne kommunikative Aufgaben wie z.B. Zeitangabe, Ortsangabe, Bezugnahme auf Personen und Gegenstände, Kennzeichnung von Zusammenhängen zwischen Äußerungen, Querverweise auf frühere Textstellen, Charakterisierung von Eigenschaften beim Beschreiben usw. Zum Wortschatz der Charakterisierung von Eigenschaften gehören z.B. die schon genannten Farbbezeichnungen wie *gelbrot* oder Konsistenzbezeichnungen bzw. Konsistenzverwendungsweisen wie *fest*, *weich*, *quammig*, *quappig* oder *determinabel*. Zu den lexikalischen Mitteln, um Querverweise auf frühere oder spätere Textstellen zu geben, gehören z.B.: *angeregt*, *erwähnt*, *ersagt*, *gedacht* oder *vorhergehend* und vielfältige Kombinationen wie z.B. *ob gedacht*, *oft ersagt*, *vor angeführt*, *vor angeregt*, *eingangs erwähnt* oder *mehr erwähnt*.

Die zuletzt genannte Gruppe der Querverweise stellt einen Wortschatzsektor mit Wörtern und Verwendungsweisen dar, die besonders aufschlußreich sind für bestimmte Texttraditionen, hier z.B. der Verwaltungssprache. Querverweisausdrücke wie *ob gedacht* oder *mehr ermelt* waren zuerst im Deutsch der mittelalterlichen Urkunden des 13. und 14. Jahrhunderts üblich, sie wurden dann im 15., 16. und 17. Jahrhundert in zahlreiche andere Texttypen übernommen, ihr Gebrauch nimmt dann vor allem seit dem 18. Jahrhundert – mit dem Prinzip der natürlichen Schreibart – stetig ab. Das letzte und konservative Rückzugsgebiet für den variationsreichen Gebrauch dieser Ausdrücke um 1800 ist die Verwaltungssprache. Es ist ein bemerkenswerter textsortengeschichtlicher Befund, wenn Goethe diese Art von Ausdrücken auch in den naturwissenschaftlichen Schriften nicht selten und in bemerkenswerter Variation verwendet.

Quer zu all diesen lexikologischen Gesichtspunkten liegt der Gesichtspunkt der Herkunft eines Wortes bzw. einer Verwendungsweise. Die Fremdwörter, die Goethe gebraucht, sind zum einen interessant im Hinblick auf das Sprachstadium und seinen etablierten Fremdwortschatz, zum anderen aber auch im Hinblick auf die sich wandelnde Stellung Goethes zu den Fremdwörtern (z.B. *frugal*, *epitomieren*, *ebauchieren*).

Manche der hier in erster Näherung skizzierten Organisationsprinzipien des Goethe-Wortschatzes hängen mit einzelnen oder gar mehreren anderen Organisationsprinzipien zu-

sammen. Die Fremdwörter lassen sich z.B. vielfach bestimmten Sachbereichen zuordnen. Der Wortbestand einzelner Sachgebiete läßt sich weiter untergliedern im Hinblick auf verschiedene funktionale Leistungen. Innovative Bildungen lassen sich vielfach im Hinblick auf typische Wortbildungsmuster charakterisieren usw.

Eine Konzeption der Architektur eines Wortschatzes, seiner Organisationsprinzipien, seiner internen Untergliederung und der Zusammenhänge zwischen Wortschatzsektoren ist die Grundlage für alle elektronische Auszeichnung und Markierung. Alle Aspekte der Gliederung des Wortschatzes müssen aus den zu erfassenden Texten selbst erarbeitet werden. Es verbietet sich dabei, apriorische Systeme der Wortschatzstrukturierung wie die von Dornseiff oder Hallig/v. Wartburg zu verwenden. Dafür gibt es vor allem, aber nicht nur, drei Gründe. Erstens lassen diese Systeme keinen Raum für die Offenheit und die Zusammenhänge zwischen Strukturierungsgesichtspunkten: ein Ausdruck steht an einem Ort des Systems, obwohl er im Hinblick auf einen weiteren Strukturierungsgesichtspunkt auch an einem weiteren Ort des Systems stehen könnte oder sogar stehen müßte. Zweitens sehen diese Systeme in der Regel nicht vor, daß ein Ausdruck unterschiedliche Verwendungsweisen haben kann und mit jeder Verwendungsweise in eine andere Rubrik der Wortschatzstrukturierung gehören kann. Jede Form der Wortschatzstrukturierung, die Polysemie nicht oder ungenügend berücksichtigt, ist wirklichkeitsfremd, nicht nur im Hinblick auf Goethe. Und drittens ist in den Standardversionen dieser Systeme kaum Platz für die zeit- und kulturtypischen Besonderheiten eines Wortschatzes, die sich in der Regel erst aus der genauen Analyse des Aufbaus einzelner Wortschätze und ihrer Sektoren ergeben. Nimmt man die z.B. bei Hallig und v. Wartburg zwar formulierte, aber in der Praxis wenig befolgte Forderung ernst, das System den Erfordernissen des jeweiligen Wortschatzes anzupassen, dann endet man schnell bei der Aufgabe, ein neues System der Wortschatzarchitektur zu entwerfen. Und das ist auch kein bedauerlicher Schaden, sondern die eigentliche Aufgabe der Lexikologen.

Wir kommen nun zur Frage, welchen Nutzen und welche erweiterten Benutzungsperspektiven es bieten würde, die hier skizzierten Gesichtspunkte der Wortschatzstrukturierung in einem elektronischen Goethe-Wörterbuch als einer lexikologischen Datenbasis mit multiplem Zugriff zu kennzeichnen und explizit zu markieren.

#### 4 Ein elektronisches Goethe-Wörterbuch: Einrichtung und Nutzungsperspektiven von lexikologischen Datenbasen

Die Hauptgesichtspunkte, die für ein elektronisches Goethe-Wörterbuch in Form einer lexikologischen Datenbasis mit mehrfachem Zugriff sprechen, ergeben sich zum einen aus der skizzierten Natur der Wortschatzarchitektur, zum anderen aber auch aus den möglichen Benutzerinteressen und möglichen weiterführenden Fragestellungen von Benutzern. Ein elektronisches Wörterbuch als lexikologische Datenbasis sollte zum einen der Komplexität des Gegenstandes gerecht werden, zum anderen für möglichst viele Benutzerinteressen und möglichst viele Arten von Fragestellungen vorbereitet sein. Als eine passende Umgebung bietet sich derzeit eine SGML-kodierte Datenbasis an. Bevor wir auf erweiterte Nutzerperspektiven eingehen, möchten wir einige Grundgedanken dieser Erfassungsumgebung darlegen.

#### 4.1 Mehrfachnutzung von SGML-Daten: gedruckte und elektronische Produkte

Ein Grundgedanke der Textauszeichnung mit Hilfe von Markup-Systemen wie SGML (Standard Generalized Markup Language) ist es zum einen, den gezielten Zugriff auf einzelne Textelemente und darauf beruhende Auswertungen zu ermöglichen, zum anderen können aus ein und derselben Datenbasis unterschiedliche Endprodukte für unterschiedliche Nutzerinteressen hergestellt werden. Dabei sind zunächst zwei Hauptgruppen von Produkten zu unterscheiden: gedruckte Produkte und elektronische Produkte. In beiden Hauptgruppen lassen sich weiterhin Produkte mit unterschiedlicher Nutzungsausrichtung und Dokumentationsiefe erstellen.

Zunächst zu den gedruckten Fassungen. Ein sehr einfacher Artikel sieht in der gedruckten Fassung des Goethe-Wörterbuchs folgendermassen aus:

**entschmeicheln** durch Schmeicheleien entlocken [Chor:] Wer die Schönste für sich begehrt,|.Schmeichelnd wohl gewann er sich|Was auf Erden das Höchste;|Aber ruhig besitzt er's nicht;|Schleicher listig e. sie ihm Faust II 9488 **Syn** entlocken

Die einzelnen Textelemente wie Lemma, (semantische) Leitbemerkung, Belegtext, Stellenangabe oder Synonym weisen jeweils eigene typographische Kennzeichen auf. In einer elektronischen SGML-Datenbasis werden den einzelnen Textelementen strukturelle Kennungen zugewiesen, mit deren Hilfe man (i) auf die unterschiedlichen Textteile zugreifen und (ii) ihre Eigenschaften manipulieren kann. Ein einfaches Beispiel ist folgender Artikel:

```
<art>
<le>entschmeicheln</le>
<sem>durch Schmeicheleien entlocken</sem>
<bel><komm>[Chor:]</komm><belt>Wer die Schönste für sich
begehrt,|&ausl;
Schmeichelnd wohl gewann er sich|Was auf Erden das Höchste;|Aber
ruhig besitzt er's nicht:|Schleicher listig e. sie
ihm</belt><belst>Faust II 9488</belst></bel>
<syn>entlocken</syn>
</art>
```

Die strukturellen Kennungen in den Spitzklammern haben hierbei folgende ‚Bedeutung‘:

<art></art>	Anfang und Ende eines Wortartikels;
<le></le>	Anfang und Ende des Lemmas;
<bel></bel>	Anfang und Ende eines Belegteils;
<komm></komm>	Anfang und Ende eines Kommentars im Beleg;
<belt></belt>	Anfang und Ende des Belegtexts;
<belst></belst>	Anfang und Ende einer Beleg-Stellenangabe;
<syn></syn>	Anfang und Ende der Angabe von Synonymen.

Diese Art der strukturellen Kennzeichnung von Textelementen ist noch unabhängig von typographischen Entscheidungen im Hinblick auf ein Endprodukt. Will man einen solchen Text drucken, dann sind die strukturellen Kennungen per Programm in typographische Steuercodes zu überführen. Die Herstellung von Satzdaten aus einer SGML-Datenbasis ist eine Aufgabe, die z.B. mit Programmen wie FrameMaker oder Tustep, aber auch in den Satzabteilungen graphischer Betriebe gelöst werden kann. Eine solche Umsetzung beinhaltet etwa, daß die Kennung für ein Lemma (<le>freilich</le>) per Programm umgesetzt



wird in typographische Befehle für den Anfang bzw. das Ende von Fettdruck oder daß eine eindeutige Kennung für Auslassungen im Belegtext (&ausl;) per Programm entsprechend den redaktionellen Vorgaben ausgetauscht wird (zwei Punkte, drei Punkte, drei Punkte mit runder Klammer, drei Punkte mit eckiger Klammer usw.).

Man kann derzeit leider nicht prognostizieren, daß solche Umsetzungen ganz ohne Anstrengungen und Reibungsverluste vonstatten gehen, aber die Tatsache, daß SGML sich im Bereich der Technischen Dokumentation durchsetzt, wo es oft um mehrere Tausend Seiten und viele Hunderttausend Mark oder Dollar geht, deutet darauf hin, daß die Anwendung dieses Standards sich langfristig ‚lohnt‘ und daß dieser Standard auch im graphischen Gewerbe zunehmende Verbreitung findet. Im Bereich der Geisteswissenschaften sind es vor allem die weitreichenden Koordinationsvorschläge und die Vorgaben der TEI (Text Encoding Initiative), die auf SGML zurückgreifen.

Für den wissenschaftlichen Einsatz ist die Nutzung einer lexikalischen SGML-Datenbasis, z.B. auf CD-ROM, besonders interessant. Vor allem die Erfahrungen mit dem Oxford English Dictionary auf CD-ROM zeigen, daß die Auswertungs- und Nutzungsmöglichkeiten der elektronischen Version einen qualitativen Sprung darstellen. Zwar werden die Bibliotheken und die Liebhaber des gedruckten Buches (zu denen wir uns selbst mit Entschiedenheit zählen) kaum auf die gedruckte Fassung verzichten wollen, aber für die Zwecke der Auswertung und der Bearbeitung wissenschaftlicher Fragestellungen ist die SGML-Fassung auf CD ein unschätzbares Hilfsmittel.

Aus einer SGML-Datenbasis lassen sich – ebenfalls per Programm, z.B. mit Tustep oder durch Verwendung geeigneter ‚Filter‘ – auch eine oder mehrere Internet-Fassungen des Datenbestandes erstellen. Die Möglichkeiten des differenzierten Zugriffs und der komplexen Abfrage sind derzeit bei Internet-Materialien zwar nicht besonders gut ausgebaut, dennoch sind vereinfachte Internetversionen ein gutes Mittel, um Material einem weiteren Benutzerkreis kostenfrei oder kostengünstig zugänglich zu machen und gleichzeitig Informationen über höherwertige ‚Vollversionen‘, die z.B. auf CD-ROM vertrieben werden, zu verbreiten.

## 4.2 Erweiterte Benutzungsperspektiven eines elektronischen Goethe-Wörterbuchs

Eine elektronische SGML-Datenbasis mit den Daten des Goethe-Wörterbuchs bietet zum einen differenzierte Nutzungsmöglichkeiten für Benutzer(innen) mit unterschiedlichen Interessen, zum anderen erlaubt sie auch den Einbau von Informationen, die eine Grundlage sind für erweiterte Forschungsmöglichkeiten. Differenzierte Nutzungsmöglichkeiten sind zunächst ‚nur‘ ein Aspekt der Bequemlichkeit und der Zugriffsökonomie, wogegen erweiterte Forschungsmöglichkeiten – z.B. durch Erschließung der Wortschatzarchitektur – einen echten Mehrwert darstellen.

### 4.2.1 Dokumentationstiefen und Informationstypen

Neben den erweiterten Auswertungs- und Abfragemöglichkeiten bietet eine SGML-Datenbasis den Vorteil, daß sich die Auswahl der Informationstypen und die Dokumentations-tiefe von Benutzerseite variieren lassen:

– ein(e) Benutzer(in) kann wählen zwischen dem elektronischen Gegenstück der gedruckten Fassung mit Belegzitaten, einer Fassung ohne Belegzitate, aber mit vollständigem Set lexikographischer Definitionen, im GWb ‚Leitbemerkenungen‘ genannt, und z.B. einer

Version, die nur das wesentliche semantische Gliederungsgerüst wiedergibt, indem man nicht nur auf alle Belege, sondern auch auf alle nicht mit einer eigenen Gliederungsmarke versehenen Leitbemerkungen und Unterleitbemerkungen verzichtet, oder, noch radikaler reduziert, indem man sich nur die Grunddifferenzierungen der höchsten Gliederungsebene(n) darstellen läßt;

- ein(e) Benutzer(in) kann bestimmte Typen von Einträgen einblenden oder ausblenden, z.B. Verweise auf das Grimmsche Wörterbuch, auf Synonyme, auf andere lexikalische Referenzinstanzen oder auf bestimmte Gewährsleute für Goethes Wortgebrauch (z.B. Herder, Schiller, Humboldt);

- die SGML-Version bietet die Möglichkeit, Eintragstypen zu ergänzen, die bislang im GWb nicht vorgesehen waren bzw. nicht konsequent gehandhabt wurden, z.B. Sachgebietskennzeichnungen wie ‚numismatisch‘, thematische Kategorisierungen wie ‚kosmogonisch‘, Hinweise zur Herkunft wie ‚Fremdwort‘ oder pragmatische Kennzeichnungen wie ‚Schimpfwort‘. Ein feindifferenziertes Netz solcher Kennzeichnungen ist ein bedeutendes Hilfsmittel, um die spezifische Architektur des Goethe-Wortschatzes zu überschauen.

#### 4.2.2 Vernetzung des elektronischen Goethe-Wörterbuchs mit anderen Dokumenten

Das GWb verweist auf andere einschlägige Informationsmittel bisher nur sporadisch, in problematischen Einzelfällen, etwa wenn für eine prima facie eigenwillige Lemmaansetzung oder sonderbar anmutende Bedeutungserklärung auf entsprechende Vorgänger in lexikalischen Standardwerken hingewiesen werden soll, aber auch wenn es gilt, sich von nur vordergründig plausiblen Erläuterungen, etwa in gängigen Kommentaren oder neuesten Editionen, ausdrücklich abzusetzen. Ein systematischer lexikalischer Abgleich mit anderen Wörterbüchern, die den Zeitraum um 1800 abdecken, ist zwar als Vorarbeit der Artikelautoren obligatorisch, nicht aber als Rubrik im gedruckten Artikel. Gleichwohl wäre es immerhin denkbar, in einem elektronischen Goethe-Wörterbuch Verweise auf solche Informationsmedien vorzusehen, z.B. auf die vorhandenen Kommentare, auf die allgemeinsprachlichen Wörterbücher, die den Zeitraum abdecken (Adelung, Campe, DWb), auf entsprechende fachsprachliche Wörterbücher (zur Anatomie, Bergbaukunde, Botanik, Chemie, Kameralistik, Mineralogie, Ökonomie und Önologie, bis hin zur Zoologie), schließlich auch auf den Sprachgebrauch von Zeitgenossen, von Klopstock über Ignaz von Born bis Schopenhauer. Vor einem solchen Hintergrund des Sprachüblichen wie auch des mehr oder minder bekannten Besonderen und Innovativen ließe sich deutlicher erkennen, was allgemeiner Sprachgebrauch der Zeit und was spezielle Sprachschöpfung, individuelle Eigenart Goethes – oder Herders oder Schillers – ist. Selbstverständlich ist eine solche Vernetzung kein Projekt, das man, vom GWb ausgehend, systematisch quer durch den gesamten Wörterbuchbestand vorantreiben könnte oder sollte. Aber es ließe sich in einer elektronischen Datenbasis immerhin die Möglichkeit vorsehen, Arbeitsergebnisse, die von den Bearbeiterinnen und Bearbeitern sowieso ermittelt werden, auch zu dokumentieren. In der gedruckten Fassung können diese Informationen problemlos unterdrückt werden, um beim eingeführten Aufbau und Erscheinungsbild zu bleiben, ebenso in den Standard-Ansichten der elektronischen Version. Aber: für diejenigen Benutzer, die Angaben dieser Art verwenden können und wollen, wären sie ebenso leicht zuzuschalten.

#### 4.2.3 Aktualisierbarkeit und Revidierbarkeit

Eine SGML-Datenbasis bietet gegenüber der gedruckten Fassung eine weitere wichtige Eigenschaft: Sie ist in allen Wörterbuchstrecken laufend aktualisierbar und den Benutzern leichter und schneller zugänglich zu machen. Auch hierzu ein Beispiel: Sollte die Redaktionskonferenz des GWb etwa aufgrund der Hinweise eines Rezensenten zu der Auffassung gelangen, daß unter den synonymischen Verweisen zur Verwendungsweise 2b des Verbs *bewegen* auch das Verb *rühren* erscheinen sollte, dann könnte man einen solchen Eintrag in einer Datenbasis leicht ergänzen. Ein solcher Eintrag käme den Benutzern von elektronisch aktualisierten Versionen schneller zugute als den Beziehern gedruckter Nachtragsbände.

Die Frage der beschleunigten Aktualisierung ist vor allem im Bereich der synonymischen Vernetzung interessant, die besonders schwer zu überblicken ist und bei der sich oft erst im Lauf der weiteren Bearbeitung von Wörterbuchstrecken neue Einsichten ergeben. Denkbar wäre weiterhin, daß – zumindest in den Grenzen der verfügbaren Arbeitskapazität – auch ältere Wörterbuchstrecken an neue redaktionelle Richtlinien angepaßt und revidiert werden könnten. Die zum Teil fehlenden oder wenig expliziten Bedeutungsbeschreibungen des ersten Bandes zum Beispiel könnten auf diese Weise im Lauf der Zeit und nach Maßgabe der vorhandenen Arbeitskapazität den aktuellen Richtlinien angepaßt werden. Allerdings: angesichts knapper Mittel und reduzierter Mitarbeiterzahlen in den Redaktionen des GWb sind dies derzeit bloße Wunschträume.

#### 4.2.4 Ein elektronisches Goethe-Wörterbuch als kostenlose Internet-Version

Zu den elektronischen Produkten, die aus einer SGML-Datenbasis per Programm erstellbar sind, gehören auch Internet-Versionen von Wörterbüchern. Wir plädieren hier dafür, solche Versionen im Internet kostenlos anzubieten, vor allem wenn ein Wörterbuch-Unternehmen mit beträchtlichen öffentlichen Mitteln gefördert wird. Eine Internet-Version des Goethe-Wörterbuchs ist zunächst ein Zugeständnis an diejenigen Steuerzahler, die keinen Zugang haben zu Bibliotheksexemplaren des GWb oder zu GWb-Exemplaren in Privatbesitz. Um nur zwei Beispiele zu nennen: Journalisten oder Politiker, die dauernd oder zeitweise keinen Zugriff auf eine größere öffentliche Bibliothek haben, die aber auf lexikalischem Weg nach einem passenden Goethe-Zitat suchen, könnten über das Internet in einer entsprechend aufbereiteten, vereinfachten Datenbasis recherchieren. Oder: Schüler, die keine Schulbibliothek mit Goethe-Wörterbuch haben – so etwas soll es tatsächlich noch geben! –, könnten in einem entsprechenden Kurs lernen, (1) wie man das Internet für Zwecke der Recherche benutzt, (2) wie und wofür man das Goethe-Wörterbuch nutzen kann (z.B. die bei der Lektüre älterer Texte durch Bedeutungswandel allenthalben lauernden Mißverständnisse vermeiden lernen; Zitate finden). – Wir nennen den lexikalischen Zugriff auf „passende Zitate“ hier deshalb, weil es eine interessante und zunächst auch erschütternde Erfahrung für den englischen Computerphilologen Roy Wisbey war, daß seine monumentale Shakespeare-Konkordanz vor allem für die Suche nach Shakespeare-Zitaten verwendet wurde. Wir meinen aber, man muß realistischerweise auch mit diesem Benutzerinteresse rechnen.

Die angedeuteten unentgeltlichen Zugriffsmöglichkeiten im Internet kollidieren prima facie mit dem Verlagscopyright und den damit verbundenen finanziellen Interessen. Auf der anderen Seite, bei genauerer Überlegung, kann ein „GWb light“ im Internet als attraktive Werbung für die Druckfassung oder für eine elektronische Vollversion mit erweiterten Nutzungsmöglichkeiten auf CD-ROM fungieren.

Im Vordergrund sollten allerdings die kulturpolitischen Aspekte einer Internet-Version und die Verpflichtung gegenüber ‚dem Steuerzahler‘ stehen. Das Unternehmen ‚Goethe-Wörterbuch‘ hat die öffentliche Hand in den vergangenen Jahrzehnten viele Millionen Mark gekostet. Es ist einer der Widersprüche unserer Zeit, daß viele Leistungen, die mit öffentlichen Geldern erarbeitet wurden, heute ausschließlich den Verlagen ‚gehören‘ und daß über wichtige kultur- und forschungspolitische Fragen letztlich in den Marketingabteilungen der Verlage entschieden wird. Es ist an der Zeit, daß Verlage und die Repräsentanten der öffentlichen Hand (z.B. die Akademien) neue Formen vereinbaren, wie solche Leistungen einem erweiterten Kreis von Nutzern zugänglich gemacht werden können.

#### 4.2.5 Architektur und onomasiologische Erschließung des Goethe-Wortschatzes

Der eigentliche Mehrwert einer lexikalischen Datenbasis besteht in der Möglichkeit, die vielfältigen Aspekte der lexikalischen Organisation eines Wortschatzes zu erfassen und in einen Zusammenhang zu bringen. Alphabetische Wörterbücher, so lautet ein auch nach Hermann Paul immer wieder vorgebrachter und völlig berechtigter Kritikpunkt, tragen nichts zur Kenntnis der Architektur der dokumentierten Wortschatze bei. Daß sicherer Zugriff durch alphabetische Dokumentation und Erschließung der Architektur nicht unverträglich sind, zeigt z.B. das „Wörterbuch zu Goethes West-östlichem Divan“ von Christa Dill: hier ist die alphabetische Abfolge der Wortartikel ergänzt durch ausführliche Darstellungen zum Divan-Wortschatz und zur Divan-Bildwelt sowie durch eine umfassende tabellarische Darbietung der „Wortfelder und Sinngruppen“ in Goethes ‚Divan‘. Wie läßt sich diese Wortschatz-Erschließung von einem einzelnen Werk übertragen auf die Weite der gesamten Goetheschen Sprachwelt, die, wie erwähnt, über 90 000 dokumentierte Wörter mit einem Vielfachen an Verwendungsweisen umfaßt? Die elektronische Dokumentation von Wortschatzen erlaubt es zunächst, mehrere lexikologische Organisationsprinzipien nebeneinander zu verwenden und von vornherein verschiedene Zugriffs- und Abfragemodi vorzusehen. Voraussetzung dafür ist jedoch eine durchgängige Vorstrukturierung des Wortschatzes, z.B. durch eine konsequente Merkmalsmarkierung – „What you mark is what you get!“

Zu den Organisationsprinzipien, die alternativ bzw. komplementär zur alphabetischen Anordnung Berücksichtigung verdienen, gehören:

- Textsortenzuweisungen (z.B. Dichtung, Privatbrief, amtliches Schreiben);
- funktional-pragmatische Kategorien (etwa vom bedeutungsentblößten Funktionsverb oder rein verstärkenden Adjektiv über wertende Verwendungen bis zum derben Schimpfwort);
- semantische Kategorien und Gebrauchsbereiche (z.B. Konsistenzbezeichnungen; Farbwörter im Zusammenhang der Farbenlehre; sprechhandlungsbezeichnende Ausdrücke im Bereich der Rechtssprache);
- Sachgebiete und Themenbereiche (z.B. Bergbau, Botanik, Ästhetik);
- Neologismen und Hereditäten (genuin Goethesche Wortschöpfung oder nachweisbare Übernahme von anderen Autoren, geistigen Bewegungen, z.B. Pietismus, Kantianismus, Freimaurerei, aber auch aus Dialekten);
- zeitgebundene Verwendung (z.B. nur beim jungen Goethe, nur nach 1815);
- sprachliche Mittel, die in besonders engem Zusammenhang mit Goethes Weltansicht stehen (z.B. Verben der sinnlichen Wahrnehmung und kognitiven Anschauung).

Eine konsequente Kennzeichnung des Goethe-Wortschatzes gemäß solchen Gesichtspunkten, mit den entsprechenden Kennungen in der elektronischen Aufbereitung, würde auch das Herausfiltern von Teilwortschätzen und Kreuzklassifikationen unter Fragestellungen wie den folgenden ermöglichen:

- Welche sprachlichen Mittel verwendet Goethe bei der Behandlung geologischer Gegenstände? Woher stammen sie?
- In welchen Texttypen/ In welcher Zeit kommen Querverweisausdrücke wie *obgemeldet* oder *vorgenannt* bevorzugt vor?
- Welche Schimpfwörter verwendet Goethe? Herrschen dabei Dialektausdrücke vor?
- Welche Ausdrücke bzw. Verwendungsweisen aus dem Sachgebiet der Kunst sind aus der ‚Italienischen Reise‘ belegt?

#### 4.3 Probleme und Hindernisse

Auf dem Weg zu einer SGML-Datenbasis des elektronischen Goethe-Wörterbuchs begegnen eine Reihe von offenen Fragen und Problemen. Sie liegen auf ganz unterschiedlichen Ebenen: auf der Ebene der Texterfassung und der Datenkonvertierung, auf der juristisch-wirtschaftlichen Ebene und auf der Ebene der bisher befolgten Prinzipien der Textgestaltung.

Ein erstes Problem ist die Frage der Texterfassung und der Datenkonvertierung. Das Goethe-Wörterbuch, das gerade seinen dritten Band vorlegen konnte, ist Mitte der Achtziger Jahre, also erst während der Erarbeitung des zweiten Bandes, zur elektronischen Datenerfassung übergegangen, und zwar mithilfe gängiger Schreibprogramme (Word, Winword), die vom graphischen Betrieb in Satzdaten konvertiert werden. Will man diese Daten verwerten, dann stellen sich u. a. folgende Fragen: Wie aufwendig ist die Umwandlung in SGML-Daten? Wer macht das? Wer bezahlt das? Und: Was tun mit den überhaupt nicht EDV-erfaßten Lieferungen von Band I und II? Das Scannen würde nur ASCII-Daten ohne bzw. ohne zureichende Strukturinformationen liefern, also einheitlich den reinen Wortlaut ohne zureichende Differenzierung zwischen den einzelnen lexikographischen Textelementen.

Ein weiteres Problem stellen die juristisch-wirtschaftlichen Aspekte eines elektronischen Goethe-Wörterbuches dar. Die Rechte am gedruckten GWb liegen nicht bei den Akademien, sondern ausschließlich beim Verlag, der CD- und Internetversionen, nach eigenem Bekunden, nur dann zustimmt, wenn sie ihn keinen Pfennig kosten und wenn der Absatz des gedruckten Werkes dadurch nicht beeinträchtigt wird. Die wirtschaftlichen Erwägungen hängen natürlich aufs engste mit den Fragen der Einrichtung einer Datenbasis und dem damit verbundenen Aufwand zusammen.

Ein dritter Problembereich ergibt sich aus den bisher befolgten Prinzipien der Textgestaltung. Von Anfang wird das GWb begleitet von der Losung, nicht schon *im* GWb das leisten zu wollen, was man *mit* dem GWb leisten können soll. Von daher erhebt sich die Frage, inwiefern die aus den erweiterten Benutzungsperspektiven resultierenden Postulate, wie sie soeben angerissen wurden, nicht doch schon ein Stück weit Nutzung, ergebnisorientierte Aufbereitung, in die semantische Arbeit des GWb hineinragen. Insbesondere eine weitergehende Standardisierung der Kennzeichnungssystematik, womit Ansatzpunkte für eine multiple Organisation des Wortschatzes geschaffen würden, erscheint als Eingriff in die primäre semantische Aufbereitung des Belegmaterials in den Redaktionen des GWb. Die Grenze zu einem ‚neuen GWb‘ wäre rasch überschritten. Wäre das schlimm, vor allem

angesichts der vielfältigen Möglichkeiten für die Erforschung der Goethezeitsprache, die wir oben umrissen haben? Schwer zu sagen. Ganz abgesehen davon, daß es stets mißlich ist, an einer bewährten Konzeption herumzubasteln, sind auch substantielle Verluste nicht auszuschließen: Das zupackende Etikettieren, die festlegende Verbuchung unter einfachen Kategorien und Rubriken, wie sie die Prästrukturierung des Materials für spätere Selektionsprozesse per Mausklick schließlich erfordert, ist im GWb bisher bewußt unterblieben. Standardisierte Kennzeichnungen wie ‚bergmannssprachlich‘, ‚rechtssprachlich‘, ‚metaphorisch‘, ‚polemisch‘ kommen zwar vor, werden jedoch mit Zurückhaltung eingesetzt, denn das Besondere von Goethes Sprache besteht nicht zuletzt in Grenzüberschreitungen und Ambivalenzen.

Hinzu kommt im Bereich der Textgestaltung, daß vor allem in den älteren Bänden ein sehr enger Zusammenhang zwischen unterschiedlichen Textelementen besteht, der mit dem Prinzip der elektronischen Modularisierung selbständiger Textelemente nur schwer vereinbar ist. So führt etwa die enge Korrelation zwischen Leitbemerkungen und Belegen dazu, daß man in den älteren Lieferungen die Belege nicht ohne weiteres ausblenden kann, wenn man die Leitbemerkungen verstehen will. In anderen Fällen gehören etwa eine Leitbemerkung und die entsprechenden Unterleitbemerkungen so eng zusammen, daß man nicht ohne weiteres einen der beiden Bestandteile ausblenden kann.

#### 4.4 Prototypen

Ein erster Arbeitsschritt beim Versuch, ein Teilstück eines elektronischen GWb herzustellen, bestand darin, die Verwendbarkeit der vorhandenen Winword-Daten zu überprüfen. Die Ergebnisse deuten darauf hin, daß mit Hilfe der verwendeten typographischen Auszeichnungen wie ‚doppelt unterstrichen‘ und mit Hilfe textsyntaktischer Kombinationsregeln wie ‚Fettdruck unmittelbar nach Neuer Absatz‘ immerhin Textelemente wie Leitbemerkungen, Belege oder Lemmata strukturell markiert werden können. Hinderlich dabei ist allerdings, daß typographische Auszeichnungen nicht eindeutig und anhand von textologischen Kriterien verwendet wurden, sondern anhand des Erscheinungsbildes in den späteren gedruckten Artikeln (z.B. Fettdruck nicht nur beim Lemma, sondern auch beim Kürzel *Syn*, mit dem die Synonyme eingeleitet werden). Man muß also versuchen, den mehrdeutigen Gebrauch von typographischen Auszeichnungen aufgrund der Abfolge und der Stellung von Textteilen im Artikel aufzulösen. Wir haben die Winword-Dateien hierfür zunächst in WordPerfect-Dateien umgewandelt, diese dann mit Hilfe des Tustep-Hilfsprogramms *konvert* in ASCII-Dateien, in denen typographische SteuerCodes explizit gemacht sind und per Programm manipuliert werden können (z.B. Zeichenketten wie *#fett\_ein*). Diese typographischen Markierungen wurden dann mit den Textmanipulationsmöglichkeiten von Tustep in SGML-Kennungen umgewandelt und, soweit möglich, textstrukturell aufgelöst. Manuelle Nachbearbeitung und Überprüfung der Ergebnisse wird wohl nicht zu umgehen sein, schon weil die Prinzipien der Texterfassung im Lauf der Bearbeitung des GWb nicht ganz einheitlich waren.

Für den auf diese Weise gewonnenen SGML-Text und seine Textstrukturmuster, die in einer bisher sehr elementaren Document Type Definition (DTD) erfaßt sind, haben wir in Panorama Pro 1.5 versuchsweise unterschiedliche ‚Ansichten‘ eingerichtet: z.B. eine Version, die der gedruckten Vollversion entspricht, und eine Version, bei der nur das extrahierte Skelett der semantischen Leitbemerkungen sichtbar ist und bei der nach Art des kleinen LEXER die Belege ausgeblendet sind.<sup>3</sup> Zu den nächsten Arbeitsschritten gehören vor

allem die Konzeption unterschiedlicher Internet-Darstellungen aufgrund der SGML-Datenbasis und die detaillierte Konzeption eines Markierungssystems, mit dem die spezifische Architektur des Goethe-Wortschatzes erschlossen werden kann.

## 5 Zusammenfassung und Ausblick

Wir sind ausgegangen vom Grundgedanken, daß ein elektronisches Wörterbuch als komplex strukturierte lexikologische Datenbasis erweiterbare Nutzungsperspektiven in zweierlei Hinsicht bietet. Zum einen kann eine solche Datenbasis ein Abbild der komplexen Zusammenhänge im Wortschatz sein, zum anderen kann sie ein Informationssystem sein, das bestimmte Benutzerinteressen besser und flexibler bedient als eine gedruckte Version. Am Beispiel des Goethe-Wortschatzes und des gedruckten Goethe-Wörterbuchs haben wir mit einigen wenigen Überlegungen den Grundgedanken eines mehrschichtig organisierten Wortschatzes erläutert und anzudeuten versucht, wie sich eine solche komplexe Organisation in elektronischer Form erfassen läßt und welche Arten des Erkenntniszugewinns sich daraus ergeben.

Ein wichtiger kultur- und forschungspolitischer Vorschlag war es, aufgrund der Stammdaten des elektronischen Wörterbuchs eine kostenlos angebotene Internet-Version herzustellen, um den Nutzerkreis des Goethe-Wörterbuchs zu erweitern. Vergleichbare Forderungen sind an alle Wörterbuch-Unternehmen zu stellen, die im wesentlichen mit öffentlichen Mitteln gefördert werden. Es steht zu hoffen, daß in absehbarer Zukunft nicht nur die technisch-konzeptionellen Aufgaben und Schwierigkeiten, sondern auch Fragen der Verwertungsrechte in ersprießlicher Weise gelöst werden.

Die Erschließung der Goetheschen Wortschatzarchitektur wird man sich jedenfalls als einen langen Weg der zunehmenden Verfeinerung vorstellen dürfen: „Wenn man denckt fertig zu seyn, gehts erst recht an“.<sup>4</sup>

## 6 Literatur

- Coombs, J.H./Renear, A.H./DeRose, S.J. (1993): Markup systems and the future of scholarly text processing. In: Landow, G.P./Delany, P. (eds.): *The digital word. Text-based computing in the humanities*. – Cambridge, Mass./ London, 85–118 (Zuerst: *Communications of the ACM* 30, 1987, 933–947).
- Deutsches Fremdwörterbuch (1988). Begonnen von H. Schulz, fortgeführt von O. Basler, weitergeführt im Institut für deutsche Sprache. Siebenter Band: Quellenverzeichnis, Wortregister, Nachwort. Hg. von A. Kirkness. – Berlin/ New York: de Gruyter.
- Dill, Ch. (1987): *Wörterbuch zu Goethes West-östlichem Divan*. [Mit einer Einführung zu Wortschatz und Bildwelt und einem Anhang zu Wortfeldern und Sinngruppen.] – Tübingen: Niemeyer.
- Fritz, G. (1998): *Historische Semantik*. – Stuttgart/Weimar: Metzler (= Sammlung Metzler).

<sup>3</sup> Eine ausführlichere Darstellung der technischen Realisierung dieser Konzeption eines elektronischen Goethe-Wörterbuchs soll in der Zeitschrift ‚Sprache und Datenverarbeitung‘ erscheinen.

<sup>4</sup> Goethe, WA, Abt. Briefe I 113,15.

- Goebel, U./Lemberg, I./Reichmann, O. (1995): Versteckte lexikographische Information. Möglichkeiten ihrer Erschließung dargestellt am Beispiel des Frühneuhochdeutschen Wörterbuchs. – Tübingen: Niemeyer (= Lexicographica, Series Maior 65).
- Goldfarb, Ch.F. (1990): The SGML handbook. – Oxford.
- Ide, N./Véronis, J. (eds.)(1995): Text encoding initiative. Background and context. – Dordrecht.
- Kluge, F. (1899): Etymologisches Wörterbuch der deutschen Sprache. Sechste verbesserte und vermehrte Auflage. – Straßburg: Trübner.
- und Götze, A. (1934): Etymologisches Wörterbuch der deutschen Sprache. Elfte Auflage. – Berlin: de Gruyter.
- Powitz, G. (1988): Das „Catholicon“ – Umriss der handschriftlichen Überlieferung. In: *Litterae medii aevi*. Festschrift für Johanne Autenrieth zu ihrem 65. Geburtstag. Hg. von M. Borgolte und H. Spilling. – Sigmaringen: Thorbeke, 209–223.
- Schadewaldt, W. (1946): Das Goethe-Wörterbuch. – Berlin 1946. Wieder in: Goethe, *Jahrbuch der Goethe-Gesellschaft* 11, 1949, 293–305.
- Tschirch, F. (1960): Bedeutungswandel im Deutsch des 19. Jahrhunderts. Zugleich ein Beitrag zum sprachlichen Verständnis unserer Klassiker. In: *Zeitschrift für deutsche Wortforschung* 16, N.F. 2, 7–24.
- Welter, R. (1998): Zwischen Bedeutung und Benutzer. Zur Mikrostruktur des Goethe-Wörterbuchs. In: Grosse, R. (Hg.): *Bedeutungserfassung und Bedeutungsbeschreibung in historischen und dialektologischen Wörterbüchern*. – Stuttgart/Leipzig, 145–149 (= Abhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig. Philologische-historische Klasse, 75/1).

*Thomas Gloning, Marburg  
Rüdiger Welter, Tübingen*