

# LEXICOGRAPHICA

International Annual for Lexicography  
Revue Internationale de Lexicographie  
Internationales Jahrbuch für Lexikographie

Edited by  
Fredric F. M. Dolezal, Alain Rey, Thorsten Roelcke,  
Herbert Ernst Wiegand, Werner Wolski, Ladislav Zgusta

in conjunction with the  
Dictionary Society of North America (DSNA)  
and the  
European Association for Lexicography (EURALEX)

Coordinating Editor: Andrea Lehr

10/1994

*Sonderdruck*

**Max Niemeyer Verlag  
Tübingen**



## DRW to FAUST

### Ein Wörterbuch zwischen Tradition und Fortschritt

- 1 Vorbemerkung
- 2 Beschreibung des Wörterbuchs
- 2.1 Konzeption des Wörterbuchs und Corpus-Erstellung
- 2.2 Quellenexzerption und Aufbau des Archivs
- 2.3 Die Herstellung von Wörterbuchartikeln
- 2.3.1 Abarbeitung des Zettelarchivs
- 2.3.2 Die Herstellung eines Wörterbuchbeleges
- 2.3.3 Die Gestaltung des Manuskripts
- 2.3.4 Die Herstellung des druckfertigen Manuskripts
- 2.3.5 Satz und Druck des Wörterbuchtextes
- 3 Desktop publishing im DRW
- 3.1 Präliminarien der Computerisierung
- 3.2 Der Einstieg
- 3.3 Textverarbeitung am Großrechner
- 3.4 Der Textsatz mit  $\text{\TeX}$
- 4 Die Textfassung über die EDV
- 4.1 Das Druckmanuskript
- 4.2 Die Belegerfassung über die EDV
- 4.3 Der Aufbau einer Quellensigle-Datei
- 4.4 Von der Belegdatei zur Manuskriptdatei
- 5 Die weitere Nutzung der Dateien
- 5.1 Recherchemöglichkeiten als Hilfsmittel der Wörterbucharbeit
- 5.2 Vom einfachen Suchbefehl zum Volltext-Retrieval-Programm
- 5.3 Das Problem der Inzuchtbelege
- 5.4 Erweiterung der Recherchegrundlagen durch OPTOPUS
- 5.5 Corpuserweiterung durch Exzerption?
- 6 Von der zeilenorientierten Textverarbeitung zur strukturierten Dateneingabe
- 6.1 Die Anfänge
- 6.2 Der Aufbau von FAUST
- 6.3 Die Quellen- und Titeldokumentation
- 6.3.1 Der Aufbau
- 6.3.2 Vom Herrschaftswissen zum Expertensystem
- 6.3.3 Faksimiles als Buchersatz
- 6.4 Die Wortartikeldokumentation
- 6.5 Die Erfassung eines Wortartikels
- 6.5.1 Der Artikelkopf
- 6.5.2 Die Verweisstruktur
- 6.5.3 Der Beleg
- 6.6 Das Druckformat
- 7 Die weitere Entwicklung der Datenverarbeitung
- 7.1 Wörterbuchparsing
- 7.2 Autarkie der Produktion
- 7.2.1 Von der Textdatei zum Druck
- 7.2.2 Ausblicke
- 8 Schlußbetrachtung
- 9 Literaturverzeichnis

## 1 Vorbemerkung

Dieser Aufsatz geht auf einen Vortrag zurück, den ich im Jahr 1991 im Rahmen des von HERBERT ERNST WIEGAND geleiteten lexikographischen Forschungskolloquiums in Heidelberg hielt. Der Anregung, diesen Vortrag in einen Aufsatz umzuformen, mochte ich damals noch nicht folgen: Zu sehr war die elektronische Datenverarbeitung im DEUTSCHEN RECHTSWÖRTERBUCH in der Entwicklung oder im Umbruch begriffen, als daß ich Endgültiges hätte niederlegen wollen. Die Präsentation der letzten Entwicklungen auf einer lexikographischen Tagung in Wien im Oktober 1993 und ein weiterer Vortrag vor dem Kolloquium H.E. WIEGANDS zeigte mir, daß nun die Zeit für ein Resümee gekommen sei. Ich schreibe daher einen persönlichen Erfahrungsbericht über die Einführung der elektronischen Datenverarbeitung in einem der großen historischen Wörterbücher Deutschlands aus der Sicht des Wörterbuchleiters, der zugleich an jedem Schritt dieser Entwicklung unmittelbar beteiligt war. Aus Erfahrung weiß ich, daß ein

solcher Bericht mit sehr unterschiedlichen Voraussetzungen gelesen wird.<sup>1</sup> Meine Absicht ist es, einige Möglichkeiten des Computereinsatzes bei der Herstellung eines Wörterbuchs aufzuzeigen. Dies mag denjenigen, die sich erst am Anfang dieses Weges befinden, einen Überblick über damit möglicherweise verbundene Schwierigkeiten, aber auch über das Sinnvolle des Computereinsatzes trotz dieser Schwierigkeiten vermitteln. Computerlinguisten und Programmierer hingegen können vielleicht neue Aufschlüsse über die alltägliche Praxis dessen gewinnen, womit sie sich in manchmal sehr theoretischer Weise beschäftigen müssen. Meiner eigenen Erfahrung nach irren Wissenschaftler zunächst ziemlich orientierungslos im Dickicht der Hard- und Software umher, bevor sie manchmal ein wenig irrational, zufällig und Terminzwängen unterliegend ihre Entscheidungen treffen. Vielleicht vermag ein solcher Bericht zur Klärung der eigenen Wege beizutragen.

## 2 Beschreibung des Wörterbuchs

Das DEUTSCHE RECHTSWÖRTERBUCH (DRW),<sup>2</sup> das sich mit seinem Untertitel als ein Wörterbuch der älteren deutschen Rechtssprache benennt,<sup>3</sup> ist ein historisches Wortwörterbuch, das die deutsche Allgemeinsprache in ihrem rechtlichen Bezug und mit ihren juristisch-fachsprachlichen Bestandteilen<sup>4</sup> vom Beginn ihrer schriftlichen Überlieferung bis etwa zum Jahr 1800 behandelt. Der zeitliche wie auch räumliche Erfassungsbereich des für das DRW relevanten Quellenmaterials macht dem Lexikographen deutlich, welche Fülle von sonst kaum in dieser Verbindung anzutreffenden Problemen sich daraus ergibt: Der Beginn der schriftlichen Überlieferung der „deutschen“ Rechtssprache liegt im 6. Jahrhundert; der Bearbeitungszeitraum des DRW reicht von dieser Zeit bis – spätestens – zum Jahr 1832<sup>5</sup> (mit Einschränkungen hinsichtlich der Behandlung von Kompositen, die erst nach 1700 belegt sind). Der räumliche Bearbeitungsbereich ergibt sich aus dem Gebrauch des im Titel benutzten Wortes „deutsch“, das im Sinne JACOB GRIMMS<sup>6</sup> das Westgermanische mit dem Altsächsischen, Altenglischen, Altfrisischen, Mittelniederländischen und dem Hoch- und Niederdeutschen beinhaltet.

Die Produktion eines historischen Wörterbuchs erfolgt in verschiedenen Schritten, die hier als Einstieg in die Geschichte des DRW kurz behandelt werden sollen.

### 2.1 Konzeption des Wörterbuchs und Corpus-Erstellung

Das DRW ist gegen Ende des letzten Jahrhunderts, in einer Zeit der Hochblüte der deutschsprachigen Lexikographie, begründet worden.<sup>7</sup> Bei seiner Konzeption durch eine Kommission

1 Dies veranlaßt mich auch, zumindest einige termini technici der unterschiedlichen Wissensgebiete, die hierbei eine Rolle spielen, in Fußnoten zu erklären. Allerdings spiegeln diese Erklärungen lediglich den Wissensstand wider, den ich erwerben mußte, um die Computerisierung des Rechtswörterbuchs vornehmen zu können.

2 Literatur zum DRW ist verzeichnet in SPEER 1991, I Anm. 1, der noch LAUFS 1993 hinzuzufügen ist.

3 Vgl. zur lexikographischen Konzeption des DRW SPEER 1989, 107 ff., und SPEER 1991, 681ff., mit weiterführender Literatur.

4 So verstehe ich im folgenden den Begriff „Rechtssprache“, vgl. dazu SPEER 1989, 113ff.

5 „Im Laufe der Zeit ist die zeitliche Grenze immer weiter hinausgeschoben und schließlich auf das Jahr 1832 gelegt worden, GOETHE'S Todesjahr. In der juristischen Auslegungskunst geschult, interpretierte man in der Arbeitsstelle die früher häufig für den zeitlichen Rahmen des Rechtswörterbuch gebrauchte Formel: „Von ULFILAS bis GOETHE“ in dem Sinne um, daß eben hier nicht das Geburts-, sondern das Todesjahr Goethes gemeint sei.“ DICKEL in: DICKEL/SPEER 1979, 23.

6 Vgl. SPEER 1989, 110ff.

7 Vgl. dazu DICKEL/SPEER 1979, 20ff.

der Preußischen Akademie der Wissenschaften in Berlin wurde relativ großzügig verfahren und vieles blieb ungeklärt: Eine so ausgearbeitete Wörterbuchplanung mit Aufschluß über alles Wissenswerte, wie dies in den letzten Jahren OSKAR REICHMANN für das FRÜHNEUHOCHDEUTSCHE WÖRTERBUCH (FWB)<sup>8</sup> vorgelegt hat, lag damals außerhalb des Bereiches des Möglichen. Die Folge waren und sind Änderungen der Konzeption mitten im Verlauf der Wörterbuchproduktion, auf die ich hier nicht näher eingehen kann.<sup>9</sup> Das Quellencorpus sollte in richtiger Einschätzung des Vorkommens der rechtsrelevanten Sprache außer den reinen Rechtstexten wie Rechtsbüchern und Gesetzestexten auch Quellen religiöser und dichterischer Art, Reiseberichte und Wirtschaftstexte enthalten – und dies aus dem gesamten westgermanischen Bereich.

Dieser Quellenbestand ist in einem Quellenheft und vier Ergänzungsheften dokumentiert. Die insgesamt rund 7 800 Quellenabkürzungen repräsentieren eine erheblich höhere Anzahl von Buchbindereinheiten.<sup>10</sup> Die Exzerptionsdichte je Quellensigle ist höchst unterschiedlich. Auch ohne die erst jetzt möglich gewordene Auswertung über die EDV läßt sich sagen, daß manche Quellen vermutlich überhaupt nicht exzerpiert worden sind (das erste Quellenverzeichnis stellte wohl einen Exzerptionsplan dar, der abgearbeitet werden sollte), daß andere Quellen nur sehr wenig Belege ergeben haben, manche hingegen sehr gut erschlossen sind.

## 2.2 Quellenexzerption und Aufbau des Archivs

Im Verlauf von rund siebzig Jahren<sup>11</sup> kamen rund 2,5 Millionen Belegzettel zusammen, die von einer Vielzahl von in- und ausländischen Wissenschaftlern für das Wörterbuch exzerpiert wurden.<sup>12</sup> Die Zahl der tatsächlich belegten Fundstellen ist jedenfalls wesentlich höher, da einzelne Zettel auch Verweise auf Wortregister wie demjenigen zum Corpus der altheutschen Originalurkunden oder dem Corpus van middelnederlandsche Teksten mit teilweise hunderten von Fundstellen zu einem Wort enthalten können.<sup>13</sup>

Diese Zettel wurden in die Forschungsstelle des DRW<sup>14</sup> nach Heidelberg geschickt und dort einer Stichwortform zugeordnet, die häufig ein Konstruktlemma darstellt.<sup>15</sup> Bei den unterschiedlichen Sprachen und Sprachvarietäten sowie dem langen Exzerptionszeitraum läßt es sich nicht vermeiden, daß Zettel, die unter sprachgeschichtlichem Aspekt *einem* Konstruktlemma hätten subsumiert werden müssen, an den unterschiedlichsten alphabetischen Stellen zu

8 Vgl. seine Einleitung zum FWB.

9 Vgl. aber DICKEL/SPEER 1979, 23ff.

10 Dies zeigt bereits die Anzahl der Abkürzungen für Zeitschriften und Jahrbücher: Es sind unter „Jb“ (für Jahrbuch) 68 und unter „Z“ (für Zeitschrift) 79 Abkürzungen verzeichnet.

11 Die systematische Exzerption von Quellen wurde 1974 beendet, da in der unmittelbar vorangehenden Zeit so exzessiv exzerpiert wurde, daß die Zahl der dem Archiv neu einverleibten Zettel diejenige der zu Wortartikeln verarbeiteten um ein Vielfaches überstieg.

12 Zum Teil waren es die Herausgeber von wichtigen Texteditionen, die ihre Edition für das DRW ausgewertet hatten. So hat FELIX LIEBERMANN in seiner charakteristischen und teilweise schwer lesbaren Schrift etwa 2.100 Archivzettel aus seiner immer noch maßgebenden Edition der altenglischen Gesetze beigeuert. Überwiegend erfolgte die Exzerption aber wohl in der Heidelberger Forschungsstelle. Die jährlichen Berichte in der Savigny-Zeitschrift für Rechtsgeschichte, Germanistische Abteilung, vermögen trotz ihrer Kürze ein interessantes Bild der an diesem Werk beteiligten Wissenschaftler und Studenten zu vermitteln – man trifft häufig auf Namen, die später in den verschiedensten Wissenschaftsbereichen bekannt geworden sind.

13 Für das Wort „Mann“ sind dies in den genannten Registern allein 2.448 Fundstellen.

14 Dies war die einzige Forschungsstelle, da der Heidelberger Rechtshistoriker RICHARD SCHRÖDER, der auch Mitglied der Preußischen Akademie war, die Leitung übernommen hatte.

15 Zu diesem Begriff vgl. die bei SPEER 1989, 111ff., angegebene Literatur.

finden sind.<sup>16</sup> Innerhalb der jeweiligen Stichwortform sind die Zettel aber völlig ungeordnet in Archivkästen zu jeweils etwa 1.600 Zetteln eingelegt worden.<sup>17</sup>

Damit deutlich wird, was diese Zettel für die Wörterbucharbeit bedeuten, was also Exzerption im DRW im Unterschied zu anderen Wörterbüchern<sup>18</sup> meint, sei dies kurz beschreiben: Exzerption ist im DRW in erster Linie das Herausschreiben von *Stichwörtern* – und nicht: *Belegtexten* – rechtlich relevanter Art aus einer Textedition<sup>19</sup> auf Archivzettel. Diese Zettel enthalten mindestens das Stichwort, die Quellenangabe, manchmal auch die Datierung des Textes, seltener den – mehr oder minder umfangreichen – Belegtext. Es handelt sich bei dem Ergebnis der Exzerption also nicht wie bei REICHMANN<sup>20</sup> oder bei der Belegsammlung des GRIMMSCHEN WÖRTERBUCHES<sup>21</sup> um dem lexikographischen Erkenntnisinteresse genügende Textausschnitte, die ohne weitere Überprüfung in das Artikelmanuskript übernommen werden können. Legt man den Exzerptionsbegriff zu Grunde, den REICHMANN 1990 entwickelt hat,<sup>22</sup> so ist die Hauptarbeit der Vorbereitung des Schreibens von Wortartikeln für das DRW noch nicht geleistet worden: Die „Exzerptoren“ des DRW haben in der Regel nur Fundstellen herausgeschrieben, während die eigentliche Exzerption<sup>23</sup> zugleich mit der Abfassung des Wortartikels erfolgt. Aber auch dort, wo der Archivzettel eine Belegabschrift enthält, muß er vom Bearbeiter des Wortartikels überprüft werden, da die tägliche Erfahrung gelehrt hat, wie ungenau die Abschriften vorgenommen worden sind. Die verschiedenen, teilweise schwer lesbaren oder bereits verblaßten Handschriften der Exzerptoren stellen den Bearbeiter ebenfalls vor besondere Probleme. Darüberhinaus sind die meisten Zitate zu knapp, so daß eine vorläufige Sortierung durch die Einsichtnahme in die Quelle häufig wieder umgestoßen werden muß.

- 
- 16 Freilich setzt dieses „Finden“ zum Teil äußerst zeitraubende etymologische Recherchen bis ins Indogermanische hinein voraus, um die Lemmaform endgültig klären und die Zuordnung der unterschiedlich angesetzten Lemmata hierzu festlegen zu können – eine Arbeit, die den Scharfsinn der philologischen Mitarbeiterinnen immer wieder herausfordert und die Nichtphilologen häufig auf Irrwege führt.
- 17 Der „Archivkasten“ ist im DRW die grundlegende Berechnungseinheit für die Projektplanung geworden: acht Archivkästen ergeben zur Zeit eine Lieferung mit 160 Druckspalten.
- 18 Viele Mißverständnisse rühren daher, daß Lexikographen sich nicht immer der Unterschiedlichkeit ihrer Arbeitsvoraussetzungen bewußt sind und beispielsweise unter Exzerption durchaus unterschiedliche Methoden und Techniken verstehen können.
- 19 Vereinzelt auch aus handschriftlichen Quellen, die bei der Artikelarbeit nicht mehr eingesehen werden können.
- 20 Aus der Einleitung zum FWB, 48ff., ergibt sich nur implizit, daß hier zur Exzerption immer auch das „Ausschreiben“ der Belegtexte gehört. Vgl. aber zur Gesamtproblematik REICHMANN 1990 a passim.
- 21 BAHR 1962.
- 22 REICHMANN 1990 a, 1607, These 19: „Exzerpieren heißt a) den Belegtext übernehmen, b) die Wortart bestimmen, c) die Eckformen der Flexion bestimmen, d) Beispielsyntagmen herauslösen, e) syntagmatische Grundformen bestimmen, f) bedeutungsverwandte Wörter aus dem Belegumfeld gewinnen, g) den Grad der Lexikalisierung von Wortbildungen und der Phraseologisierung von Einheiten beurteilen.“
- 23 Daher läßt sich der Satz REICHMANNs (1990 a, 1607, These 18): „Mit dem Gesagten wird keineswegs bestritten, daß die Exzerption auch technische Aspekte hat und z.B. pro Stunde zu einem durchschnittlichen Ausstoß von rund 20 Belegzetteln führen muß“ auf die Exzerptionstätigkeit im DRW leider nicht anwenden. Das von Reichmann geforderte Ergebnis setzt die kontinuierliche Exzerption eines Quellentextes für eine bestimmte Alphabetsstrecke bzw. für alle Wörter des Alphabets voraus, während im DRW bei der Arbeit an einer Manuskriptstrecke im optimalen Fall nur einige wenige Belege zu einem Wort aus einem Buch exzerpiert werden können.

## 2.3 Die Herstellung von Wörterbuchartikeln

### 2.3.1 Abarbeitung des Zettelarchivs

Die Belegtexte des DRW beruhen auf diesen Archivzetteln, die auf die exzerpierten Quellenwerke verweisen. Letztlich ermöglicht ein mit Belegtext ausgefüllter Zettel es dem Mitarbeiter nur, eine Vorentscheidung über die Weiterbearbeitung dieser Quellenstelle zu treffen. Da die Mehrzahl der Archivzettel jedoch keinen Belegtext enthält, kann die Wichtigkeit der durch den Zettel erschlossenen Quellenstelle für den Wortartikel höchstens aus der Textsorte – falls die Quellensigle dem erfahrenen Bearbeiter Aufschluß über sie gewährt – und der zeitlichen Einordnung im Vergleich zur sonstigen Beleglage abgelesen werden. Zum Teil vermag auch die Handschrift des Exzerptors<sup>24</sup> einen Hinweis darüber zu geben, ob die Fundstelle für den Wortartikel ergiebig sein wird. Die auf dem Zettel vorhandene Quellenangabe kann ein erhebliches Problem darstellen, da besonders in der Anfangsphase der Exzerption die Abkürzungen teilweise erst nachträglich – d. h. nach dem Einsortieren der Zettel in das Archiv – festgelegt worden waren. Insofern ist eine gewisse Erfahrung im Umgang mit dem Archiv erforderlich, um die – in solchen Fällen überholten und manchmal recht kryptischen – Quellenangaben auf dem Archivzettel in eine korrekte Quellenabkürzung umwandeln zu können.

### 2.3.2 Die Herstellung eines Wörterbuchbeleges

Für die Herstellung eines Beleges anhand eines Belegzettels bedeutet dies:

- Entzifferung des Zetteltextes
- Verifizierung der Quellensigle
- Nachschlagen der Signatur des Quellenwerkes
- Transport des Werkes vom Regal zum Schreibtisch oder Bestellung bei einer Bibliothek
- Einsichtnahme in den Quellentext
- gegebenenfalls Umlesen auf eine neuere Edition
- Suchen der Fundstelle<sup>25</sup>
- sinnfällige Begrenzung des herauszuschreibenden Textausschnittes
- genaue Abschrift des Belegtextes<sup>26</sup>
- dabei Verifizierung der Datierung (insb. Verhältnis Entstehungs-/ Überlieferungszeit und Textkritik)
- Zurückstellen oder Abgeben des Quellenwerkes.

### 2.3.3 Die Gestaltung des Manuskripts

Aus der herausgeschriebenen Menge von Belegen wird ein Wortartikel mit einer oder mehreren Bedeutungsangaben – jeweils in chronologischer Reihung der Belege – gefertigt. Das von den Mitarbeitern zur Endredaktion vorgelegte Manuskript bestand zu Beginn meiner Tätigkeit im DRW – wie vermutlich schon zu Beginn der Wörterbucharbeit – aus DIN-A4-Seiten, auf die die Belegtexte und Quellenangaben mit Hand oder Schreibmaschine geschrieben waren.

24 So bei den Rechtshistorikern HEINRICH BRUNNER oder RICHARD SCHRÖDER, die aber auch vor Abschreibefehlern nicht gefeit waren.

25 Handelt es sich z.B. bei Urkundenbüchern um die Angabe einer Seite oder der Urkundennummer – und welchen Band des Mecklenburgischen Urkundenbuches muß ich beispielsweise bestellen, wenn mein Beleg in der Urkunde Nummer 13.544 steht?

26 Unter Berücksichtigung der Regeln, nach denen diakritische Zeichen oder Abkürzungen im DRW aufgelöst und Schreibweisen vereinheitlicht werden (Kleinschreibung aller Wörter und der lateinischen Zahlen außer der Eigennamen).

Diese Technik implizierte eine gewisse Endgültigkeit und Unveränderbarkeit des Wörterbuchtextes schon in einem recht frühen Stadium der Bearbeitung.<sup>27</sup> Denn eine Korrektur war nur durch das umständliche Zerschneiden und Zusammenkleben nach der neuen Anordnung der Bedeutungen und Belege möglich. Eine erhebliche Erleichterung brachte die vom damaligen Wissenschaftlichen Leiter, GÜNTER DICKEL, eingeführte Neuerung,<sup>28</sup> daß jede Bedeutungserklärung und jeder Beleg separat auf Zettel im Format der Archivzettel geschrieben werden mußte. Diese Bestandteile des Artikels konnten ohne die Zuhilfenahme von Schere und Klebstoff leicht neu geordnet werden.

### 2.3.4 Die Herstellung des druckfertigen Manuskripts

Die in einem Archivkasten hintereinander geordneten Bedeutungs- und Belegzettel wurden in der Endredaktion vom Wissenschaftlichen Leiter und mir überarbeitet und schließlich von einer wissenschaftlichen Mitarbeiterin nach den Angaben des Verlages mit Schreibmaschine geschrieben, wobei die Satzauszeichnung mit Farbstift erfolgte. Mehrfache und sehr zeitaufwendige Korrekturgänge begleiteten die Herstellung der Satzvorlage, bis das Manuskript an den Verlag abgeschickt werden konnte.

### 2.3.5 Satz und Druck des Wörterbuchtextes

Der Verlag übergab das Manuskript einer Setzerei, in der erfahrene Setzer den Monotype-Bleisatz des Textes herstellten. Im Verlauf der Zeit wurde der Satz freilich immer problematischer, da Verlag und Setzerei in der damaligen DDR angesiedelt waren und die Setzmaschinen wohl noch aus der Vorkriegszeit stammten, so daß entsprechende Verschleißerscheinungen bei den Drucktypen das Druckbild beeinträchtigten. Auch die Grundlinien der Zeilen verrutschten öfters; das Druckbild war nicht immer so perfekt, wie wir es uns gewünscht hätten. Vom Verlag kamen die Fahnenkorrekturen nach ein bis zwei Jahren in das DRW zurück – die Frist zwischen der Ablieferung unseres Manuskriptes und dem Satz konnte auf Grund der wirtschaftlichen Situation in der damaligen DDR niemals genau vorhergesagt werden. Die Fahnenkorrektur war ein nochmaliger äußerst zeitraubender Arbeitsvorgang, der mehrere Personen in der Forschungsstelle wochenlang von der Arbeit an Wortartikeln abhielt.

## 3 Desktop publishing im DRW

### 3.1 Präliminarien der Computerisierung

Das DRW ist bei der Umstellung auf die computergestützte Lexikographie die eben beschriebenen Schritte rückwärts gegangen. Dies scheint mir der Hauptgrund dafür zu sein, daß die Computerisierung des DRW nicht nur von keinem zeitweiligen Stillstand der Produktion begleitet war, sondern daß sich bereits in dieser Phase Beschleunigungseffekte ergaben. Wenn ich diesen Ablauf historisch darstelle, so muß ich mit der Erstellung des Druckmanuskriptes mit Hilfe der EDV beginnen. Doch vor diesem Beginn standen in der Geschichte des Verhältnisses zwischen DRW und EDV eine Frage und eine Antwort:

27 Bei einigen Wortartikeln läßt sich die Durchbrechung der chronologischen Reihenfolge der Belege nicht anders deuten, als daß der Zettelkasten der Reihe nach abgearbeitet worden war, ohne daß eine Endredaktion dieses Zufallsergebnis korrigiert hätte.

28 Seine maßgebliche Beteiligung an der Wörterbuchreform gegen Anfang der Siebzigerjahre geht nur ungenügend ans DICKEL/SPEER 1979, 20-27, und seinen jährlichen Berichten im Jahrbuch der Heidelberger Akademie der Wissenschaften hervor.

In den Sechzigerjahren besuchte eine Gruppe von rechtshistorisch interessierten Jurastudenten die Forschungsstelle des DRW. Einer dieser Studenten fragte den damaligen Forschungsstellenleiter, wie es denn mit dem Einsatz der EDV in der Wörterbucharbeit beschaffen sei. Die Antwort meines Vorgängers war kurz und vernichtend: „Junger Mann, wir treiben hier Wissenschaft!“ Erzählenswert wird diese Anekdote freilich erst dadurch, daß dieser Jurastudent sie selbst weitererzählte, als er Jahre danach in Amt und Würden war – als Referent des baden-württembergischen Wissenschaftsministeriums war er zuständig für die Heidelberger Akademie und damit indirekt für das Deutsche Rechtswörterbuch. Er schloß seine Erzählung mit der ironischen Vermutung, im DRW würden die Manuskripte wohl immer noch mit dem Federkiel geschrieben. Vielleicht lag es auch an diesem Vor-Urteil, daß ich bei meinen 1986 einsetzenden Bemühungen, das DRW zu computerisieren, weder in der Akademie der Wissenschaften noch im Ministerium auf Widerstand stieß, sondern im Gegenteil alle nur mögliche Unterstützung genoß.

Allerdings war ich bereits dreizehn Jahre lang Leiter der Forschungsstelle des DRW, bevor ich die Computerisierung des Wörterbuchs in Angriff nahm. Woran also lag es, daß ich dieses Projekt erst jetzt begann? Der Hauptgrund war wohl die ungenügende Information über die Möglichkeiten des Computereinsatzes im lexikographischen Arbeitsprozeß sowie eine zwar vage, aber eben deshalb für einen Geisteswissenschaftler abschreckende Vorstellung von den Kosten einer solchen Umstellung. Zwar kam mir im späteren Verlauf dieses Prozesses der stetige Preisverfall auf dem Hardware-Markt zugute. Dennoch verlangte es zunächst ein Umdenken bei allen Beteiligten, daß nun auch ein geisteswissenschaftliches Institut in größerem Umfang Geldbedarf für Geräte anmeldete, wie dies sonst nur bei naturwissenschaftlichen Einrichtungen üblich war.

Der Grund, dem Ganzen näher zu treten, war zwingend: 1986 schied die Mitarbeiterin, die neben ihrer wissenschaftlichen Arbeit das Druckmanuskript geschrieben und die Satzauszeichnungen eingefügt hatte, aus Altersgründen aus. Die allgemein gestiegene Akzeptanz des Computereinsatzes in den Geisteswissenschaften schien – mehr unbewußt – auch mich erreicht zu haben, so daß ich die Lösung eines solchen Personal- und Produktionsproblems in diesem Bereich suchen konnte. Niemand freilich, der mit dem DRW zu tun hatte, besaß irgendeine Erfahrung mit Computern.

### 3.2 Der Einstieg

Auf dem Weg durch das schwierige Gefilde der Datenverarbeitungstechnik wäre ich zunächst nicht weit gekommen, hätte ich nicht einen Führer gefunden, der mich – der Vergleich sei erlaubt – durch die Unterwelt dieser modernen Technik geführt hat wie dies VERGIL mit DANTE getan hatte.<sup>29</sup> Eine Mitarbeiterin hatte mir von einem Kollegen in einer studentischen Theatergruppe erzählt, der seine germanistische Dissertation mit dem Computer schriebe, und mir weiterhelfen könne. In AXEL HECKER hatte ich einen Geisteswissenschaftler gefunden, der imstande war, die Bedürfnisse eines Wörterbuchunternehmens in die Möglichkeiten der damaligen Computertechnik zu transponieren. Er beriet mich hinsichtlich der verschiedenen Betriebssysteme am Universitätsrechenzentrum (URZ) und ebenso der Textverarbeitungsmöglichkeiten am Großrechner. Und ich war seines Mitgefühls und Verständnisses sicher, wenn ich an Grenzen stieß und sie nicht akzeptieren wollte. Es war ein ausgesprochener Glücksfall, daß wir sehr

<sup>29</sup> Der Vergleich liegt auch deshalb nahe, weil das erste Betriebssystem, mit dem ich es zu tun haben sollte, zwar nicht INFERNO, wohl aber HADES (für: Heidelberger Automatisches Datenverwaltungs- und Edittier-Subsystem) hieß.



schnell eine gemeinsame Basis gefunden hatten, auf der die unterschiedlichen Voraussetzungen, von denen wir beide ausgingen, kaum eine Rolle spielten

Zunächst ging es nur darum, mit Hilfe der EDV ein Typoskript herzustellen, das dem Verlag und der Druckerei als Satzvorlage dienen konnte. Dazu mußten vier verschiedene Schriftarten bzw. -größen dargestellt werden, daneben aber auch diakritische Zeichen, die in jedem historischen Text selbstverständlich sind: Dehnungszeichen, übergeschriebene Vokale oder Konsonanten mit Vokalwert (v und w), es gab Sonderzeichen wie das Pfennigzeichen oder das Pfund (die libra, nicht zu verwechseln mit dem Zeichen für Pfund Sterling: £), sowie altenglische Runenzeichen wie das stimmhafte und stimmlose Thorn und die Abkürzung für „and“.

Es war nicht leicht, den verschiedenen Computerfirmen und auch den Textverarbeitungsspezialisten im URZ begreiflich zu machen, worum es hierbei ging. Noch war die Rede nicht vom Computersatz eines Wörterbuchs, sondern nur von der Darstellung der für uns unbedingt erforderlichen Schriftzeichen, so wie es im Typoskript – wenn auch nur „händisch“ mit Bleistift – möglich war. Die Mißverständnisse häuften sich. So schien es in den Gesprächen kein Problem zu sein, übergeschriebene Buchstaben herzustellen – was unsere Gesprächspartner allerdings darunter verstanden, war nur die Möglichkeit, nach dem Druck eines Zeichens ein Zeichen zurückzugehen und ein anderes Zeichen auf der gleichen Zeilenhöhe darüberzudrucken. Ein Textverarbeitungsspezialist im URZ konfrontierte mich mit dem Satz: „Bevor Sie mir Fragen stellen, lernen Sie erst einmal, unsere Sprache zu sprechen!“

### 3.3 Textverarbeitung am Großrechner

Nach vielen Mühen konnte die Entscheidung für die Nutzung der Textverarbeitungsmöglichkeiten des Rechenzentrums getroffen werden. Eine Verbindung zu dessen Großrechner war durch die Anbindung an eine IBM-Steuereinheit<sup>30</sup> im Nachbargebäude möglich geworden. Zwei Terminals<sup>31</sup> konnten angeschafft werden, eines für mich und eines für die mit einer Diplombibliothekarin besetzte Dokumentationsstelle des DRW, über die auch der Bestellverkehr mit der Heidelberger Universitätsbibliothek sowie der interne bibliothekarische Geschäftsablauf abgewickelt werden sollte.

Für längere Zeit war ich der einzige, der in den Produktionsprozeß des Wörterbuchs auf diese Weise eingebunden war. Es ging zunächst nur um das Schreiben des Manuskripts mit Hilfe der EDV, und das hatte ich – gerade weil ich die Erprobung der neuen Technik an niemanden delegieren konnte und wollte – selbst übernommen. Konkret bedeutete dies, daß alle Probleme zunächst *meine* Probleme waren und von mir – wenn auch häufig durch die Konsul-

30 Eine Steuereinheit ist ein Gerät, das die Verbindung zwischen mehreren Terminals und einem räumlich entfernten Rechner herstellt – in unserem Fall über ein Modem, das eine normale Telefonleitung für den Datentransfer nutzt. Der Nachteil der Anbindung an den Großrechner über ein Modem bestand damals in der Langsamkeit der Datenübertragung, d.h. daß die Texteingabe am Terminal erst zum Großrechner geschickt und dort verarbeitet wurde, bevor neuer Text eingegeben werden konnte. Die Antwortzeiten waren bei Hochbetrieb recht lang, so daß ich eine nur scheinbar paradoxe Ansicht für realistisch halte: Daß nämlich ein erheblicher Teil der Arbeitszeit am Bildschirm mit dem Warten auf Programmabläufe verbracht wird. Heute spielen weder die Steuereinheit noch die Beschleunigung über Hochgeschwindigkeitsmodems eine Rolle, da wir für die seltenen Fälle, in denen wir die Ressourcen des Großrechners noch benötigen, einen schnellen ISDN-Anschluß mitbenutzen können, bis einmal ein Glasfaser-Hochgeschwindigkeitskabel die Heidelberger Altstadt mit dem Universitätsgebiet in der Rheinebene verbinden wird, in dem das Rechenzentrum steht.

31 Ein Terminal ist eine nichtintelligente Ausgabereinheit (ein PC ist intelligent, da er eigene Rechenleistung besitzt), auf deren Bildschirm die Tastatureingaben dargestellt und von wo sie an den Großrechner zur Bearbeitung weitergegeben werden.

tation anderer – gelöst werden mußten. Diese zwangsläufig hergestellte Verbindung zwischen der Projektleitung und der konkreten Erprobung der EDV erscheint in der Retrospektive immer noch sinnvoll,<sup>32</sup> auch wenn dies eine erhebliche zusätzliche Arbeitsbelastung neben der redaktionellen Arbeit bedeutete.

Die Probleme fingen ungeachtet aller Vorklärunen bei der Eingabe von Text an. Die von AXEL HECKER im URZ geschriebenen Makros<sup>33</sup> für das Textverarbeitungssystem WSCRIPT<sup>34</sup> sorgten zwar für den Druck von Sonderzeichen und übergeschriebenen Vokalen. Auch der Schriftartenwechsel wurde durch die Eingabe von Steuerungszeichen bestimmt. Aber: Die Steuereinheit war auf die amerikanische Tastatur eingestellt, so daß die Umlaute und das „ß“ auf der Tastatur nicht vorgesehen und auf dem Terminal nur in einer kodierten Form zu sehen waren. Im URZ wiederum schienen andere Zeichensätze<sup>35</sup> zu gelten, so daß Umwandlungstabellen geschrieben werden mußten, damit ein auf meiner Tastatur eingegebenes Zeichen am Großrechner dasselbe Zeichen erzeugte. Das Ergebnis war schließlich ein auf dem Drucker im URZ gedrucktes Manuskript im DIN-A4-Format mit den richtigen Schriftarten (recte, kursiv, halbfett, petit), keineswegs aber bereits allen Sonderzeichen. Das am Nachmittag im URZ gedruckte Manuskript sollte ursprünglich von uns dort (am anderen Ende der Stadt) abgeholt werden, erst eine Reihe von Verhandlungen ermöglichte die Verteilung mit der universitätsinternen Post, so daß der Text am nächsten Morgen in der Forschungsstelle verfügbar war.<sup>36</sup>

### 3.4 Der Textsatz mit T<sub>E</sub>X

Ich schickte dem Verlag HERMANN BÖHLAUS Nachfolger in Weimar, der das DRW seit dem Beginn der Produktion verlegt hatte, eine Probeseite mit der Frage, ob dies als Vorlage für den Setzer genüge und wie viele dieser Seiten eine Lieferung ergäben. Die Antwort zeigte mir, daß eine unbefangene Begegnung mit neuen Techniken auch richtungweisend sein kann: Der Verlag fragte zurück, ob wir nicht mit geringfügig anderen Schriften und mit anderer Seitengröße eine Vorlage für den Offsetdruck liefern könnten. Mit Vergrößerung und Verkleinerung ließen sich – so meinte der Verlag – Seiten herstellen, die im Verlag zerschnitten und zweiseitig, mit Kolummentiteln versehen, zu Lexikonseiten zusammengefügt werden konnten. Wir überlegten uns die Praktikabilität dieses Verfahrens, aber während ich noch versuchte, mich mit diesem Weg

32 Vgl. das Schlußwort.

33 Ein Makro ist eine Bündelung von Befehlen, die für einen häufig wiederkehrenden Bedarf zusammengestellt werden können und bei Aufruf der Reihe nach abgearbeitet werden.

34 „WSCRIPT ist ein Lizenzprogramm der University of Waterloo, Canada. Es entspricht bei niedrigerem Preis weitgehend dem IBM-Programm *Document Composition Facility (DCF)/SCRIPT*. WSCRIPT ist ein leistungsfähiger durch Optionen (beim Programmaufruf) und Befehle (in der eigentlichen Eingabe) gesteuerter Textformatierer und darüberhinaus eine textverarbeitende interpretierende Programmiersprache.“ So WALSDORFF 1987, S. 1. Entscheidend ist der Unterschied zu Textverarbeitungs-systemen wie WORD und WORDPERFECT: Die Formatierung des Textes muß explizit vorgenommen werden und formatiert den Text erst in einem weiteren Bearbeitungsschritt, während die letztgenannten Systeme mit – in der Regel – nicht sichtbaren Steuerzeichen für Absatzgestaltung, Hervorhebung von Zeichen usw. arbeiten.

35 Es gibt verschiedene Zeichensätze, die entweder im ASCII- oder EBCDI-Format vorliegen. Das letztere wird in der Regel auf Großrechnern benutzt, während das ASCII-Format auf den PC's des Industriestandards – sprich: IBM und Kompatiblen – verwendet wird. Im ASCII-Format gibt es wiederum verschiedene länderspezifische Zeichensätze. Nachdem wir auf den PC's des DRW später lange mit der Codepage 437 gearbeitet hatten, habe ich wegen der Verfügbarkeit der Bildschirmdarstellung von bestimmten Sonderzeichen auf die Codepage 850 (ISO-Latein) umgestellt.

36 Auch diese Verfügbarkeit gehörte zu den Voraussetzungen, um den EDV-Einsatz optimieren zu können. Erhält der Bearbeiter seinen Text erst nach einigen Tagen, so führt dies in jedem Fall wegen des zu langen Zeitraums zwischen Bearbeitung und Korrektur zu Verzögerungen.

anzufreunden, dachte AXEL HECKER schon weiter. Er hatte seine Dissertation mit TEX geschrieben und kannte die Möglichkeiten dieses von DONALD E. KNUTH – ursprünglich für den Mathematiksatz – entwickelten Textsatzsystems.<sup>37</sup> Er traute sich zu, die entsprechenden TEX-Makros zu schreiben, mit denen der Satz des Manuskriptes im bisherigen Layout erfolgen konnte. Neben den beschriebenen Sonderzeichen waren dies der zweiseitige Druck mit Kolumnentiteln und die Spaltennumerierung (später wurde noch auf Wunsch des Verlages die Bogennumerierung hinzugefügt).

Einige Probleme tauchten erst im Lauf der Zeit auf: Die Formatierung war gelungen, über den beiden Spalten einer Seite standen die Lemmata der Wortartikel, die auf dieser Seite zu finden waren. Aber es sollten lebende Kolumnentitel<sup>38</sup> sein. Wenn am Beginn einer Seite Lautverweise standen, die nicht den Status eines Wortartikels besaßen, so durfte das Programm nicht den Verweiseintrag oder den letzten Lemmaeintrag der vorigen Seite in den linken Teil des Kolumnentitels stellen, sondern mußte das nächste Lemmazeichen benutzen. Es bedurfte einiger Programmierarbeit, um dieses Ergebnis zu erzielen. Aber auch noch andere Möglichkeiten mußten berücksichtigt werden, die zu diesem Zeitpunkt noch nicht aktuell waren: Der Beginn eines neuen Buchstabens, der Beginn eines neuen Bandes – all dies ist in der damals erstellten Makroprogrammierung bereits enthalten.

Es wurde zunächst versucht, einen vergrößerten Ausdruck dieses Manuskripts als Vorlage für den photomechanischen Druck zu nutzen. Damit sollte die geringe Auflösung von 300 dpi<sup>39</sup> auf einem der damals noch teuren Laserdrucker verbessert werden, indem das Format zunächst um 20 % vergrößert ausgegeben und anschließend wieder auf das ursprüngliche Format verkleinert wurde. Dadurch konnte die Auflösung auf 360 dpi erhöht werden. Gegenüber der etwa dreifachen Auflösung bei modernen Lichtsatzanlagen kam diese Druckqualität nicht in Frage. Die Lösung ließ sich nur bei einer Setzerei suchen, die TEX-Dateien verarbeiten konnte. Die Universitätsdruckerei STÜRTZ in Würzburg bot diese Möglichkeiten. Nun aber sollte das Druckbild dem bisherigen Monotype-Bleisatz-Bild des DRW so ähnlich wie möglich sein. Wir schickten also Vorlagen an die Druckerei und bekamen daraufhin Metric-Fonts geliefert, die in TEX eingebunden wurden und Zeichen erzeugten, die die Dicke<sup>40</sup> der später bei STÜRTZ benutzten Monotype Times-Roman-Schrift darstellten.<sup>41</sup> Dies genügte aber, um den Umbruch am Schreibtisch vorzunehmen, da der Zeilenfall, den wir in Heidelberg sahen, identisch mit demjenigen war, den die Lichtsatzanlage ausgab. Alle Korrekturvorgänge konnten und mußten somit abgeschlossen sein, bevor die Datei das Haus verließ. Da TEX den Zeilenfall jeweils für einen Absatz berechnet, mußte der sofort zur Verfügung stehende – und potentiell endgültige – Zeilenfall in späteren Bearbeitungsschritten nur noch in den jeweils veränderten Absätzen kontrol-

37 TEX in einer Fußnote zu beschreiben, ist ein unmögliches Unterfangen. Ich verweise daher auf die umfangreiche Literatur hierzu – SCHWARZ 1987, WONNEBERGER 1987, KOPKA 1991, vor allem aber natürlich auf KNUTH 1987 und die laufenden Veröffentlichungen der DANTE, Deutschsprachige Anwendervereinigung TEX e.V. (Postfach 10 18 40, 69008 Heidelberg, oder: dante@dante.de).

38 Bei lebenden Kolumnentiteln bezieht sich der erste Eintrag über den Spalten auf den Artikel, mit dem die Spalte beginnt, auch wenn der Lemmaeintrag selbst auf der vorherigen Spalte steht und dort bereits als zweiter Eintrag im Kolumnentitel zu finden ist.

39 dpi ist eine Maßeinheit für die Auflösung von Zeichen: dots per inch. 300 dpi sind 300 Punkte je Zoll; heute sind Laserdrucker mit 600 dpi Auflösung keine Seltenheit mehr.

40 Termims technicus für die Breite eines Schriftzeichens.

41 Metric-Files erzeugen graphische Schriftzeichen für den Bildschirm- oder Druckertreiber, die genauso breit wie die Originalschriften, aber mit diesen nicht identisch sind. Wären sie identisch, so benötigte man die Druckerei nicht mehr. Inzwischen ließe sich diese Einschränkung durch die Verwendung von Postscript-Schriften, die leicht erhältlich sind, aufheben; vgl. dazu 7.2.

liert werden. Es handelte sich zwar einerseits um eine gravierende Arbeitsverlagerung von der Setzerei auf die Forschungsstelle, andererseits aber mußte die vom Redaktor abgegebene Druckvorlage auch früher schon korrekt sein, so daß der Vorteil bei weitem überwog. Wir mußten den sorgfältig geschriebenen Text nach dem Satz nicht wiederum korrigieren, sondern konnten auf diesen letzten zeitraubenden Korrekturvorgang verzichten. Und die lange Vorlaufzeit für den Druck eines Manuskriptes konnte zur Optimierung der inhaltlichen Qualität genutzt werden, da die Verweilzeit in der Forschungsstelle sich verlängerte und dadurch über einen längeren Zeitraum hinweg inhaltliche Korrekturen angebracht werden konnten.

Voraussetzung für das Erreichen dieses Zieles waren freilich auch Kenntnisse typographischer Details:<sup>42</sup> Das DRW wurde mit einer 10-Punkt-Schrift gedruckt. Die entsprechende Schrift von Stürz wirkte aber ganz anders – ich mußte erst den Unterschied zwischen point und Didot<sup>43</sup> lernen. Und schließlich die Sonderzeichen: Die beiden altenglischen Runenzeichen und das libra- wie das Denarzeichen konnte Stürz später zur Verfügung stellen – aus Fonts, die sonst nur für Texte in phonetischer Umschrift gebraucht wurden.<sup>44</sup> Es funktionierte schließlich, auch wenn der Weg bis dahin sehr schwierig war. Die Vorbereitung des Textsatzes war damit gelungen, worin bestand der nächste Schritt?

## 4 Die Texterfassung über die EDV

### 4.1 Das Druckmanuskript

Das in Zettelform vorliegende Mitarbeitermanuskript wurde von mir seit Ende 1986 an einem Terminal in eine Datei übertragen; die Bearbeitung mit T<sub>E</sub>X erfolgte auf dem Großrechner des Rechenzentrums. Ursprünglich arbeitete ich in HADES mit WSCRIPT und mußte den Text umständlich mit den für WSCRIPT gewählten vierstelligen Steuerungszeichen eingeben. Jedoch waren diese Steuerungszeichen immer noch komfortabler als die originalen oder makrogesteuerten T<sub>E</sub>X-Befehle. Zur Illustration füge ich hier Beispiele ein, wobei das erste Beispiel die WSCRIPT-Steuerungszeichen zeigt:

```

app.mitteilen
&ue.v.&ue.
&fe.I.&fe. &ue.etw. (offiziell) mitteilen, verkünden, Bescheid geben.&ue.
dez wir mit gu&otem willen ^^m^i^t^ ^t^a^i^l^e^n^^ wellen, dem gotzheus
ze V., diu [&ua.Ansehl der&ue.] chint, diu V.W. het
&ue.1325 Inderedorf I 58.&ue.

```

42 Deren Mangel bei dem sogenannten Desktop publishing manchmal zu ästhetisch fragwürdigen Ergebnissen führt! Inzwischen gibt es allerdings auch auf diesem Gebiet Hilfsmittel, vgl. GULBINS/KAHRMANN 1993.

43 Point ist die amerikanische typographische Größeneinheit (2.8453 pt = 1 mm), während Didot eine der europäischen Größeneinheiten (2.6596 dd = 1 mm) ist. Aber auch hier gibt es Unterschiede: Der alte Didot-Punkt besitzt eine Größe von 0,37606 mm, der neue (seit 1978) ist genau 0,376 mm groß. Darüberhinaus gibt es noch den DTP-Point von 0,3528 mm. Mit Recht schreiben GULBINS/KAHRMANN 1993, 51: „Wenn also von einem Punkt oder Point gesprochen wird, muß man wissen, welches Maßsystem gemeint ist, auch wenn die Unterschiede zunächst gering erscheinen. Die Unterschiede... addieren sich auf einer Seite zu Zeilen auf!“ Vgl. auch SCHWARZ 1987, 51ff. für T<sub>E</sub>X.

44 Bei der Einbindung dieser Zeichen in T<sub>E</sub>X mußte ich freilich erst Umrechnungen aus dem mir bis dahin nicht bekannten Oktal- in den Dezimalcode (diese Codes beschreiben die Adressierung eines Zeichens für den Computer) vornehmen, bis das Ergebnis korrekt war und der Durchlauf von T<sub>E</sub>X nicht mehr zu Programmabbrüchen führte.

Dieser Eingabetext wurde durch ein Konvertierungsprogramm,<sup>45</sup> das AXEL HECKER in der Programmiersprache C geschrieben hatte, in TEX-Makros verwandelt. Zugleich damit erfolgte die Kontrolle der korrekten Verwendung dieser Zeichen.<sup>46</sup> Die TEX-Eingabedatei sieht so aus:

```
\keymark{mitteilen}\keypar{mitteilen}
\ua{}v.\ue{}
\fa{}I.\fe{} \ua{}etw. (offiziell) mitteilen, verk"unden, Bescheid ge-
ben.\ue{}
das wir mit g\vuv{u}{o}tem willen \.m\,i\,t\, \,t\,a\,i\,l\,e\,n\, wel-
len, dem gotzhaua ze V., diu <\ua{}Anzahl der\ue{}> chint, diu V.N. hat
\ua{}1325 Inderadorf I 58.\ue{}
```

wobei z.B. bei der Zeile `\keymark{mitteilen}\keypar{mitteilen}` der `\keymark`-Befehl dafür sorgt, daß der darauf folgende Eintrag für die Generierung des Kolumnentitels genutzt wird, während der `\keypar`-Befehl den Klammerinhalt mit den Formatierungen für den Lemma-Eintrag versieht. Die Zeichenfolge `\vuv{u}{o}` wiederum bewirkt, daß der Vokal in der zweiten Klammer in kleinerem Schriftgrad über den Vokal in der ersten Klammer gedruckt wird.

Solange ich am Großrechner arbeiten mußte, war ich wegen der Tastaturbelegung an die für WSCRIPT vereinbarten Steuerungszeichen gebunden. Erst als ich Ende 1988 einen IBM PS/2 Personalcomputer<sup>47</sup> mit einer Zusatzkarte für eine Terminalemulation am Großrechner<sup>48</sup> anschaffen konnte, war ich an diese Beschränkungen nicht mehr gebunden und konnte alle auf der PC-Tastatur verfügbaren Zeichen sowie die über die Alt-Taste und den Zahlenblock erreichbaren Zeichen nutzen, um die aus vier Zeichen bestehenden alten Steuerungszeichen durch *ein* – im normalen Text nicht benötigtes – Zeichen zu ersetzen. Ich schrieb daher in KEDIT,<sup>49</sup> das auf dem PC den am Großrechner verwendeten Texteditor ersetzte, ein weiteres Konvertierungsprogramm, das die in der darauffolgenden Verarbeitungsstufe erforderlichen Umwandlungen bewirkte. Die PC-Eingabe für den obigen Text sieht so aus:

```
@mitteilen
<v.> {I.} <etw. (offiziell) mitteilen, verkünden, Bescheid geben.> daz
wir mit guotem willen ^^m^i^t^ ^t^a^i^l^e^n^^ wellen, dem gotzhaus ze
```

45 Ein Konvertierungsprogramm sorgt dafür, daß Zeichen- und Steuerungskonventionen, die unter einem System gelten, in diejenigen überführt werden, die ein anderes System versteht. Letztlich sind es lediglich Befehle, daß ein oder mehrere Zeichen automatisch in andere Zeichen umgewandelt werden sollen.

46 Ist versehentlich die Steuerung für die Kursive nicht ausgeschaltet oder diejenige für die Rechtsstellung falsch verwendet worden, so erfolgt die Fehlermeldung: Falsches Steuerungszeichen in Zeile

47 Dies freilich ist ein Beispiel für eine Entscheidung, die nur anfänglich die beste zu sein schien. Durch die besondere Rechnerarchitektur (Microchannel) benötigte dieses Gerät, wie sich später herausstellte, grundsätzlich ganz andere und grundsätzlich sehr teure Zusatzkomponenten, wollte man das Gleiche wie mit einem normalen PC erreichen. Schließlich mußte das Gerät 1994 ausgesondert werden, nur weil der Graphikchip defekt war. Da dieser auf der Hauptplatine integriert ist, hätte diese ausgewechselt werden müssen, während sonst der Kauf einer neuen Karte für wenige hundert Mark genügt hätte.

48 Diese ermöglichte es mir, den PC auch als nichtintelligentes Terminal für die Verbindung zum Großrechner zu nutzen.

49 Dieses Programm ist ein Texteditor und war bis 1994 über eine Campus-Lizenz vom URZ erhältlich. Vgl. KEDIT User's Guide. Version 5.0. Spring 1992. Mansfield Software Group, Inc., P.O. Box 532, Storrs, CT 06268. Der Übergang von XEDIT am URZ auf KEDIT am PC war relativ einfach, so daß sich nur die Frage nach dem Sinn des Einsatzes eines Textverarbeitungsprogrammes (wie WORD oder WORDPERFECT) statt eines einfachen Texteditors nie stellte. Dies lag freilich auch daran, daß der eigentliche Satz von einem Textsatzsystem übernommen wurde, das auf explizit eingegebene Steuerungszeichen angewiesen war.

V., diu [<Anzahl der>] chint, diu V.N. hat  
<1325 Inderdorf I 58.>

Schon eine rein statistische Aufstellung zeigt, wie sinnvoll eine solche Konvertierung ist, da sie die „händische“ Eingabe von Text erspart. Ein willkürlich ausgewählter Textausschnitt von 1.000 Zeilen umfaßt als T<sub>E</sub>X-Eingabedatei 50.088 Bytes, als WSCRIPT-Datei 44.125 Bytes und als auf dem PC eingebene Datei 41.188 Bytes. Gäbe es diese Konvertierungsmöglichkeit nicht, so mußte zur Erreichung desselben Ergebnisses ein Viertel mehr an Zeichen eingegeben werden.

## 4.2 Die Belegerfassung über die EDV

Als Ende 1987 weitere Terminals angeschafft worden waren,<sup>50</sup> mit deren Hilfe alle Mitarbeiterinnen und Mitarbeiter des DRW ihre Texte am Großrechner eingeben konnten, hatte ich bereits eine nicht unerhebliche Menge von Wortartikeln in eine Textdatei eingegeben. Dies wird die Akzeptanz der Mitarbeiter gegenüber der neuen Technik trotz aller damit verbundenen Ängste gesteigert haben. Denn sie konnten sicher sein, daß ich immer als „troubleshooter“ zur Verfügung stand oder mich zumindest darum kümmerte, daß die Probleme gelöst wurden. Selbstverständlich war diese Akzeptanz aber nicht. Immer wieder werden wir gefragt, wie wir dies geschafft hätten: Die Schwellenangst ist gerade bei der älteren Generation von Wissenschaftlern erheblich. Letztlich war es wohl die Einsicht, daß angesichts des Publikationsdrucks eine solche Technik eine sinnvolle Möglichkeit der Arbeitserleichterung und Beschleunigung darstellte. Die Schulung in der neuen Technik war nur kurz, wichtiger war für uns learning-by-doing.

Dazu gehörte zunächst einmal das Erlernen der Steuerungszeichen, die in einer Liste von etwa zwanzig Zeilen neben jedem Arbeitsplatz lagen. Da die Nichtbeachtung oder irrtümliche Falschschreibung dieser Zeichen unweigerlich zu Fehlermeldungen – manchmal allerdings sehr unverständlicher Art<sup>51</sup> – führte, erzog der Computer uns einerseits zu erhöhter Genauigkeit, war aber andererseits häufig nicht bereit, die Fehlerquelle genau zu nennen. Zeitraubendes Suchen nach dem Fehler war die Folge, und es war manchmal entmutigend, ihn beispielsweise in der Verwechslung von Komma und Semikolon an einer viel früheren Stelle als der der tatsächlichen Fehlermeldung zu finden. Die Texteingabe am Großrechner stellte in diesem Stadium – abgesehen von der komfortablen Weiterverarbeitung mit T<sub>E</sub>X – immer noch nicht viel mehr als einen Schreibmaschinenersatz dar.

Die Gewöhnung an die oben beschriebene Technik der Belegerstellung (auf Einzelzettel) bestimmte zunächst auch die Struktur der Belegaufnahme in der EDV. Die einzelnen Belege wurden ebenfalls separat in eine Textdatei am Großrechner geschrieben, die jeder Mitarbeiter für sich angelegt hatte. Um damit wie bisher arbeiten zu können, schrieb ich ein T<sub>E</sub>X-Makro,

50 Die Entscheidung für Terminals und gegen PC's beruhte allein auf finanziellen Gründen, da Terminals wegen der fehlenden Rechen- und Speicherleistung erheblich preisgünstiger waren. Zu einem späteren Zeitpunkt konnten sie für alle Mitarbeiter durch stand-alone-Geräte, also selbständige Rechner, ersetzt werden, so daß wir nicht mehr auf die Datenübertragung zum Großrechner angewiesen waren.

51 Fehlermeldungen konnten durch das Konvertierungsprogramm erfolgen, aber auch durch T<sub>E</sub>X, wobei die Kryptizität dieser Meldungen allen T<sub>E</sub>X-Benutzern zur Genüge bekannt ist. Ein Beispiel: „T<sub>E</sub>X capacity exceeded, sorry [main memory size=65535]. If you really absolutely need more capacity, you can ask a wizard to enlarge me.“ Die massivste Fehlermeldung allerdings besteht im Systemabsturz, bei dem unter Umständen die gesamte Systemkonfiguration überprüft werden muß, um herausfinden zu können, warum ein Programmablauf, der auf einem Gerät problemlos abgelaufen war, auf einem anderen einen Systemabsturz bewirkte.

das die Druckausgabe auf Din-A4-Seiten, aber mit einem Rahmen im bisherigen Zettelformat, mit dem Stichwort rechts oben und dem Namen des Bearbeiters links oben, ermöglichte. Die Seiten wurden bei uns auseinandergeschnitten und täglich an die Bearbeiter verteilt. Ich erwähne dies, weil es deutlich werden läßt, wie fixiert wir auf die bislang benutzten Techniken und Formalien waren, und wie lange es dauert, bis die Chancen einer solchen Umstellung für eine grundsätzliche Infragestellung der Tradition genutzt werden können. Andererseits aber erleichterte die Bewahrung dieser Tradition sicherlich die Akzeptanz der neuen Technik, gerade weil sie zunächst nur einen Schreibmaschinenersatz darzustellen schien.

Hatten wir die Belege in diesen Dateien mehr oder weniger ungeordnet je nach den Zufälligkeiten des Bearbeitungsmodus eingegeben, so bedurfte es nun noch der korrekten Anordnung der Bedeutungserklärungen je Wortartikel und der chronologischen Anordnung der Belege je Wortbedeutung. Die Lösung dieses Problems hätte eigentlich einer strukturierten Dateneingabe bedurft, dies aber war in einer zeilenorientierten<sup>52</sup> Textdatei nicht möglich. Ein Ersatz wäre – wenn auch sehr kompliziert – mit Hilfe einer Makroprogrammierung denkbar gewesen. In diesem Stadium setzte sich jedoch die konservative Methode der Artikelredaktion durch: Die Belegzettel wurden in der Endredaktion gegebenenfalls im Manuskriptkasten umgelegt; die dort entstandene Reihenfolge bildete ich durch Umkopieren in der Manuskriptdatei ab.

Konnte es bei dieser Nutzung der EDV bleiben oder worin konnte der zusätzliche Nutzen der Datenverarbeitung für die Wörterbucharbeit liegen?

#### 4.3 Der Aufbau einer Quellensigle-Datei

Der oben beschriebene Zustand der Archivzettel erschwerte die Arbeit immer wieder, immer wieder wurden die gleichen umständlichen Nachschlagehandlungen in der Quellenkartei<sup>53</sup> nötig, mußten Änderungen dieser Karteieinträge jedem mitgeteilt werden. Der Gedanke, daß die Verifizierung einer Quelle – ein erhebliches Problem in einem über einen Zeitraum von sieben Jahrzehnten gewachsenen Zettelarchiv – mit Hilfe der EDV erheblich einfacher gestaltet werden könne als mit dem Nachschlagemedium Quellenkatalog, lag nahe.

Mit dem Vorliegen maschinenlesbarer Wörterbuchtexte wurde es möglich, die darin vorkommenden Quellensiglen und Datierungen herauszufiltern. Der Texteditor KEDIT bietet eine Makroprogrammierung an, die an die Befehlssprache REXX<sup>54</sup> angelehnt ist, und mit deren Hilfe ich Schriftauszeichnungen und andere textinterne Kriterien nutzen konnte, um Datierungen und Quellensiglen vom übrigen Wörterbuchtext zu separieren.<sup>55</sup> Bereits die ersten Ergebnisse ließen deutlich werden, welche Inkonsistenzen in der Zitierweise es trotz aller Korrekturvorgänge gab. Das Bedürfnis, die gültigen Quellensiglen mit den Datierungen, falls eindeutig festlegbar, und den verbindlichen Zitierweisen in einer Datei verfügbar zu haben, schien mir auf Grund dieser Erfahrung so unabdingbar zu sein, daß ich die aus den Manuskripten herausgezogenen Quellensiglen „händisch“ komplettieren ließ. In dieser Datei stehen alle Quellenabkür-

52 Alle Befehle konnten sich nur auf eine oder mehrere Zeilen und deren Inhalt beziehen, nicht aber auf einen Textblock, der beispielsweise durch das Lemmazeichen am Anfang und ein bestimmtes Endezeichen als Wortartikel strukturiert erfaßt worden wäre.

53 Dieser Katalog enthält im wesentlichen alle notwendigen Angaben zu den einzelnen Quellenangaben, ist aber als Kartei umständlich zu benutzen. In einem früheren Stadium der Bemühung um Rationalisierung hatten wir die Karteikarten verfilmen und kopieren lassen, so daß sie in Buchform an jedem Arbeitsplatz verfügbar waren. Dies erforderte aber, daß die Aktualisierung ebenfalls an jedem Arbeitsplatz – und zwar in Ermangelung von Hilfskräften von jedem Wissenschaftler selber – vorgenommen werden mußte.

54 Hierzu vgl. COWLISHAW 1988 und KEES 1993.

55 Dies stellt die ersten Anfänge eines Wörterbuchparsings dar; vgl. unten 7.1.

zungen, dann aber auch eine Vielzahl von Verweisen von den vorläufigen auf die endgültigen Quellenabkürzungen sowie von Reihenabkürzungen auf Einzelabkürzungen und bei Zeitschriftenbänden teilweise die Band- und Jahrgangsangaben. Ferner ist bei den Abkürzungen zum Teil bereits eine Datierung und der Standort des Quellenwerkes (eigener Buchbestand, Dauerleihgabe der Universitätsbibliothek Heidelberg, Fernleihe mit Angabe der Bibliothekssigle) angegeben. Damit kann einerseits der Zugriff auf das Quellenwerk beschleunigt werden, andererseits aber können die bislang nicht zu vermeidenden unterschiedlichen Schreib- und Zitierweisen bei den Quellenabkürzungen ausgeschlossen werden. Diese Fassung des Quellenkatalogs wurde zunächst auf dem Großrechner zentral, dann aber nach dem Übergang auf stand-alone-Geräte auf jedem EDV-Arbeitsplatz bereitgestellt und automatisch mit dem Laden der jeweiligen Manuskriptdatei mitgeladen, so daß bei der Bearbeitung des Archivmaterials die Quellensigle aus dieser Datei in die Artikel- bzw. Belegdatei hineinkopiert werden konnte.<sup>56</sup>

#### 4.4 Von der Belegdatei zur Manuskriptdatei

Immer noch wurden die Artikelteile in relativ beliebiger Reihenfolge in eine Datei geschrieben und erst im Archivkasten zum Wortartikel zusammengestellt. Dies änderte sich erst 1990, als alle Mitarbeiter mit Personalcomputern anstelle der nichtintelligenten Terminals ausgestattet waren. Die Einschränkungen der Arbeit durch die Langsamkeit der Datenübertragung an das URZ wie auch durch die notwendigen Wartungs- und Umstellungsarbeiten fielen nun fort. Wichtiger war freilich, daß die Texteingabe erleichtert werden konnte.

Der Vorschlag, statt der atomistischen Eingabe von Bedeutungserklärungen und Belegtexten eine fortlaufende Manuskriptdatei zu schreiben, kam dabei von der Seite der Mitarbeiter. Sie gewannen dadurch wieder den etwas in den Hintergrund gedrängten Überblick über das Gesamtmanuskript, das ihnen wie auf Schreibmaschinenseiten zeilenweise zur Verfügung stand. Zugleich war damit aber die Möglichkeit gegeben, den Manuskripttext jederzeit ausdrucken zu können, sei es in einem einspaltigen Korrekturformat mit größeren Schriften – wegen der besseren Lesbarkeit – oder im zweisepaltigen Wörterbuchformat. War der Ausdruck gelungen, so war nunmehr während der Bearbeitungszeit selbst eine Kontrolle über den für die Alphabetstrecke bereits verbrauchten Platz möglich geworden. Eine Überschreitung des Raumes um einen gewissen Prozentsatz konnte und mußte toleriert werden, da Kürzungen, Streichungen von Belegen und teilweise auch von Wortartikeln während der Endredaktion den Gesamtumfang sowieso verringerten. Die Übersicht über den verbrauchten Platz sowie die Übersicht über den verflossenen Zeitraum für die Strecke trug zu einer – in jedem Wörterbuchunternehmen erforderlichen – Selbstdisziplinierung bei, die es vorher nicht in diesem Maße geben konnte.

## 5 Die weitere Nutzung der Dateien

### 5.1 Recherchemöglichkeiten als Hilfsmittel der Wörterbucharbeit

Der Umgang mit Textdateien und die neuen Recherchemöglichkeiten in diesem Medium werden jeden, der es mit Textverarbeitung in größerem Stil zu tun hat, zu einer Reflexion der eige-

<sup>56</sup> Nur am Rande soll erwähnt werden, daß wir zusätzliche Angaben zu Quellenwerken und Sachhinweise unterschiedlichster Art über ein Programm ASKSAM verwaltet hatten. Als wir wenig später alles auf FAUST umstellten, konnten wir die in ASKSAM bereits enthaltenen Angaben in eine ASCII-Datei exportieren und in FAUST importieren, so daß die darauf verwendete Mühe nicht umsonst war. ASKSAM ist ein Textretrievalprogramm.



nen Tätigkeit und der Überlegung zwingen, wie diese Möglichkeiten in den normalen Arbeitsablauf eingebaut werden können. Für die Wörterbucharbeit im DRW bedeutete dies neben der Vereinheitlichung unterschiedlicher Zitierweisen im Manuskript auch die Erschließung der maschinenlesbaren Wörterbuchtexte für die zukünftigen Wortartikel. Ein Wörterbuch ist eine Textsorte, die erheblich mehr Informationen zur Verfügung stellen kann, wenn der Zugriff auf die Information nicht an die alphabetische Reihenfolge der Lemmaeinträge gebunden ist.<sup>57</sup> Mit anderen Worten: Der alphabetische Nachschlagevorgang erfordert eine Verweisstruktur innerhalb des Wörterbuchs, wenn über diesen einen Lemmaeintrag hinaus weitere Informationen aus anderen Wortartikeln herangezogen werden sollen. Der Verweis auf Wortbildungsmuster sowie die alphabetische Reihung der Kompositen nach dem Bestimmungswort stellen hierbei nur einen kleinen Teil der Möglichkeiten dar. Synonymen- und Antonymenverweise decken ebenfalls nicht die Breite des möglichen Spektrums ab. Im DRW – als einem Wortwörterbuch mit enzyklopädischen Funktionen – spielen daneben auch Sachverweise eine Rolle, etwa von dem Typ, daß unter „Kauf“ auf verwandte Rechtsinstitute wie „Tausch“, „Leihe“, „Miete“ oder „Pacht“ verwiesen werden kann.<sup>58</sup>

Eine solche Verweisstruktur muß aus systemimmanenten Gründen immer unvollständig bleiben, denn sie setzt voraus, daß entsprechend dem Weltwissen des Lexikographen Markierungen in ausreichender Anzahl gesetzt worden sind.<sup>59</sup> Etwas anders ist es mit maschinenlesbaren Texten. Da sie in der Form eines Stromes von Einzelzeichen vorliegen, läßt sich eine Suche nach jeder beliebigen Kette von Zeichen vornehmen; die Möglichkeit, durch Jokerzeichen (Wildcards)<sup>60</sup> Zeichenketten mit variierenden Teilen herauszufiltern oder Links- oder Rechtstrunkierungen<sup>61</sup> vorzunehmen, erweitert diese Suchmöglichkeiten erheblich. Dadurch unterliegen die Recherchen in maschinenlesbaren Texten in der Regel nur denjenigen Beschränkungen, die in der Person des Recherchierenden liegen, etwa in der Notwendigkeit, Fragen formulieren zu können. So müssen, um aus dem gesamten Material sämtliche möglichen Wortformen zu einem bestimmten Lemma herausfinden zu können, die verschiedenen Graphien bekannt sein – zumindest annäherungsweise.

## 5.2 Vom einfachen Suchbefehl zum Volltext-Retrieval-Programm

Die Folge dieser Suchmöglichkeiten war es, daß zunächst mehr Belegmaterial zur Verfügung stand und die Verweise vom Typ „Beleg siehe unter“ sich häuften. Es war freilich relativ unkomfortabel, in KEDIT einen normalen „Locate“-Befehl für eine solche Recherche zu nutzen, da nur die aktuelle Datei<sup>62</sup> den Suchbereich darstellte. Abhilfe schuf ein Volltext-Retrieval-Programm<sup>63</sup> mit dem sprechenden Namen „WordCruncher“,<sup>64</sup> das es ermöglichte, Texte mit

57 Grundlegend dazu GOEBEL/LEMBERG/REICHMANN 1995.

58 Im äußersten Fall könnte dies in die Richtung der Frame-Theorie zielen, die WEGNER 1989, 893ff., behandelt hat.

59 Vgl. dazu unten 6.5.2.

60 Das sind Zeichen, die ein oder mehrere Zeichen vertreten, so daß bei einem Suchvorgang alle Zeichenketten gezeigt werden, die an der Stelle des Jokerzeichens einen beliebigen Buchstaben haben: *m?t* als Suchstring ergibt unter anderem *mit*, *mut*, *mrt* usw.

61 Z.B. um Wortbildungsmuster herauszufinden: Bei der Linkstrunkierung wird etwa jede Zeichenkette gezeigt, die mit *heit* endet, bei der Rechtstrunkierung jede Kette, die mit *Mier* beginnt.

62 Deren Umfang wegen des beschränkten Speicherplatzes im Arbeitsspeicher nur einen jeweils geringen Teil des Gesamtmaterials darstellen konnte.

63 Im Informatik-Duden (DUDEN 1988) taucht „Retrieval-“ weder als Stichwort noch im Sachregister auf. Ein solches Programm ermöglicht die Recherche nach beliebigen Zeichenketten, wobei die verschiedenen Programme sich danach unterscheiden, ob die Indexierung sofort bei der Importierung der Texte vorge-

Markierungen für bestimmte Bereiche (etwa Seiten oder Abschnitte) zu versehen und sie einmal indexieren zu lassen, so daß die Recherche über eine alphabetische Zeichenkettenliste erfolgen konnte. Die Sortierreihenfolge war beliebig wählbar, so daß auch die Einsortierung der Umlaute, der Sonderzeichen sowie die Ignorierung bestimmter sonstiger Zeichen, die im Wortinneren vorkommen konnten (z.B. der Bindestrich) möglich war. Auf die Erstellung einer Stopwortliste, durch die bestimmte Zeichenketten aus der Indexierung herausgenommen werden, verzichtete ich, da jede Zeichenkette in einem Wörterbuch einmal recherchefähig sein kann, insbesondere zur Eingrenzung eines Rechercheergebnisses. Inzwischen stehen rund 35 Megabyte maschinenlesbarer Texte<sup>65</sup> über den WordCruncher<sup>66</sup> zu Recherchen zur Verfügung. Allein ein Drittel davon sind Wortartikel des DRW,<sup>67</sup> die entweder gleich maschinenlesbar hergestellt oder aber maschinell eingelesen wurden.

### 5.3 Das Problem der Inzuchtbelege

Die Ausnutzung der Textdateien für die Beleggewinnung bringt ein Problem mit sich, das in konventionell genutzten Wörterbüchern nur selten eine Rolle spielt: Die sogenannten Inzuchtbelege,<sup>68</sup> bei denen aus bereits vorhandenen Belegtexten weitere Belege für andere Lemmata gewonnen werden können. In der Lexikographie der Standardsprache werden Inzuchtbelege nicht gern gesehen, da die Gefahr besteht, daß bestimmte Belege und damit Textsorten immer wieder zitiert werden und die Belegverteilung damit nicht mehr repräsentativ für die Allgemeinsprache ist. Im DRW ist dies wegen seines Objektbereichs – der Fachsprache des Rechts und des darauf bezogenen Teils der Allgemeinsprache – grundsätzlich anders. Da – jedenfalls der Absicht nach – immer nur Rechtswörter behandelt werden, stellt jeder Text, der ein Rechtswort enthält, auch eine mögliche Quelle für weitere Rechtswörter dar. Je höher die rechtliche Relevanz der Quelle und des konkreten Belegausschnittes ist, desto mehr Rechtswörter werden sich in diesem Belegtext häufen. Eine Wiederverwendung gerade dieses Beleges ist legitim, weil jedes Wort rechtliche Relevanz besitzen kann.<sup>69</sup> Die Zufälligkeit der alpha-

---

nommen wird, was die Suche erheblich beschleunigt, oder ob der Text für jeden Suchvorgang gesondert durchsucht wird. Die indexierten Zeichenketten können alphabetisch oder nach individuell wählbaren Sortierkriterien geordnet werden. Durch Antippen der Buchstabenfolge im Index nähert man sich immer mehr dem gesuchten Wort und bekommt es schließlich in dem bekannten KWIC-Format (Keyword in Context) mit einem meist wählbar großen Textstück gezeigt. Diesen Text kann man bei Bedarf markieren und in eine Datei zur beliebigen Weiterverarbeitung kopieren.

64 Das Programm wurde an der Brigham Young University entwickelt. Der Vertrieb erfolgte von der Electronic Text Corporation, jetzt von Johnston & Company, P.O. Box 446, Amerikan Fork, UT 84003, und in Deutschland von der Firma Incom GmbH., Herzogsfreudenweg 16, 53125 Bonn. Wir verwenden die Version 4.5, in deren deutscher Fassung allerdings die unverzichtbare Möglichkeit, erweiterte Fundstellenangaben ausgeben zu lassen, nicht implementiert ist. Die vom NIEMEYER Verlag geplante, aber inzwischen auf die Hamburger GOETHE-Ausgabe beschränkte Elektronische Textbibliothek wurde mit der „View“-Funktion dieses Programmes, die nur die Auswertung bereits indexierter Texte ermöglicht, ausgeliefert.

65 Die Größe der Quelldateien ist nur während der Indexierung beschränkt; einmal indexierte Dateien können danach zu einer Datei zusammengeführt werden, deren Größe nur durch den verfügbaren Platz auf dem Massenspeicher begrenzt wird.

66 Zur Weiterentwicklung der Volltextrecherche in FAUST vgl. unten 6.5.3.

67 Es sind ca. 14.100 Wortartikel von Band VI Heft 9/10 bis Band IX Heft 3/4, die 209.070 verschiedene Zeichenketten und rund 12 Megabyte umfassen.

68 Dieser ebenso sprechende wie unkonventionelle Begriff stammt aus dem Umkreis des FWB.

69 In der Definition des Diebstahls in § 242 des deutschen Strafgesetzbuchs ist nahezu jedes Wort als Rechtswort interpretierungsfähig und -bedürftig. Vgl. dazu auch SPEER 1989, 117ff.

betischen Reihenfolge darf nicht darüber entscheiden, ob ein Belegtext, der bereits für ein Wort am Beginn des Alphabets verwendet wurde, für die weiteren Wortartikel als Fundstelle ausscheidet.<sup>70</sup>

#### 5.4 Erweiterung der Recherchegrundlagen durch OPTOPUS

Die eigenen Texte des DRW boten vielfache Nachschlagemöglichkeiten. Sie liessen aber auch deutlich werden, daß die maschinelle Recherche jeglicher manueller Exzerption – mit gewissen Einschränkungen – weit überlegen ist. Und damit entstand die Frage, wieweit es möglich wäre, die wichtigsten Quellen des DRW maschinell zu nutzen. Ende der Achtzigerjahre war in der Rechtsgeschichte, die zunächst als Textlieferant hätte in Frage kommen können, noch kaum ein Gedanke an die freie Verfügbarkeit von Textsatzdateien – denn um diese mußte es zunächst gehen. So blieb als einziger Weg die interne Herstellung maschinenlesbarer Quellentexte. Hierfür kam zunächst die KDEM<sup>71</sup> in Frage, die ich nur vom Hörensagen her kannte, deren Preis aber damals noch in die Hunderttausende gehen sollte. In Heidelberg gab es kein derartiges Gerät, so daß auch die Mitnutzung eines Fremdgerätes kaum in Frage kam. Bei einer Firmenpräsentation, zu der das URZ eingeladen hatte, lernte ich OPTOPUS kennen, ein frei trainierbares Programm zur Schriftenerkennung.<sup>72</sup> Es gibt zwei sehr unterschiedliche Verfahren bei dem sogenannten OCR (Optical Character Recognition): Einmal sind verschiedene Schriftarten mit ihren Algorithmen im Programm fest eingebaut, andere Schriften lassen sich höchstens bei einer gewissen Ähnlichkeit mit den programmierten Schriften erkennen. Dafür benötigen diese Programme kein Training. Für Vorhaben mit den heutigen Schreibmaschinen- und den üblichen Druckschriften ist dies sicherlich ein gutes Verfahren. Hat man es mit historischen Drucken oder mit nichtlateinischen Schriften zu tun, so kommt nur das zweite Verfahren in Betracht: Die Trainierbarkeit des Programms. Das OCR-Programm zeigt jedes Zeichen, das es als selbständig<sup>73</sup> erkannt hat, auf dem Bildschirm und erwartet die Zuordnung dieses graphischen Zeichens zu einem oder mehreren<sup>74</sup> ASCII-Zeichen. In der Praxis bedurfte es einer gewissen Erfahrung, um zu erkennen, wann beispielsweise Drucktypen, die durch zuviel Druckerschwärze oder abgenutzte Drucktypen als zusammengesetzte oder als zerbrochene Buchstaben erschienen, gesondert trainiert werden sollten.

Da der Preis einer OPTOPUS-Arbeitsstation nur etwa ein Zehntel der KDEM betrug, konnte das Gerät – für alle Forschungsstellen der Heidelberger Akademie – angeschafft werden und fand seinen Platz in den Räumen des DRW. Nach den üblichen Anfangsschwierigkeiten konnte mit dem Scannen<sup>75</sup> begonnen werden, zunächst durch mich, im weiteren Verlauf durch studentische Hilfskräfte. So wurden bisher 31 Lieferungen des DRW maschinenlesbar gemacht, von denen freilich noch nicht alle korrigiert sind. Denn es ist ein Irrtum, wenn man als Ergebnis

70 Im DRW gibt es zwei Möglichkeiten, solche Belege zu verwerten: Einmal ohne Hinweis auf einen bereits gedruckten Wortartikel, dann aber auch als bloßen Verweis vom Typ *Beleg s. unter...* Zur Behandlung dieser letzteren Belegform seit 1994 siehe unten 6.5.3.

71 Kurzweil Data Entry Machine, heute in dem Modell Kurzweil K5200.

72 Entwickelt und vertrieben von der Firma MAKROLOG in Wiesbaden, inzwischen in einer Version für WINDOWS verfügbar.

73 Dh. durch einen Leerraum vom vorigen und nächsten Zeichen getrennt – was selbstverständlich klingt, aber bei kursiven Schriften und Ligaturen bereits Schwierigkeiten verursachen kann.

74 Beispielsweise bei Ligaturen oder bei Sonderzeichen, die im ASCII-Zeichensatz nicht enthalten sind und daher kodiert eingegeben werden müssen.

75 Dieser Gebrauch des Wortes als Erzeugen der Graphikdatei einer Buchseite und anschließendes maschinelles Umsetzen in eine ASCII-Datei scheint sich eingebürgert zu haben, obwohl eigentlich nur das Erstellere damit gemeint war.

eines solchen Einlesens eine fehlerfreie Datei erwartet. Bei einer Erkennungsquote von 99,7 oder noch mehr Prozent, die das Gerät nach genügendem Training meldet, sind auf tausend Zeichen immer noch drei nicht erkannt worden. Ganz zu schweigen von den Zeichen, die das Gerät irrtümlich falsch zugeordnet hat. Eine Korrekturmöglichkeit besteht darin, Bedingungen zu formulieren, unter denen bestimmte Zeichenkombinationen möglich oder nicht möglich sind. Gelingt dies, so können einige Fehler automatisch durch Nachbearbeitung beseitigt werden. Das entsprechende KEDIT-Makro, das ich hierfür geschrieben habe, umfaßt zusammen mit dem optopuseigenen, aber für das DRW modifizierten Nachbearbeitungsprogramm TREN-NEX etwa vierhundert Change-Befehle, die beispielsweise dafür sorgen, daß in einer Zeichenkette, die aus Zahlen besteht, ein „l“ oder ein „i“ durch die „1“ oder ein „O“ durch eine „0“ ersetzt wird. Besonders wichtig ist hierbei die Reihenfolge der Befehle, denn unter Umständen werden durch eine falsche Reihenfolge mehr neue Fehler produziert als alte bereinigt.

Eine gänzlich andere Möglichkeit besteht darin, den Wörterbuchtext am Computer abschreiben zu lassen. Die elektronische Fassung des OXFORD ENGLISH DICTIONARY ist so hergestellt worden und das Jahrhundertwerk des THESAURUS LINGVAE LATINAE wird den gleichen Weg gehen. Ich lasse gegenwärtig durch eine Firma, die ein entsprechendes Vorhaben für den LANGENSCHIEDT-Verlag durchgeführt hat, die Wirtschaftlichkeit dieses Verfahrens für das DRW prüfen. Letztlich ist die Grundlage für ein solches Angebot das Lohngefälle zwischen Europa und den Niedriglohnländern, da die Abschrift in Südostasien hergestellt wird. Allerdings steht die fiskalische Haushaltspolitik einem solchen Vorhaben meist entgegen: Die Maschinenlesbarkeit kann innerhalb der Forschungsstelle durch die vorhandenen Arbeitskräfte schenbar kostenneutral durchgeführt werden, während die Verlagerung dieses Vorhabens nach außen gesondert finanziert werden muß. Es unterliegt kaum einem Zweifel, daß eine Zusatzfinanzierung durch die öffentliche Hand im gegenwärtigen Zeitpunkt wohl nicht erreichbar ist.

### 5.5 Corpuuserweiterung durch Exzerption?

Die Maschinenlesbarkeit einiger Teile des DRW und einiger altfriesischer Rechtsquellen, die wir mit Erlaubnis des Verlages VANDENHOECK & RUPRECHT in Göttingen „gescannt“ hatten,<sup>76</sup> war natürlich nicht Selbstzweck. Und auch andere Dateien kamen im Lauf der Zeit hinzu: Das gesamte altenglische Textmaterial, das mit ca. 40 Megabyte die Grundlage für das neue altenglische Wörterbuch ist, wurde uns zum Preis für den Datenträger zur Verfügung gestellt. In Linz wurden einige nur handschriftlich vorliegende Rechtsquellen des 16. und 17. Jahrhunderts maschinell eingegeben, aus Duisburg und Kleve<sup>77</sup> erhielten wir Stadtrechtsquellen, und vom Institut für Bayerische Literaturgeschichte in München bekamen wir mittelhochdeutsche Texte, deren Ergiebigkeit für die Rechtssprachgeschichte allerdings nicht sehr hoch ist.<sup>78</sup> Das FWB stellte uns Artikeldateien zur Verfügung, die wir in den WordCruncher eingaben. Und in den letzten Tagen erklärte der NIEMEYER-Verlag freundlicherweise seine Bereitschaft,<sup>79</sup> dem DRW

76 BUMA/EBEL 1963ff.

77 Da dies nicht selbstverständlich ist, sei hier Herrn Prof. MIHM aus Duisburg (vgl. MIHM/ELMENTHALER 1990) sowie seiner Frau (vgl. MIHM 1994) und Herrn Prof. FLINK vom Stadtarchiv Kleve (vgl. FLINK 1991) ausdrücklich gedankt. Gerade das Klever Stadtrecht war im DRW nur über handschriftliche Exzerpte von RICHARD SCHRÖDER repräsentiert, der diese Quelle im dortigen Archiv ausgewertet hatte. Sie war bis jetzt unpubliziert.

78 Dennoch lohnt sich die maschinelle Verfügbarkeit, um die Wichtigkeit oder gerade auch die mangelnde rechtliche Relevanz der in den gängigen Wörterbüchern meist nur als Fundstellen repräsentierten Quellen für einen Rechtswortartikel schnell und ohne Gang an das Bücherregal feststellen zu können.

79 Dies scheint auch eine zunehmende Flexibilität im Umgang mit diesen Datenträgern anzuzeigen: Häufig

eine Textdatei der wichtigen Beichtsumme des Bruders Berthold<sup>80</sup> zu überlassen. Der Zugriff auf jede einzelne Zeichenkette dieser Texte war damit möglich. War dies der Ersatz der 1974 eingestellten Exzerption? Laßt sich das Quellencorpus des DRW nicht in einem einmaligen Kraftakt maschinenlesbar machen, so daß die Bearbeitungszeit entscheidend verringert werden kann?

Hier muß ich nochmals auf die Struktur des Quellenmaterials, aus dem die Belegauschnitte hergestellt werden, eingehen. Sie unterscheidet sich zumindest in zwei Punkten grundlegend von derjenigen in anderen Wörterbüchern. Einmal ist es die Vielzahl der ausgewerteten Quellen mit ca. 7.800 Quellensiglen, zum anderen aber die Exzerptionsdichte, die gerade für Zeitschriftenbände, aus denen jeweils nur einige Seiten einschlägige Textsorten enthielten, sehr niedrig war. Einige tausend Texte aus einem (Druck-)Zeitraum von 500 Jahren lassen sich nicht durch *ein* Wörterbuchunternehmen maschinenlesbar machen. Es bedürfte dazu eines nationalen Instituts, das dies in großem Umfang planen und auch als Clearingstelle<sup>81</sup> für derartige Dateien dienen könnte. Das DRW kann sich nicht nur wegen der bereits verstrichenen Bearbeitungszeit, sondern aus diesen grundlegenden Vorbedingungen seiner Quellenbasis heraus nicht auf ein Quellencorpus stützen, das gegebenenfalls maschinenlesbar gemacht werden könnte.

Das Gesamtcorpus des DRW wird also niemals maschinenlesbar sein. Die Textdateien, die ich oben genannt habe und deren Zahl sich auch in Zukunft sicher vermehren wird, werden immer nur zusätzliches Material zum Zettelarchiv liefern können. Dies entkräftet auch die grundsätzlichen Bedenken, die REICHMANN gegenüber diesen Möglichkeiten anmeldet.<sup>82</sup> Überdies: Die maschinelle Erfassung von Rechtstexten führt zu einer Belegvermehrung, die wegen der Struktur der Rechtssprache und der Rechtstexte in sich sinnvoll ist. Angesichts der häufig definitionsartigen<sup>83</sup> Verwendung wichtiger Rechtswörter in derartigen Texten kann ein schneller Überblick über solche Belege das Fachwissen des Bearbeiters vermehren und ihm Kriterien für die weitere Belegauswahl an die Hand geben. Es hat sich bereits gezeigt, daß diese Möglichkeiten vor allem dazu dienen können, Exzerptionlücken zu schließen.<sup>84</sup>

---

überwogen bisher bei den Verlagen die urheberrechtlichen Bedenken gegenüber der Überlassung maschinenlesbarer Texte an wissenschaftliche Forschungsunternehmungen, auch wenn diese die Nutzung ausschließlich zu wissenschaftlichen Zwecken zusagten.

80 Vgl. STEER 1987ff.

81 So daß jeder Wissenschaftler, der einen Text bearbeiten will, dort erst einmal nachfragen kann, ob er nicht bereits maschinenlesbar ist.

82 REICHMANN 1990 a, 1601f.: „Der zweite Konstellationstyp [bei dem lexikographische Bemühungen um Qualitätssteigerung sich auf eine quantitative Vermehrung des Materials stützen] ist dann gegeben, wenn die quantitativen Möglichkeiten, die die moderne Technik bereitstellt, auf die Leiter und Mitarbeiter eines lexikographischen Unternehmens eine solche Faszination ausüben, daß man der Maschine die Erleichterung des schlechten methodischen Gewissens anvertraut, das oft durch eine freilich immer als unzureichend beurteilbare Quellengrundlage hervorgerufen wird. Sie kann auch dann gegeben sein, wenn sich Nicht-Lexikographen infolge der Beherrschung einiger technischer Mittel plötzlich im Vorteil gegenüber den Fachvertretern sehen und sich auf den Weg begeben, eine schnelle lexikographische Mark zu verdienen...“.

83 Vgl. dazu EBEL 1974.

84 Vgl. dazu 6.5.3.

## 6 Von der zeilenorientierten Textverarbeitung zur strukturierten Dateneingabe

### 6.1 Die Anfänge

Bei der Arbeit mit den verschiedenen Programmen, die sich im Lauf dieser Zeit zu einem durchaus unterschiedlichen Konglomerat von Hilfsmitteln zusammengefügt hatten, kam immer stärker der Wunsch auf, alle diese Möglichkeiten in *einem* System zu verbinden. Denn in diesem Stadium mußte aus Manuskriptdateien heraus immer wieder auf andere Dateien oder andere Programme zurückgegriffen werden, um alle maschinenlesbar verfügbaren Informationen zu erhalten. Der Nachschlagevorgang in dem jeweiligen Programm erforderte zwar immer nur kurze Zeit und war erheblich schneller als auf dem konventionellen Weg beispielsweise in der Quellenkartei. Aber bei einer Vielzahl aufeinanderfolgender Recherchen in verschiedenen Dateien summierte sich allein schon die Zeit, die zum Laden und dann wieder zum Verlassen der jeweiligen Programme benötigt wurde. Zwar war es mir nach längerem Bemühen möglich, bestimmte Programme so in das Textverarbeitungsprogramm einzubauen, daß sie durch eine Tastenkombination gestartet und in ihnen eine Recherche durchgeführt werden konnte; das Ergebnis wurde schließlich in eine ASCII-Datei geschrieben, die sofort auf dem Bildschirm zur Übernahme in die Manuskriptdatei angezeigt wurde. Es war eine Konstruktion, die historisch gewachsen und nicht systematisch entwickelt worden war, die aber viele Bedürfnisse der lexikographischen Alltagspraxis befriedigen konnte. Dennoch: Die ständige Arbeit mit der neuen Technik schärfte den Blick für weitere Einsatzmöglichkeiten. Teilweise war es auch ein ästhetisches Bedürfnis, diese Lösungen eleganter und komfortabler zu gestalten, so daß jeder Schritt weitere nach sich zog. Dies ging so weit, daß ich begann, die Programmiersprache „C“ zu lernen – ein Unterfangen, das ich allerdings nach einigen Wochen aus Zeitmangel aufgeben mußte.

Die Weiterentwicklung einer Systemlösung im DRW wurde aber auch durch die Fragen und Wünsche gerade *derjenigen* meiner Mitarbeiterinnen gefördert, die eine ungeprüft optimistische Vorstellung von den Möglichkeiten der elektronischen Datenverarbeitung besaßen. Seinerzeit war es die Frage des Verlags nach der Brauchbarkeit der Manuskriptdatei als Vorlage für den photomechanischen Satz gewesen, die den Übergang auf TEX motiviert hatte. Nun wurde ich mit dem Wunsch nach Funktionen der EDV im DRW konfrontiert, die ich im Wissen um die damit verbundenen Schwierigkeiten zunächst von mir wies. Aber die Anregungen wirkten weiter und forderten mich letztlich auch heraus.

Ein grundlegendes Dilemma allerdings blieb trotz aller Bemühungen ungelöst: War es erforderlich, jeden Mitarbeiter derart mit den Funktionen des Betriebssystems vertraut zu machen, daß er sich selbständig in diesem System bewegen konnte? Oder erhöhte es nicht vielmehr umgekehrt die Akzeptanz gegenüber der Datenverarbeitung, wenn möglichst wenig Hintergrundwissen benötigt wurde? Ich hatte mich zu dem letzteren Weg entschlossen und mit Hilfe des bekannten Programmpaketes NORTON UTILITIES ein Menü entworfen, das es den Mitarbeitern ermöglichte, alle erforderlichen Programme ablaufen zu lassen, ohne einen DOS-Befehl oder einen direkten Startbefehl für ein Programm eingeben zu müssen.<sup>85</sup> Heute bin ich der Meinung, daß keine dieser beiden Lösungen optimal ist: Ohne Kenntnisse des Betriebssystems oder der jeweiligen Programmstruktur wird man immer wieder vor Problemen stehen, die man alleine nicht lösen kann. Andererseits aber ist es kaum zu rechtfertigen, wenn jede Per-

---

85 In jedem Menü-Bildschirm gab es verschiedene Wahlmöglichkeiten, die durch Eingeben der davorstehenden Ziffer gewählt wurden und gegebenenfalls in das nächste Fenster mit weiteren Wahlmöglichkeiten weiterleiteten.

son den muhsamen Weg zur Erlangung derartiger Kenntnisse *neben* der wissenschaftlichen Arbeit gehen muß<sup>86</sup>

So kam es schließlich zum Übergang von der menügesteuerten Nutzung verschiedener Programme zur Nutzung einer Datenbank.<sup>87</sup> Erste Anfänge hierzu reichten noch in die Zeit zurück, in der wir am Großrechner arbeiteten. Zunächst versuchte ich, mit Hilfe eines Mitarbeiters des URZ auf dem Großrechner das Datenbankprogramm ORACLE<sup>88</sup> für die lexikographische Arbeit nutzbar zu machen. Voraussetzung hierfür schien es damals zu sein, sich ein für allemal an die Strukturen des Großrechners zu binden.<sup>89</sup> Schwerwiegender war freilich, daß ORACLE als relationale Datenbank nicht flexibel genug war. Die Schwierigkeiten bei diesem Versuch häuften sich. Bestimmte Lösungen, die hätten möglich sein müssen, ließen sich nicht verwirklichen, weil ein Modul des Programmes dem URZ nicht zur Verfügung stand und erst aus den USA besorgt werden mußte. Als es schließlich ankam, paßte es nicht zu der Programmversion, die das URZ benutzte. Meine Geduld wurde auf eine so harte Probe gestellt, daß ich diesen Versuch für gescheitert erklärte und mich darauf einstellte, für die weitere Zukunft mit der Kombination verschiedenster Programme im DRW zurechtzukommen zu müssen.

Der Verzicht auf eine solche Lösung wäre allerdings für die Arbeit an unserem Wörterbuch problematisch geworden. Der stand-alone-Betrieb von PC's führt in einem Unternehmen, in dem alle letztlich an der gleichen Aufgabe arbeiten, zu gravierenden Problemen bei der Konsistenz der Daten. Es ist in dieser Zeit mehrfach vorgekommen, daß wichtige Veränderungen für dieselbe Alphabetstrecke an Dateien in jeweils unterschiedlichem Bearbeitungsstand vorgenommen wurden, so daß die Herstellung eines endgültigen Textes überaus schwierig wurde. Die Lösung konnte nur in einer Datenbank gefunden werden, in der jedes Objekt nur einmal vorhanden ist und daher nicht gleichzeitig von zwei Personen verändert werden kann. Dies wiederum setzte voraus, daß wir von der Einzelplatzlösung abgehen und ein Netzwerk<sup>90</sup> einrichten mußten. Bei der Planung und später Einrichtung des Netzwerks half mir HEINZ-DETLEV KOCH, ein Student der Computerlinguistik, der mich davor bewahrte, alle Einzelheiten des NOVELL Netware-Systems verstehen zu müssen. Auch heute komme ich mit einem kleinen Teil der eigentlich für einen Supervisor<sup>91</sup> erforderlichen Kenntnisse aus.<sup>92</sup> Als die Finanzie-

86 Ideal wäre in einer Institution wie der Heidelberger Akademie der Wissenschaften mit ihren über 40 Forschungsstellen die Einrichtung einer Stelle für einen Informatiker, der die wissenschaftlichen Mitarbeiter von den täglich auftretenden Problemen mit Hard- und Software entlasten könnte. Die Lösung solcher Probleme durch Wissenschaftler, die sich erst mit dem Gegenstand vertraut machen müssen, ist ebenso unwirtschaftlich wie die Heranziehung von Mitarbeitern der EDV-Firmen, deren hoher Stundenlohn keineswegs für eine ausreichende Kenntnis der Systemkonstellation und damit der Problemlösungsmöglichkeiten garantiert!

87 Ich verwende diesen Begriff unspezifisch für ein Programm, das eine strukturierte Dateneingabe erfordert und ermöglicht.

88 Literatur hierzu findet sich bei MISGELD 1991.

89 Inzwischen gibt es Implementationen von ORACLE für PC's.

90 Bei einem Netzwerk sind alle Einzelarbeitsplätze in irgendeiner Weise – es gibt verschiedene Netzwerktopologien – über Kabel miteinander verbunden und können miteinander kommunizieren, ohne daß dazu ein Diskettenaustausch erforderlich wäre.

91 Diejenige Person, die das Netzwerk verwaltet, die höchste Zugangsberechtigung hat, und für alles verantwortlich ist.

92 Vielleicht gibt es mehr Funktionen, die für eine Optimierung des Ablaufes wichtig wären und die ich bisher übersehen habe. Die Aporie zwischen dem Zwang, auf möglichst schnelle Weise das zu erreichen, was unbedingt erforderlich ist, und dem Wunsch, auch das kennenzulernen, was weiterführen könnte, läßt sich nicht aufheben. Es gab Fälle, in denen ich erst relativ spät von Programmfunktionen Kenntnis nehmen konnte, die die Arbeit schon wesentlich früher hätten erleichtern können.

rung des Netzwerks genehmigt war, konnten die Überlegungen über den Einsatz einer Datenbank oder eines vergleichbaren Programmes konkreter werden.

In einer Zeitschrift las ich schließlich die Anzeige einer Computerfirma, die für ein Dokumentationsprogramm mit dem Namen FAUST warb. Die versprochene Flexibilität des Programmes beeindruckte mich nach den bisherigen Erfahrungen. Eine Demonstrationsdiskette ließ den Schluß zu, daß dieses Programm die meisten der Anforderungen, die wir an ein solches Programm stellen mußten, würde erfüllen können. Die Angelegenheit war für mich so wichtig, daß ich selbst zu dieser Firma fuhr, um dort meine Wünsche zu erklären. Es überraschte mich, daß ich in einer kleinen Stadt in der Nähe von Nürnberg einen Familienbetrieb antraf, der aus einigen wenigen Leuten bestand. Aber das Gespräch, das ich dort führte, überzeugte mich von der Möglichkeit, das Programm an die Bedürfnisse der Wörterbucharbeit anzupassen. Einiges war nicht so einfach realisierbar, wie es mir damals erschien. Aber anderes, was ich als Wunsch geäußert hatte, ist inzwischen in die neueste Version des Programmes eingeflossen. Bei einer großen Softwarefirma wäre dies sicherlich nicht möglich gewesen. Und auch die vielen im Lauf der Zeit nötig gewordenen Telefonate über Spezialfragen oder auch die Arbeitsgespräche in der Firma stellten im besten Sinne das dar, was als *Support* bei großen Softwarefirmen häufig vermißt wird oder seit neuestem teuer bezahlt werden muß.

FAUST wird von den Entwicklern selbst nicht als Datenbank im eigentlichen Sinn des Wortes beschrieben. Entwickelt wurde es für gänzlich andere Zwecke als die Produktion eines Wörterbuchs: für die Verwaltung von Bibliotheken, Archiven, Museen, für Wirtschaftsbetriebe wie Banken und Versandhäuser – allein dies zeigt schon, wie vielfältig der Einsatz dieses Programmes von den Herstellern konzipiert war. Und dies machte mir Mut, die Entscheidung für den Kauf einer Netzwerkversion des Programmes mit der Bildarchiv-Komponente<sup>93</sup> zu wagen. Die Entscheidung fiel mir nicht leicht – ich mußte sie allein treffen, und es ging um einen Betrag, der drei Viertel des gesamten Sachmittelletats des DRW für ein Jahr ausmachte. Letztlich hat sich diese Entscheidung als richtig herausgestellt. Es bedurfte freilich einer erheblichen Anstrengung, bis alle Anforderungen unseres Wörterbuchs an ein solches Programm erfüllt waren, und manchmal war ich der Meinung, ich müßte auch diesen Versuch aufgeben.

## 6.2 Der Aufbau von FAUST

FAUST verwaltet die Daten in Dokumentationen,<sup>94</sup> innerhalb derer es verschiedene Objektarten<sup>95</sup> gibt. Diese Objektarten „werden in einer dynamischen Erfassungsmaske erfaßt und gespeichert. Eine solche Maske heißt dynamisch, weil sie nicht starr festgelegt ist, sondern noch während der Erfassung für alle Objekte der Objektart, aber auch für ein individuelles Objekt geändert werden kann. Im Unterschied zu einem konventionellen Datenbanksystem ist man also mit der Erfassungsmaske nicht an einen starren Rahmen gebunden, sondern kann die

93 Dazu vgl. unten 6.3.3.

94 Der Einfachheit halber zitiere ich aus dem Handbuch für FAUST: „Dokumentationen nennen wir die in einem zusammenhängenden Informationsbestand gesammelten Informationen. Diese Informationsbestände, sprich: Dokumentationen, existieren separat voneinander. Sie können ganz unterschiedlichen Aufbau haben und in unterschiedlichen Laufwerken und Pfaden, im Netzwerk sogar auf verschiedenen PCs angelegt werden. FAUST ist dennoch in der Lage, bis zu 13 solcher separaten Dokumentationen gleichzeitig zu öffnen, um zum Beispiel übergreifend darin zu recherchieren.“ (FAUST I 1).

95 „Als Objekte bezeichnen wir die Einheiten einer Dokumentation, die einzeln erfaßt werden können: Dies können Kundenadressen, Firmenporträts, Mitarbeiterdaten, aber auch Bücher, Videos, Bilder usw. sein. Gleichartige Objekte werden als Objektart bezeichnet; in der Dokumentation eines Museums würden etwa Bilder, Skulpturen, Ausstellungskataloge, Bücher, Dias usw. als unterschiedliche Objektarten betrachtet. FAUST verarbeitet pro Dokumentation bis zu 254 verschiedene Objektarten.“ ebd.



Rahmen nach Bedarf in dieser oder jener Richtung ausdehnen.<sup>96</sup> Die Anzahl der Felder je Objektart ist beliebig; die Struktur der Felder geht vom frei beschreibbaren Feld (mit einer Obergrenze von 64 KB) bis zum numerischen und Datumsfeld oder zu einem Interaktionsfeld, über das man in andere Programme wechseln kann. Ich werde auf die einzelnen Feldtypen und ihre Brauchbarkeit für die lexikographische Arbeit noch zurückkommen.

Vor der Planung der Wörterbuchstruktur für FAUST wurden die zur Verfügung stehenden Hilfsmittel bibliographischer Art<sup>97</sup> in die Datenbank integriert. Möglich wurde dies durch das Zusatzprogramm DOWNLOAD zu dem Schwesterprogramm von FAUST, LIDOS, das die DIN-Norm 1506 (ISO 2709) „Format für den Austausch von bibliografischen Daten“ nutzte, um Fremddaten nach FAUST zu importieren. In der neuesten Version von FAUST<sup>98</sup> ist dieser Weg verlassen worden, weil diese DIN-Norm sich nicht durchgesetzt hatte. Nunmehr ist der Import von Fremddaten über einfache Felddescriptoren<sup>99</sup> möglich, ohne den komplizierten Umweg über das DOWNLOAD-Programm nehmen zu müssen. Eine solche Importdatei kann mit jedem beliebigen Texteditor erstellt werden und enthält als Markierungen lediglich die Feldnamen sowie die Bezeichnung der Dokumentation und der Objektart.

## 6.3 Die Quellen- und Titeldokumentation

### 6.3.1 Der Aufbau

Die notwendigen Angaben zu den Quellensiglen, also die Sigle selbst, die Signatur und – falls vorhanden – die Datierung und Zitierweise trennten wir zunächst von den bibliographisch vollständigen Titelaufnahmen, da wir bei der Aufnahme sämtlicher bibliographischer und sonstiger Angaben für jede Quellensigle eine Überfrachtung der Maske und damit die Unübersichtlichkeit der Information befürchteten. Daneben wurde eine Dokumentation eingerichtet, in der diese umfassenden Angaben eingefügt werden sollten.

Es ist hier nicht erheblich, aus welchen Feldern diese Objekte bestehen. Wichtig ist allerdings, daß jedes Objekt in sehr verschiedener Weise recherchiert werden kann. Von besonderer Bedeutung sind für ein alphabetisch strukturiertes Wörterbuch natürlich Recherchemöglichkeiten über ebenfalls alphabetisch geordnete Indexlisten. So gibt es einen Index der Quellensiglen, bei dem die Recherche so erfolgen kann, daß die Liste mit den Cursortasten Zeile für Zeile abgefahren wird. Da dies bei großen Listen nicht praktikabel ist, kann aber auch in ein Suchfeld diejenige Buchstabenfolge eingetragen werden, die eindeutig genug für das gesuchte Objekt ist. Der Vorteil von FAUST ist es nun, daß im Unterschied zu dem Programm WordCruncher der Index nicht statisch ist. Veränderungen der Quelldatei für den WordCruncher werden erst wirksam, wenn die Datei neu indexiert wird. In FAUST wird jeder Index bei jeder Veränderung der darüber erfaßten Objekte automatisch aktualisiert. Weitere Recherchemöglichkeiten seien hier nur aufgezählt; sie dürften denjenigen entsprechen, die in vergleichbaren Programmen verfügbar sind: Volltext-, Stringrecherche mit Berücksichtigung von Klein- und Großschreibung, die numerische Recherche sowie die Recherche nach Objektnummern oder aller Objekte. Für besondere Projekte ist zweifellos die Indexrecherche die attraktivste Möglichkeit, da sie durch den Bezug auf die jeweiligen Felder der Objekte die häufigsten Abfragewünsche erleichtert. Ist die *Recherche* das Mittel, um einzelne Objekte nach individuellen Kriterien zu

96 ebd.

97 Die Quelldatei und die in ASKSAM enthaltenen Daten, vgl. oben 4.3.

98 Seit März 1994.

99 Sie enthalten eine eindeutige Kennzeichnung des Feldnamens und ermöglichen die Zuordnung des Feldinhaltes zu dem betreffenden Feld in FAUST.

suchen, die dem Recherchierenden bereits bekannt sein müssen, so stellen die *Referenzen* in FAUST die Möglichkeit dar, Informationen zu erhalten, die über die eigene Fragekompetenz hinausgehen.<sup>100</sup>

Referenzen sind das Mittel, um in FAUST Datensätze<sup>101</sup> miteinander zu verknüpfen. Sie werden über Referenzfelder hergestellt, bei deren Schaffung bereits die Referenzart<sup>102</sup> angegeben werden muß (allerdings kann sie später geändert werden). Durch die Referenz wird von diesem Feld eines Objektes auf ein anderes Objekt, das nicht derselben Dokumentation angehören muß, verwiesen. Der Verweis stellt von dem Referenzfeld den Einblick in das referenzierte Objekt her. Als Beispiel: Von jedem Objekt der Quellensigle-Dokumentation wurden Referenzen auf die Dokumentation mit den vollständigen Angaben gesetzt. Das Feld *Quellensigle* in der kleineren Dokumentation verwies auf das gleiche Feld des entsprechenden Objekts der größeren Dokumentation, so daß von der aktuellen Quellensigle durch bloßen Tastendruck oder durch Voreinstellung die referenzierte umfassendere Information gezeigt wurde.

Im Lauf der Zeit wurde jedoch deutlich, daß durch einen geschickten Maskenaufbau – d. h. die Konzentration der wichtigsten Daten auf *einem* Bildschirm – das Bedürfnis der schnellen Information mit der Möglichkeit des unmittelbaren Zugriffs auf die Zusatzinformationen verbunden werden konnte.<sup>103</sup> Die Quellensigle-Dokumentation wurde mit der umfassenderen Dokumentation vereinigt, so daß nur noch *eine* Dokumentation benötigt wurde. Diese Dokumentation dient aber nicht nur der Information über die Quellentexte des DRW. Zwar gehört es zu den Funktionen dieser Dokumentation, alle bibliographischen Angaben zu den Quellen des DRW auf dem aktuellsten Stand zu halten und im Lauf der Zeit die bibliographischen inzwischen überholten Titelaufnahmen der verschiedenen Quellenhefte in ein Quellenverzeichnis zusammenzuführen. Diese Funktion wird aber nicht dadurch erschwert, daß auch Objekte gänzlich anderer Art in der Dokumentation enthalten sind, da jederzeit alle Objekte herausgefiltert werden können, die sich auf eine gültige Quellensigle beziehen. Damit kann auch die Durchsicht von Verlagskatalogen und Bibliographien, von Aufsätzen und Neuerscheinungen aller Art hier ihren Niederschlag finden. Und die Titelaufnahme einer Publikation kann auch ohne Durchsicht des Werkes sinnvoll sein, wenn sich aus dem Titel Schlagwörter oder Thesaurusbegriffe ergeben, die für das DRW wichtig sind. In diesen Fällen kann die Kombination bestimmter Titelwörter einen gezielten Zugriff ermöglichen. Die Verbindung von „Erbrecht“, einem Ländernamen und einer Zeitangabe kann einem Bearbeiter bei der Lösung eines heute noch nicht er-

100 So werden bei der Information über eine Rechtsquelle aus dem Bereich des sächsischen Rechts unter Umständen verwandte Rechtsquellen gezeigt, in denen der gleiche Rechtssatz – vielleicht aber aus einem anderen zeitlichen oder räumlichen Bereich oder mit einem anderen Wortschatz – vorkommt. Und es können Angaben vorhanden sein, in welchem Zeitschriftenband die letzte wissenschaftliche Behandlung einer Rechtsquelle zu finden ist. Dies scheint mir die einzige Möglichkeit zu sein, bei dem immens großen Quellenbestand des DRW einigermaßen auf der Höhe der wissenschaftlichen Diskussion zu bleiben.

101 Datensätze sind fest strukturierte Verbindungen von unterschiedlichen Feldern mit unterschiedlichem Informationsgehalt, die in ihrer Verbindung eine sinnvolle Information bereitstellen: Ein Belegtext, der mit einer Datierung versehen ist, eine Fundstelle aufweist und einer Wortbedeutung zugeordnet ist, stellt einen solchen Datensatz dar. Eine Vielzahl von Beleg-Datensätzen kann mit einem übergeordneten Artikel-Datensatz verbunden sein, so daß die verschiedenen und verschiedenartigen Datensätze eine Informationseinheit *Wortartikel* bilden.

102 Es gibt drei unterschiedliche Referenzarten: Einmal diejenige über die Objektnummer, dann über einen bestimmten Index und schließlich die assoziative Referenz, bei der auf beliebig viele Objekte verwiesen wird. Hinzu kommt eine Referenz mit gleichzeitiger Gegenreferenz.

103 Die Flexibilität von FAUST bewährt sich dadurch, daß der Maskenaufbau grundsätzlich immer verändert werden kann. Die Änderungen gelten dann – anders als die individuell wählbare Fenster- und Farbgestaltung des Bildschirms – für alle Benutzer von FAUST und werden sofort wirksam.

kennbaren Problems dienen, wenn er diese Kombination recherchiert. Damit stellt diese eine Dokumentation mit ihren inzwischen ca. 15 000 Objekten den Anfang eines umfassenden Informationssystems zur Rechts-, Sprach- und allgemeinen Geschichte dar.

Ebenso ist es möglich, Lesefrüchte in der Weise zu verwerten, daß man bei der Lektüre eines Aufsatzes in dem darauf bezogenen Objekt in einem besonderen Feld Fundstellen für noch zu behandelnde Stichwörter angibt. Der traditionelle Weg einer solchen Nachexzerption ist das Herausschreiben solcher Informationen auf Archivzettel jeweils mit der vollen Quellenangabe. Und mußte dieser Archivzettel auch noch an der richtigen Stelle im Archiv eingelegt werden,<sup>104</sup> so war die Information zunächst zwar gesichert, aber auch verborgen. In FAUST ist die Information nicht erst bei Abarbeitung des entsprechenden Buchstabens als Zettel verfügbar, sondern sie ist unmittelbar mit dem Abspeichern des Objektes für alle zugänglich. Ein wichtiger Anwendungsfall dieses immer aktuellen Index ist im DRW die Erfassung von Fremdwörtern aus Rechtstexten, weil neben der sehr guten Belegung der rechtssprachlichen Erbwörter die jetzt gleichberechtigt mitbehandelten Fremdwörter<sup>105</sup> unterbelegt sind. Bei der Aufnahme von Belegen aus Quellen mit reichem Fremdwortbestand, beispielsweise juristischer Literatur des 16. Jahrhunderts, macht es keine Mühe, Fundstellen für Zusatzbelege in die Datenbank einzugeben.

### 6.3.2 Vom Herrschaftswissen zum Expertensystem

Entscheidend war bei der Konzeption dieser Dokumentation jedoch eines: Das Wissen um verschiedene Quellen, um Datierungsfragen, Abhängigkeitsverhältnisse der Quellen untereinander usw. war nur sehr sporadisch schriftlich niedergelegt worden. So mißlich dies ist, so verständlich erscheint es nach langer Praxis der Wörterbucharbeit. Spezialwissen wird einem selbstverständlich, und die Mühe des Aufschreibens wie auch die Organisation des Niedergeschriebenen ist nicht unerheblich. Dieses Spezialwissen kann und soll nun in allen möglichen Feldern dieser Dokumentation untergebracht werden, verfügbar für alle und zu jeder Zeit, auch noch in Jahrzehnten, wenn die jetzigen Träger dieses Wissens schon längst nicht mehr am Wörterbuch mitarbeiten. Es wird ein Gesamtsystem des Expertenwissens zur Geschichte der deutschen Rechtssprache entstehen, bezogen auf die Bedürfnisse dieses Wörterbuches, aber im Lauf der Zeit sicherlich auch mit hohem Informationsgehalt für andere Wissenschaftsinteressen. Verknüpfungen sind möglich und es werden immer mehr Verknüpfungen sein, die gerade wegen der Breite der Quellenbasis des DRW für Geisteswissenschaftler, Volkskundler und Historiker jeder Art interessant sein können und die den Nachfragenden unter Umständen über lange Referenzketten mit den verschiedensten Objekten zum gewünschten Ergebnis führen können. Dabei muß betont werden, daß dies ein Nebenprodukt derjenigen Arbeit ist, die für die Informationserschließung im Wörterbuch sowieso geleistet werden muß. Der überschießende Nutzen ist bei dieser Konstellation groß, aber noch keineswegs in allen Bereichen zu überschauen.

### 6.3.3 Faksimiles als Buchersatz

Die Schilderung der Quellendokumentation im DRW läßt sich nicht abschließen, ohne auf eine faszinierende Möglichkeit einzugehen, die FAUST bietet: Das Bildarchiv. Ursprünglich für Museen und Archive gedacht, vermag diese Funktion gerade auch einem historischen Wörterbuch zu dienen. Eine Vielzahl von Quellen des DRW ist nur über Fernleihe aus auswärtigen

<sup>104</sup> Falls diese Arbeit nicht im Interesse einer schnelleren Publikation überhaupt unterbleiben mußte.

<sup>105</sup> Ich will hier auf den umstrittenen Fremdwortbegriff nicht eingehen: Die Tatsache, daß in der Behandlung der Worttypen seit 1989 kein Unterschied mehr gemacht wird, spricht für sich.

Bibliotheken erhältlich. Teilweise werden die älteren Druckwerke nicht mehr ausgeliehen, teilweise dauert der Bestellvorgang sehr lange, und die Bücher sind gerade wieder zurückgeschickt worden, wenn der nächste Beleg aus dieser Quelle auftaucht. Das Bildarchiv in FAUST kann dazu genutzt werden, von diesen Quellen Faksimiles anzulegen, so daß eine Buchseite als Graphikdatei in FAUST abgelegt und unter der jeweiligen Quellensigle verwaltet wird. Damit ist der sofortige Zugriff auf das Faksimile möglich, ohne einen Bestellvorgang einleiten zu müssen.<sup>106</sup> Der Speicherplatz einer Buchseite ist, gerade wenn sie komprimiert ist,<sup>107</sup> nicht so groß, daß es ernsthafte Speicherprobleme geben müßte. Schließlich können auch, wenn genügend Bilddateien zusammengekommen sind, diese entweder intern<sup>108</sup> oder aber extern auf CD-ROM geschrieben und über ein normales CD-ROM-Laufwerk im Netzwerk verfügbar gehalten werden.

Bedenkt man, wie groß der zeitliche Aufwand ist, um ein solches Quellenwerk für nur einen Beleg in die Hand zu bekommen, so wird der Nutzen dieses Bildarchivs und der Faksimilierung eines Teiles der Quellen evident.

#### 6.4 Die Wortartikeldokumentation

Der Aufbau einer Dokumentation, in der Wortartikel eingegeben werden konnten, bedurfte einer eingehenden Planung. Verschiedene Artikelteile mußten getrennt voneinander behandelt werden: Ein Wortartikel ist immer hierarchisch aufgebaut, wobei der Lemmaeintrag die hierarchische Spitze darstellt, von der alle anderen Artikelteile abhängen.<sup>109</sup> Dabei steht der Artikelkopf mit den Wortart- und sonstigen sprachlichen Angaben an erster Stelle, von ihm hängt die Bedeutungsangabe (oder bei einem polysemen Wort: die Bedeutungsangaben) ab, und von diesen schließlich die Belege.

Bei der Planung dieser Dokumentation standen am Anfang die Überlegungen, welche Feldtypen für welche Artikelteile geeignet waren. Volltextfelder, die frei beschreibbar waren, boten sich neben Feldern fester Länge an, Referenzfelder hatten je nach der Art der Referenz eine unterschiedliche Funktion. Ich begann mit dem Feld für den Lemmaeintrag, das ich mit einer festen Länge von 40 Zeichen definierte. Sogar ein Wort wie *Kaffeeschenkspersonalgerechtsame*, das nach den heutigen Kriterien für die Aufnahme von Kompositen nicht mehr bearbeitet würde, schöpft diese Zahl nicht aus. Die Recherche auf Einträge in diesem Feld erfolgt über einen alphabetischen Index; jeder Eintrag in diesem Feld wird automatisch in den Index gestellt und kann darüber gefunden werden. Bei dieser Konzeption hatte ich allerdings nicht berücksichtigt, daß die Wortartikel im DRW im Prinzip zwar alphabetisch angeordnet werden, aber nicht streng nach der Reihenfolge aller Buchstaben eines Wortes, sondern getrennt nach Wortbestandteilen ohne Berücksichtigung irgendwelcher Kompositionsfugen. Ob ein Wort als *Lehngericht* oder als *Lehnsgericht* belegt ist, wird für die alphabetische Reihung nicht berücksichtigt; sie erfolgt nach *Lehn-* und *-gericht*. Als Konsequenz mußte die Ordnungsfunktion des

106 Freilich steht dem manchmal das Kopierverbot entgegen. Aber da in der Regel Mikroverfilmungen möglich sind, bedarf es dann nur noch des Scannens der Mikrofilmaufnahme, so daß sie in eine Graphikdatei überführt wird.

107 FAUST besitzt interne Kompressionsmöglichkeiten, die eine Bilddatei auf ca. 40 Kilobyte komprimieren. Auf einer CD-ROM mit 600 Megabyte Speicherplatz könnten somit ca. 1.500 Faksimiles archiviert werden.

108 Das sogenannte Inhouse-CD-ROM-Publishing wird finanziell bald interessant werden.

109 Ich behandle diesen Bereich ganz bewußt völlig untheoretisch, da es hier nicht auf den lexikographischen Status dieser Artikelteile ankommt, sondern lediglich auf die pragmatische Frage: Wie kann ich einen Wörterbuchartikel in einer Datenbank abbilden?

Lemmas von der Abbildung des Lemmazeichens getrennt werden. Ein Feld nahm die Zeichenfolge auf, nach der sortiert werden sollte, ein anderes enthielt die Zeichenkette, die gedruckt wurde. Dies löste auch das Problem der Sortierung solcher Lemmata, die mit Indizes als Homonyme oder mit Klammern als erschlossene Konstruktlemmata gekennzeichnet werden mußten.<sup>110</sup>

Die Maske, die mit diesen beiden Einträgen begann, enthielt neben einem in der Länge nicht beschränkten Feld für die Bedeutungserläuterung viele weitere Felder, auf die hier nicht weiter eingegangen werden muß, da sie mit den verschiedenen Sprachen und Sprachvarietäten zusammenhängen, die im DRW behandelt werden. Es genügt, daß alle Angaben sprachlicher Art in der Maske untergebracht wurden, ohne daß sie zu unübersichtlich wirkte. Es schien nun nur noch die Zuordnung der Belege zu dem Artikelkopf zu fehlen. Aber vorher stellte sich die Frage, wie polyseme Wörter behandelt werden konnten. Nicht alle Angaben, die in dem Artikelkopf benötigt wurden, mußten bei den unterschiedlichen Wortbedeutungen auftauchen. So konnten je Wortbedeutung unterschiedliche Angaben zur Zuordnung von Grund- und Bestimmungswort zu in sich wiederum polysemen Simplizia erforderlich werden, Angaben über etymologische Zusammenhänge konnten auch hier sinnvoll sein.

Zunächst wählte ich den Weg, für den Artikelkopf und die dazugehörigen Bedeutungsgliederungen getrennte Objektarten zu schaffen. Erst während der Arbeit mit diesen Objektarten stellte sich heraus, daß dies überflüssig war und ich mit einem Objekt für Artikelkopf und Bedeutungserläuterung bei polysemen Wörtern auskommen konnte. Entscheidend war hierfür eine Haltung, die auf systematische und theoretische Unterscheidungen verzichtete und sich allein auf das konzentrierte, was pragmatisch erforderlich war, um einen Wortartikel mit möglichst wenig Aufwand schreiben zu können. In diesem Verzicht auf eine explizite theoretische Durchdringung der Artikelstruktur und der Abbildung dieser Strukturzusammenhänge in theoretisch abgesicherter Weise sehe ich den Hauptgrund dafür, daß wir mit den so entwickelten Masken arbeiten können. Als Ergebnis dieser Überlegungen und Experimente gibt es nun eine Objektart (Artikelkopf), in der neben dem Lemma alle sprachlichen Angaben und die Bedeutungen erfaßt werden. Sie enthält etwa 40 verschiedene Felder, die zum Teil nur interne Funktionen besitzen, also für den Ausdruck des Wortartikels keine Bedeutung haben. Die zweite Objektart (Beleg) enthält die einzelnen Belege zu dem Lemma und seinen verschiedenen Bedeutungen. Die Zuordnung erfolgt über Referenzen, die bei der Erfassung eines Belegobjektes automatisch zu dem betreffenden Artikelobjekt, das freilich entweder existieren muß oder während der Erfassung des Belegobjektes en passant miterfaßt werden kann, hergestellt wird. Die Referenz bewirkt, daß bei jeder Veränderung der Artikelstruktur – wenn z.B. Gliederungspunkt III zu Punkt IV 1 a wird – sämtliche Belege diesem Gliederungspunkt zugeordnet bleiben. Andererseits muß bei einer neuen Zuordnung eines Beleges zu einer anderen Bedeutung die Referenz neu hergestellt werden.

## 6.5 Die Erfassung eines Wortartikels

### 6.5.1 Der Artikelkopf

In der Erfassungsmaske, in der das Lemma und die notwendigen sprachlichen Angaben eingetragen werden, muß zunächst nur das Feld „Gliederungsebene“ einen Eintrag enthalten.<sup>111</sup> Die

110 Obwohl die Nichtberücksichtigung dieser Zeichen auch über die Definition einer Sortierfolge möglich gewesen wäre.

111 Die Definition eines Feldes in einer Datenbank als Pflichteintragungsfeld bewirkt, daß ein Datensatz nur dann

Bezeichnung für das Feld ist relativ willkürlich gewählt. Andere Felder stehen zur Verfügung, werden aber nur bei Bedarf ausgefüllt. Im Feld „Gliederungsebene“ wird der Lemmaeintrag in der Form hineingeschrieben, in der der Wortartikel alphabetisch sortiert werden soll. Eine Hochzahl, die Homonyme voneinander unterscheidet, wird codiert hinter den Lemmaeintrag gestellt, um das Sortierergebnis nicht zu verfälschen. Da FAUST bei der Sortierung Groß- und Kleinschreibung nicht unterscheidet, gelten für die Datenbank gleichlautende Substantive und Adjektive als Homonyme, die besonders gekennzeichnet werden müssen. Das Drucklemma, Angaben zur Wortart, zu den verschiedenen Normalformen der Sprachen und Sprachvarietäten, die im DRW behandelt werden, und zur *Wortherkunft*<sup>112</sup> werden hier eingetragen. Bei einem monosemen Wort erfolgt der Eintrag der Bedeutungserläuterung in diesem Objekt.

Handelt es sich um ein polysemes Wort, so werden zu dem ersten Objekt weitere Objekte dieser Objektart angelegt, die durch Referenzen mit dem ersten Objekt verbunden sind. Die im ersten Objekt stehenden Angaben werden hier nur dann ergänzt, wenn sie sich für diese Wortbedeutung davon unterscheiden.

Ist dies für einen Artikelkopf alles selbstverständlich, so gibt es zwei Bereiche, die einer genaueren Darstellung bedürfen: Der eine, weil er in allen Wörterbüchern eine wichtige Rolle spielt, der andere, weil er die zentrale Funktion eines historischen Belegwörterbuchs ausmacht – die Verweisstruktur und die Darstellung der Belege.

### 6.5.2 Die Verweisstruktur

Zu den schwierigsten Problemen der Lexikographie gehört die Verweisstruktur und die Pflicht, Verweise nicht ins Leere gehen zu lassen.<sup>113</sup> Im DRW gibt es Verweise, die die Wortbildung transparent machen sollen: Bei Kompositen Verweise auf Bestimmungswort und Grundwort und bei Ableitungen Verweise auf das Wort, von dem das Lemma abgeleitet ist.<sup>114</sup> Neben den Verweisen auf bedeutungsverwandte Wörter<sup>115</sup> gibt es Verweise auf sinnverwandte Wörter, die dem Benutzer sachlich-enzklopädische Informationen<sup>116</sup> über das durch das Wort bezeichnete Rechtsinstitut oder die bezeichnete Sache vermitteln sollen. Sie können auch in der Form vorkommen, daß im kursiven Erklärungstext ein Wort *recte* gesetzt ist.<sup>117</sup>

Die Reflexion der Verweisstruktur für diesen Aufsatz zeigte mir, daß ich bei der Übertragung dieser Struktur in FAUST immer noch zu konventionell gedacht hatte: Die Verweise auf die Wörter, die bei einem Kompositum Grund- und Bestimmungswort bilden, wurden ebenso voneinander getrennt wie der Verweis auf eine Wortableitung. Dies führte zwangsläufig zu der Konzipierung von jeweils verschiedenen Feldern für jede Verweisart. Abgebildet wurde damit

---

in die Datenbank aufgenommen (abgespeichert) wird, wenn dieses Feld einen Eintrag enthält – freilich kann dies auch ein falscher Eintrag sein, denn keine Datenbank läßt sich gegen menschliche Fehler schützen.

- 112 Die Etymologie der Lemmata wird sehr restriktiv behandelt, da sich das DRW als *komplementäres* Wörterbuch versteht, das nicht alle die Informationen geben muß, die in anderen Wörterbüchern bereits verfügbar sind.
- 113 Vgl. dazu BLUMENTHAL/LEMNITZER/STORER 1987.
- 114 Sie hatten bisher die Form: „Bestw. zu...“, „Grdw. zu...“ und „zu...“.
- 115 Die – unabhängig von dem Grad der Synonymität – mit der Sigle: „Syn.“ kenntlich gemacht werden.
- 116 Sie haben die Form „vgl...“ und gehen über die Erschließung des onomasiologischen Feldes hinaus; vgl. dazu WEGNER 1989.
- 117 Für die strukturierte Datenerfassung ist dieser Verweis praktisch nicht existent, da er im unstrukturierten Fließtext des Erklärungsfeldes vorkommt und ein Zugriff nur über eine Volltextrecherche möglich wäre. Die Konsequenz besteht darin, daß das Verweiswort vom Bearbeiter zusätzlich in ein Verweisfeld eingetragen werden muß.

jedoch eine redundante Struktur, die auch auf das Wesentliche reduziert werden konnte. Ein Verweisfeld für die Wortbildung. In dieses Feld könnte beispielsweise bei dem Kompositum *Marktmeister* zunächst der Verweis auf *Markt (I)* eingetragen werden. Das Druckformat stellt im Druck vor den Feldeintrag „zu“ in kursiver Schrift. Das Verweisfeld wird verdoppelt und als zweiter Verweis konnte *Meister (V)* eingetragen werden. Bei einer solchen Feldverdoppelung fügt das Druckformat zwischen die beiden Feldinhalte „und“ ein, abgeschlossen wird der Eintrag im Druck durch ein Semikolon. Der Hinweis, daß der eine Verweis sich auf das Bestimmungswort, der andere sich auf das Grundwort bezieht, kann als unmittelbar evident entfallen. Und da bei der ursprünglichen Konzeption diese Felder auch noch danach unterschieden wurden, ob das Verweiswort bereits abschließend bearbeitet war, konnten dadurch insgesamt sechs Verweisfelder durch nur eines ersetzt werden.<sup>118</sup>

Die Verweisstruktur kann naturgemäß nur dort konstant sein, wo die Verweisartikel keinen Veränderungen mehr unterliegen. Verweise auf Wortartikel, die noch bearbeitet werden oder deren Bearbeitung erst in einem noch nicht absehbaren Zeitraum erfolgt, laufen Gefahr, daß sie ins Leere gehen, sei es, daß das Verweiswort nicht bearbeitet wird oder anders lemmatisiert wird.<sup>119</sup> Daher darf grundsätzlich nur auf solche Wortartikel verwiesen werden, deren Aufnahme in das Wörterbuch geklärt und deren Lemmatisierung eindeutig ist. Was aber geschieht mit Wörtern, die gerade bearbeitet werden? Ein Verweis auf eine konkrete Wortbedeutung wäre hier möglich, droht aber während der gesamten Bearbeitungszeit durch Änderungen der Artikelstruktur ins Leere zu gehen.

FAUST bietet die Möglichkeit, die Verweisstruktur konstant zu halten, wenn sie sich auf Wortartikel innerhalb des Systems beziehen. Das Mittel dazu sind Referenzfelder. Als Beispiel diene ein Verweis der Form „vgl.“ unter *nacheilen (I)* steht ein Verweis: *vgl. nachfolgen (I 2)*. Beide Wörter sind noch nicht endgültig bearbeitet, ihre Artikelstruktur kann sich jederzeit ändern. Der Verweis wird über ein Referenzfeld erfaßt, dessen Eintrag aus dem Index mit den Lemmata bzw. deren Untergliederungen übernommen werden muß. Dazu muß das Lemma, auf das verwiesen wird, in FAUST als Objekt bereits bestehen. In diesem Fall zeigt die Liste den Eintrag *nachfolgen 1.2*,<sup>120</sup> der durch Tastendruck übernommen werden kann. In dem Objekt *nachfolgen 1.2* wird dabei automatisch ein Gegenverweis eingetragen: *Verweis von: nacheilen 1.0*. Damit ist jederzeit die Kontrolle möglich, von welchen Lemmata auf das aktuelle Lemma verwiesen wird. Der Verweiseintrag bewirkt zugleich, daß in einem besonderen Fenster das Verweisobjekt angezeigt wird, so daß die Überprüfung, ob der Verweis durch die Bedeutungserklärung des Verweisobjektes gerechtfertigt ist, mit einem Blick erfolgen kann. Da der Verweis auf das Feld gesetzt ist, in dem die Artikelstruktur und die Stelle dieser Wortbedeutung innerhalb des Wortartikels verwaltet wird, bleibt der Verweis immer konsistent. Wenn nämlich

118 Bei dieser kritischen Überprüfung der Datenbankstruktur war es mir möglich, die mir inzwischen vertrauteren Funktionen von FAUST derart zu nutzen, daß die Erfassungsmaske für den Wortartikel nunmehr statt aus 84 Feldern nur noch aus 28 Feldern besteht und damit übersichtlicher wird. Und die Anzahl der Feldformate, durch die bestimmt wird, in welcher Weise die Feldinhalte ausgedruckt werden, konnte ich so von 98 auf 34 reduzieren, ohne eine Einschränkung der Funktionalität hinnehmen zu müssen. Es ist einer der größten Vorzüge von FAUST, daß derartige Umstrukturierungen zu jeder Zeit möglich sind.

119 Was bereits dann der Fall ist, wenn sich bei der Bearbeitung ein Fall von Homonymie herausstellt.

120 Die Ordnungszahl unterscheidet sich von derjenigen, die später gedruckt wird. Es wäre aber unmöglich gewesen, die konventionelle Markierung der Wortbedeutungen durch römische und arabische Zahlen und dann Kleinbuchstaben zur automatischen Anordnung der jeweiligen Objekte in FAUST zu nutzen. Insofern muß die Gliederungsstruktur durch arabische, durch Punkt voneinander getrennte Ziffern erfolgen, die in der automatischen Nachbearbeitung der Textdatei durch die konventionelle Markierung ersetzt werden.

der Eintrag *nachfolgen* 1.2 auf Grund einer Umstrukturierung des Artikels in *nachfolgen* 3.5 geändert würde, so würde sich der Eintrag bei *nacheilen* 1.0 automatisch entsprechend ändern. Ebenso ist es natürlich mit einer Änderung der Artikelstruktur von *nacheilen*.

Wie aber steht es mit Verweisen auf Wortartikel, die bereits gedruckt sind und noch nicht mit FAUST erfaßt wurden? Abgesehen von der vorläufig noch nicht realisierbaren Absicht, alle Wortartikel des DRW in FAUST zu integrieren,<sup>121</sup> schien dieses Problem zunächst kaum lösbar zu sein. Hat man jedoch die Arbeitsweise des Lexikographen vor Augen, so bietet sich eine ganz einfache Lösung an: Ein Verweis erfolgt immer auf ein Wort, das der Lexikograph bei der Artikelarbeit nachgeschlagen hat. Der Lexikograph sieht also die Wortbedeutung gedruckt vor sich, während er ein Objekt erfaßt, und muß den Verweis dort eintragen. Dann kann er auch ohne großen Arbeitsaufwand aus dieser Erfassung heraus ein Referenzobjekt erfassen, in das lediglich das Lemma mit der Markierung der jeweiligen Gliederungsstruktur und die Bedeutungserklärung eingetragen werden muß.<sup>122</sup> Damit wird in FAUST eine Reihe von Objekten der Objektart Artikel geschaffen, die zunächst nur die Funktion haben, den Verweis auf sie zu ermöglichen. Bei Verweisen auf die gedruckte Strecke fällt bei dieser Methode jederzeit die inhaltliche Richtigkeit des Verweises ins Auge und der Endredaktor muß nicht mehr den gedruckten Text nachschlagen, um dies zu überprüfen.<sup>123</sup> Im Lauf der Zeit wird dadurch eine Vielzahl von Wortartikeln und -bedeutungen des gedruckten Teiles des Wörterbuchs in FAUST integriert werden. In den etwa sechs Monaten, die zwischen der Konzeption dieser Verweisstruktur und dem Abschluß dieses Aufsatzes vergangen sind, wurden bereits ca. 1.000 dieser Objekte angelegt. Und da mit der Erfassung eines solchen „Verweis“-Objektes zugleich die Gegenreferenz auf das ursprüngliche Objekt gesetzt wird, steht jetzt beispielsweise im Objekt *Kreisoberst* ein Verweis auf das Wort *Nachgeordnete*, der bei der Bearbeitung von *Kreisoberst* noch nicht hätte gesetzt werden können. Die rückläufige Vernetzung der Wortartikel führt automatisch auch zu einer Vernetzung der bereits gedruckten Artikel mit den folgenden Buchstabenstrecken – freilich nur in der Datenbankfassung des Wörterbuchs. Ebenso wirkt sich diese Technik auf die noch ungedruckte Wortstrecke aus: Von einer gegenwärtig zu bearbeitenden Wortbedeutung kann auf ein Wort am Ende des Alphabets verwiesen werden, wenn ein entsprechendes Objekt angelegt wird.<sup>124</sup> Es entspricht einer vorausschauenden Arbeitsökonomie, wenn die Objekte zu diesen Wortbedeutungen der betreffenden Wörter gleich angelegt werden und darauf verwiesen wird, auch wenn der Ort dieser besonderen Bedeutung im späteren Wortartikel noch variabel ist. Dabei wirkt sich die Flexibilität der Datenbank gegenüber dem unveränderlichen gedruckten Text des Wörterbuchs so aus, daß die gesamten und aktua-

121 Vgl. dazu unten 7.1.

122 Dies kann theoretisch auch so erfolgen, daß der Lexikograph lediglich das Lemma mit Markierung einträgt und die Übertragung der Bedeutungserklärung Hilfskräften überläßt. Da aber ein solches Referenzobjekt während der Erfassung des eigentlichen Objektes geschaffen werden kann, ist es sinnvoll, dies auch tatsächlich sofort zu tun.

123 Weitergehende Möglichkeiten von Wortschatzuntersuchungen auf Grund dieser Verweisstruktur liegen außerhalb meines Interesses, sind aber sicherlich denkbar. Die Referenzstruktur von FAUST ermöglicht es, sich von Verweis zu Verweis führen zu lassen, indem man jeweils das Referenzfeld anklickt und sich das nächste Referenzobjekt zeigen läßt. Daß solche Referenzketten auch ausgedruckt werden können, ist selbstverständlich. Die erste Erfahrung mit dem Abschreiten solcher Referenzketten im bereits publizierten Teil des Wörterbuchs läßt manchmal auch Zweifel aufkommen, ob Verweise immer sinnvoll gesetzt wurden. Für die Zukunft wird auf Grund dieser Erfahrung sicherlich die Verweispraxis reflektiert werden müssen.

124 Was sehr sinnvoll sein kann, wenn ich beispielsweise bei der Bearbeitung von *Nachklage* feststelle, daß eine der Bedeutungen dieses Wortes synonym mit *Gegenklage* und antonym mit *Vorklage* ist und in einem noch zu klärenden Verhältnis zu *Widerklage* steht.



lisierten Vernetzungen in Zukunft nur in der elektronischen Fassung des Wörterbuchs greifbar sein werden.<sup>125</sup>

### 6.5.3 Der Beleg

Vor der Erfassung eines konkreten Beleges steht die Aufarbeitung des Zettelmaterials. Die Struktur und Gestalt der Zettel läßt es sinnvoll erscheinen, zunächst eine gewisse Ordnung in die Archivzettel, die zu einer Lemmaform im Archivkasten liegen, zu bringen. Bei Wörtern mit vielen Belegen ist immer eine Reihe von Zetteln mit Belegtext versehen, so daß zunächst eine Grobgliederung nach Bedeutungen erfolgen kann.<sup>126</sup> Innerhalb dieser Gliederungen oder bei einem noch ungegliederten Wort werden die Zettel chronologisch sortiert, soweit Datierungen vorhanden sind. Bei einem Wort wie *nachgehen* mit ca. 160 Zetteln ist dies relativ schnell erledigt. Nun bedarf es aber noch des Zugriffes auf die korrekte Quellensigle und die Signatur der Quelle. Mit dem Karteikasten bzw. dem kopierten Quellenbuch bedeutete dies 160mal einen alphabetischen Nachschlagevorgang in 16 Katalogkästen oder in drei Bänden sowie die Beschriftung des Zettels mit der korrekten Quellensigle und der Signatur. In FAUST ist dies einfacher gestaltet: In der Dokumentation DRWTITEL wird in dem Recherchemenü die Recherche „Index Quellensigle“ gewählt. Dabei kann die gesuchte Sigle in ein Suchfeld mit der Mindestanzahl von Zeichen hineingeschrieben werden, die eine eindeutige Identifikation ermöglicht. Als Ergebnis wird der entsprechende Eintrag<sup>127</sup> gezeigt und kann mit einem Tastendruck in ein Rechercheergebnis übernommen werden. Die Bearbeitung der 160 Zettel für *nachgehen* erfordert einen Zeitaufwand von einer Stunde; das Ergebnis wird als Liste ausgedruckt, in der die Sigle mit Bandangaben usw. und die Signatur sowie ein Vermerk „Bestellen?“ steht.<sup>128</sup> Das Rechercheergebnis wird gespeichert und läuft in einem parallelen Fenster bei der Abarbeitung der Zettel mit, so daß die jeweiligen Quellensiglen zeilenweise aus diesem Ergebnis abgerufen werden können.

Vor der Eingabe von Belegen zu einem Lemma muß ein Objekt mit Artikelkopf bestehen, damit alle Belege diesem Objekt zugeordnet werden können.<sup>129</sup> Alle Felder des Artikelkopfes außer dem Lemmaeintrag selbst können später zu einem beliebigen Zeitpunkt der Bearbeitung ergänzt werden. Wenn es allerdings möglich ist, bereits bei der Durchsicht der Archivzettel zu einem polysemen Wort eine Grobgliederung nach Bedeutungen vorzunehmen, so können die

125 Nabeliegende Überlegungen zum Vertrieb einer solchen maschinenlesbaren Fassung sind vorläufig noch hintangestellt worden. Zunächst müssen Erfahrungen gesammelt werden, inwieweit ein Import der gedruckten Wortartikel in zumindest teilweise automatisierter Form möglich sein wird (vgl. dazu unten 7.1). Deutlich ist aber bereits geworden, daß bei der Aufnahme gedruckter Wortartikel in FAUST keine 1:1-Übernahme sinnvoll ist, sondern diese Artikel an die jetzigen Artikelnormen angepaßt werden müssen. Die maschinenlesbare Fassung des DRW wird sich damit immer weiter von der gedruckten Fassung entfernen und vor allem weitergehende Informationen bereitstellen.

126 Zu diesem Arbeitsvorgang vgl. REICHMANN 1990 b 231ff. Er ist die Grundlage historischer Wörterbucharbeit; wenn im DRW teilweise anders gearbeitet werden muß, so liegt dies nur daran, daß der Exzerptionsvorgang zeitlich auseinandergerissen ist und der Abschluß der Exzerption und damit die Voraussetzung für die erste Grobgliederung erst mit dem Herausschreiben der Belegtexte durch die Lexikographen gegeben ist.

127 Beispielsweise „QNPrivatR. II I S.“.

128 Nicht alle erforderlichen Angaben stehen schon in der Dokumentation. Sie werden gegebenenfalls vom Zettel in die Liste übertragen und der Bibliothekarin zum Eintrag in die Dokumentation gegeben.

129 Allerdings kann ein solches Objekt auch während der Erfassung des Belegobjektes geschrieben werden: Wenn ich beim Schreiben des Belegtextes merke, daß er doch keiner der bisher vorhandenen Wortbedeutungen zugeordnet werden kann, habe ich die Möglichkeit, ein neues Objekt hierfür noch während der Erfassung des Beleges anzulegen.

entsprechenden Objekte vorab geschaffen und die Belege damit gleich zugeordnet werden. Sie können dann zeilenweise aus einem Rechercheergebnis abgefragt werden, wobei die jeweilige Bedeutungserklärung in einem Fenster gezeigt wird. Damit ist es möglich, einen Beleg einer bestimmten Wortbedeutung zuzuordnen, ohne alle Bedeutungen im Gedächtnis behalten zu müssen: Die Reihe der Bedeutungserklärungen kann mit dem Cursor so lange abgefahren werden, bis eine Erklärung erscheint, die den Belegtext abdeckt. Die Zuordnung erfolgt mit einem Tastendruck, der die Herstellung einer wechselseitigen Referenz bewirkt. Damit kann bei jedem Aufruf dieser Wortbedeutung jeder ihr zugeordnete Beleg angesehen werden, umgekehrt aber auch bei jedem Aufruf eines Belegobjektes die dazugehörige Bedeutungserklärung. Dadurch kann andererseits der Erläuterungsteil ständig den hinzugekommenen Belegen angepaßt werden.

In einem frei beschreibbaren Feld wird der Belegtext eingegeben. Ein Kommentarfeld,<sup>130</sup> das bei Bedarf geöffnet werden kann, zeigt die Kodierungen der Sonderzeichen und andere Regeln. Als nächstes folgt in der Erfassungsmaske ein Feld für Kurzkommentare des Bearbeiters.<sup>131</sup> Reicht die Feldlänge nicht aus, so kann das Feld beliebig oft vervielfältigt werden. Hierauf folgt die Übernahme der Quellensigle<sup>132</sup> aus dem gespeicherten Rechercheergebnis durch Tastendruck. Die Einsicht in das betreffende Objekt der Dokumentation DRWTITEL geschieht ebenfalls durch einfachen Tastendruck; hier werden alle zu der Quellensigle gespeicherten Informationen gezeigt. Zunächst aber interessiert die Datierung der Quelle, die – wenn sie festgelegt ist – in einem besonderen Feld des referenzierten Objekts angegeben ist. Diese Angabe muß in das Feld „Datierung“ geschrieben werden,<sup>133</sup> darauf erfolgt ein Eintrag in dem Feld „Sortierdatierung“, das ein numerisches Festkommefeld<sup>134</sup> ist, in dem eine vierstellige Jahreszahl, erweitert durch zwei Stellen nach dem Komma, eine eindeutige Sortiermöglichkeit für den Computer bietet.<sup>135</sup> Ist dem Mitarbeiter die Sortierdatierung entfallen, so kann sie aus gespeicherten Texten abgerufen werden. Als nächstes Feld wird die genaue Fundstelle eingetragen. Da die Zitierweisen im DRW unterschiedlich sein können,<sup>136</sup> steht in dem referenzierten Objekt die Information zu abweichenden Zitierweisen in einem dafür vorgesehenen Feld. Das nächste Feld ermöglicht es, eine Angabe zur Textsorte oder zur regionalen Zuordnung des

130 Kommentarfelder können frei beschrieben werden und sind ein wesentliches Mittel, um die Struktur der Datenbank durchsichtig zu halten. In jeder Situation kann das Kommentarfeld aufgerufen und gelesen werden, so daß im Idealfall eine kontextsensitive Hilfe, d.h. ein Text, der ein gerade an dieser Stelle auftretendes Problem erklärt, zur Verfügung steht. Es verlangt allerdings eine erhebliche Disziplin, bei der erforderlichen Schnelligkeit der Arbeit an der Datenbankstruktur das Kommentarfeld immer zu beschreiben.

131 Für die Zukunft wird es gravierende Veränderungen in der Arbeitsweise allein schon dadurch geben, daß alles, was in FAUST erfaßt ist, jedem zu jeder Zeit zugänglich ist und mit Anmerkungen versehen werden kann. Selbstverständlich gibt es in FAUST die Möglichkeit, den Zugang zu Objekten und einzelnen Feldern dieser Objekte individuell zu beschränken. Dies würde aber auf dieser Ebene gerade den Nutzen einer Datenbank, der in deren ständiger Transparenz und Aktualität besteht, erheblich vermindern.

132 Selbstverständlich gibt es auch ein Feld, in das Quellentitel eingetragen werden, für die keine Quellensigle festgelegt worden ist.

133 Die Übernahme der Quellensigle *zusammen* mit anderen Feldern des Objektes ist in FAUST nicht möglich.

134 Also ein Feld, in dem nur Ziffern eingetragen werden dürfen.

135 ULRIKE RÖHL, die als wissenschaftliche Mitarbeiterin insbesondere für schwierige Fragen aus dem Gebiet der westgermanischen Sprachgeschichte zuständig ist, hat dieses System entwickelt: um 1666? – eine Datumsangabe, die kein Programm korrekt sortieren kann – wird als 1666,07 kodiert.

136 Gesetzeswerke können nach Buch, Artikel und Paragraph zitiert werden, aber auch, wenn dies leichter zu finden ist, nach Seiten.

Beleges vorzunehmen.<sup>137</sup> Einträge in diesem Feld werden immer ausgedruckt und zugleich durch einen Index erfaßt, so daß bei späteren Recherchen darauf rekuriert werden kann.<sup>138</sup> Ein Feld „Belegklassifikation“ dient der internen Information<sup>139</sup> über Textsorten, Sprachvarietäten usw. In Anlehnung an die onomasiologische Vernetzung des FWB folgt darauf ein Feld, mit dem diese auch im DRW hergestellt werden soll. Hier werden Oberbegriffe eingetragen, die später beispielsweise den Zugriff über den Oberbegriff „Abgabe“ auf alle Bezeichnungen hierfür ermöglichen.

Zwei wichtige Felder ermöglichen die Beibehaltung aller Informationen auch bei der Entscheidung, daß der Beleg oder der Belegtext für den Wortartikel irrelevant sind. War dies bei der zeilenorientierten Texteingabe nicht möglich, da alle Informationen unmittelbar für den Textsatz verwertet wurden – waren also weggestrichene Texte unwiederbringlich verloren – so genügt es hier, ein einstelliges Feld „Beleg weglassen“ auszufüllen, damit der gesamte Beleg im gedruckten Wortartikel nicht erscheint. Ein Eintrag im Feld „Nur Fundstelle“ bewirkt, daß der Belegtext weggelassen wird und nur die Quellenangaben gedruckt werden.

Bei der Belegeingabe begegnen im abzuschreibenden Text immer wieder Wörter, die für später zu bearbeitende Lemmata wichtig sein können. Ob diese Fundstellen als Archivzettel bereits vorhanden sind, kann in diesem Bearbeitungsstand in sinnvoller Weise kaum nachgeprüft werden. Eine gezielte Nachexzerption der Quellen während des Herausschreibens von Belegen ist in einem konventionellen Zettelarchiv aus arbeitsökonomischen Gründen (Ausfüllen eineszettels, Einordnen in das Archiv usw.) nicht möglich. In FAUST hingegen genügt es, einen Eintrag in das Feld „verwendbar für:“ vorzunehmen, um für die weitere Bearbeitung den Verweis auf diesen Belegtext zu erhalten. Auch diese Einträge werden in einem sofort aktualisierten Index mitgeführt, der in jedem Bearbeitungsstadium für die Suche nach weiteren Belegen zur Verfügung steht. Damit diese Suche bei der Belegerfassung erfolgen kann, läßt sich die Existenz eines solchen Zusatzbeleges über einen Index abfragen, aus dem die Übernahme in den Wortartikel durch Tastendruck vorgenommen kann. Der Verweis erfolgt in einem Feld „Fundartikel“ und bewirkt, daß im Artikeltext ein Beleg des Typs „Beleg s. unter Fundartikel“ steht. An dieser Stelle habe ich freilich die Struktur der Wortartikel geringfügig verändert: Da sich die Verweise auf Belege zu einem Wort, die in einem anderen Wortartikel zu finden sind, durch diese Recherchemöglichkeiten häufen, habe ich eine eigene Position geschaffen, in der am Schluß des Artikels oder eines Gliederungspunktes derartige Fundartikel in chronologischer Reihenfolge der Belegung aufgeführt werden. Somit erhält der Benutzer zunächst die Belege mit Belegtext zur Verfügung gestellt, deren Funktion es ja auch ist, die Bedeutungserklärung zu verifizieren, und erst danach folgen die Angaben weiterer Fundstellen im Wörterbuch, die er bei einem erhöhten Nachschlageinteresse verfolgen kann.<sup>140</sup>

Auf das Volltextretrieval-Programm WordCruncher gibt es den Zugriff über ein Interaktionsfeld, das dieses Programm aufruft – das Recherchergebnis wird als ASCII-Datei gespei-

137 Dies ist erforderlich, wenn die Region nicht bereits aus der Quellensigle ersichtlich ist, wie etwa bei Weistümersammlungen aus dem gesamten deutschsprachigen Bereich.

138 Z.B. die Recherche auf alle Belegobjekte, die innerhalb eines bestimmten Zeitraums in einem bestimmten räumlichen Bereich existieren.

139 Es wird also nicht ausgedruckt.

140 Dies ist eine Änderung, die nicht auf die technischen Anforderungen der Datenbank bezogen ist, sondern sich wegen der erweiterten Recherchemöglichkeiten anbietet. Ausschließlich aus datenbanktechnischen Erfordernissen heraus habe ich lediglich die Abtrennung zweier ohne Belegtext erscheinenden Fundstellen, die bisher durch ein Semikolon erfolgte, an die sonst im DRW übliche Abtrennung durch den Punkt angleichen müssen. Dies dürfte auch die häufig geäußerte Befürchtung entkräften, daß Datenbanken zur Kompromissen in der lexikographischen Konzeption zwingen könnten.

chert und kann in ein Belegobjekt in FAUST importiert werden. In der neuesten Version von FAUST kann jedes Feld „invertiert“ werden, wie die Herstellerfirma dies nennt, so daß jede Zeichenkette des Feldes in einen Index geschrieben wird.<sup>141</sup> Diese Möglichkeit wird so genutzt, daß alle jetzt im WordCruncher verwalteten Texte in FAUST importiert werden, so daß der Zugriff noch schneller und komfortabler ist. Mit der Importierung aller maschinenlesbar gemachten Bände des DRW in FAUST – zunächst als unstrukturierte Textdateien – und mit der Invertierung des Feldes „Belegtext“ stehen jederzeit alle Zeichenketten aus Belegtexten für die weitere Nutzung zur Verfügung.<sup>142</sup> In der Praxis zeichnet es sich ab, daß die im Wörterbuch zitierten Quellentexte für die Gewinnung von Zusatzbelegen und zur Schließung von Exzerptionslücken weitaus ergiebiger sind als die maschinenlesbaren einzelnen Rechtstexte. Dies beruht einfach darauf, daß im Wörterbuch eine Vielzahl von Quellen unterschiedlichster Art – wenn auch immer in sehr kleinen Textausschnitten – verarbeitet sind, während die anderen Texte (Rechtsbücher, literarische Werke usw.) den typischen beschränkten Wortschatz *einer* Quelle oder *eines* Autors aufweisen. Als ein relativ beliebiges Beispiel kann der Artikel *Mutter* in seiner gegenwärtig noch nicht abgeschlossenen Form dienen: In Bedeutung I („menschliche Mutter“) stammen von den 27 Belegen 12 aus dem maschinenlesbaren Wörterbuchtext – und es sind sehr aussagekräftige Texte. Ein anderes Beispiel sind die Fremdwörter, deren Exzerption für das Wörterbucharchiv höchst unterschiedlich vorgenommen wurde. In der Regel ist das im Archiv vorliegende Material so ergänzungsbedürftig, daß jede Belegstelle aus dem maschinenlesbaren Text bei der Bearbeitung dankbar aufgenommen wird.

## 6.6 Das Druckformat

Die Konzeption der Datenbankfelder war ein relativ einfaches Problem gegenüber demjenigen, wie der Inhalt dieser Felder für die Weiterbearbeitung mit T<sub>E</sub>X aus der Datenbank exportiert werden konnte. Erst bei dieser Arbeit wurde deutlich, welche Inkonsistenzen in den Formalien ein Wörterbuch mit einer langen Entstehungszeit enthält.<sup>143</sup> Satzzeichen, typographische Merkmale, Reihenfolgen von lexikographischen Angaben wurden relativ willkürlich verwendet – eine stringente Konzeption der Formalien des Wörterbuchs gab es nicht. Nun aber mußte für den Export der Feldinhalte festgelegt werden, in welcher Reihenfolge sie ausgegeben werden sollten, welche Satzzeichen die Informationseinheiten voneinander trennen sollten, was passieren sollte, wenn ein normalerweise ausgefülltes Feld einmal leer bliebe, wie der Anschluß an die nächsten Felder – mit Komma, Semikolon, Punkt oder Doppelpunkt – sein müsse. Der Aufbau eines typischen Wörterbuchartikels im DRW mußte transparent gemacht werden, bevor die Formulierung eines Druckformates vorgenommen werden konnte.

FAUST bietet bei der Gestaltung des Druckformates eine Vielzahl von Wahlmöglichkeiten an. Vielleicht wäre es möglich gewesen, das gesamte Layout mit diesen Komponenten zu entwerfen. Da wegen der Benutzung des Textsatzsystemes T<sub>E</sub>X dafür kein Bedarf bestand, ging es zunächst darum, diejenigen Steuerungszeichen auszugeben, die bisher schon für die Schriftartauszeichnungen und Ähnliches benötigt worden waren. Es würde zu weit führen, alle Schwierigkeiten und auch Irrwege bei der Lösung dieser Aufgabe aufzuzeichnen. Es möge genügen, daß ich im Zuge dieser Entwicklung das Konvertierungsprogramm, das AXEL HECKER

141 Dabei besteht eine Wahlmöglichkeit, ob die Aktualisierung des Index bei jedem neuen Eintrag – was eine erhebliche Rechenleistung erfordert – oder in gewissen Abständen erfolgen soll.

142 Damit erhält das Feld „verwendbar für“ den Status eines besonderen Hinweises des einen Bearbeiters für den anderen, ist also durch die jeweilige lexikographische Kompetenz besonders gewichtet.

143 Und ich vermute, daß dies bei Wörterbüchern mit kürzerer Bearbeitungszeit nicht viel anders ist.

seinerzeit für das DRW geschrieben hatte, durch ein KEDIT-Makro<sup>144</sup> ersetzen konnte, das zwar beinahe die Obergrenze von 500 Zeilen Quelltext erreicht, aber für mich den Vorteil besitzt, jederzeit auch ohne Kenntnis der Programmierung in C geändert und an veränderte Bedürfnisse angepaßt werden zu können.<sup>145</sup> Dabei bot sich auch die Gelegenheit, Steuerungsprozeduren zu überarbeiten und zu vereinheitlichen.

Es soll hier nur ein Beispiel für die zusätzlichen Probleme genannt werden, die der Übergang von der konventionellen Textverarbeitung in einer zeilenorientierten Datei zur Verarbeitung von Objekten der Datenbank mit sich brachte: Die Bestandteile der zukünftigen Textdatei sind zunächst unverbunden und zum Teil auch ohne die Möglichkeit einer Verbindung in der Datenbank als einzelne Objekte abgespeichert. Das heißt konkret: In einem Objekt finden sich keine Informationen, ob in einem anderen Objekt bestimmte Felder ausgefüllt sind oder nicht. Bei aufeinanderfolgenden Belegzitate aus *einer* Quelle, zum Teil auch mit derselben Datierung, wird im zweiten Beleg die Fundstelle idR. mit *ebd.* und gegebenenfalls der abweichenden Seitenzahl ersetzt. Da dieses zweite Objekt aber an einer gänzlich anderen Stelle innerhalb der Datenbank verwaltet werden kann (also nicht durch die um 1 hochgezählte Objekt Nummer identifiziert ist), läßt sich eine solche Fundstellenangabe nicht automatisch herstellen. Der Ausweg ließ sich nur so finden, daß fünf neue Felder für die unterschiedlichen Grade der Fundstellenidentität geschaffen werden mußten (gleiche Datierung, gleiche Region/Textsorte, gleiche Quelle mit und ohne unterschiedlicher Bandangabe, gleiche Seite – wenn alle diese Angaben identisch sind und in der Endredaktion die entsprechenden Felder markiert werden, so werden die entsprechenden Feldinhalte in dem zweiten Objekt durch *ebd.* ersetzt), bei unterschiedlicher Datierung wird also das Datierungsfeld vor *ebd.* ausgedruckt.

## 7 Die weitere Entwicklung der Datenverarbeitung

### 7.1 Wörterbuchparsing

Die elektronische Datenverarbeitung in der Lexikographie läßt sich – als Hilfsmittel – erst dann optimal nutzen, wenn alle lexikographischen Einheiten maschinenlesbar vorliegen, da nur dann ein Gesamtsystem mit allen möglichen Vernetzungen der lexikographischen Angaben untereinander entstehen kann. Im günstigsten Fall wird ein Wörterbuch von Anfang an unter diesem Gesichtspunkt geplant und erst als fertiges Produkt gedruckt. Im ungünstigsten Fall erfolgt die Umstellung auf elektronische Datenverarbeitung mitten im Produktionsprozeß; dann sind bereits Entscheidungen getroffen und Artikel formuliert worden, die mit den Anforderungen der

<sup>144</sup> Dieses Makro sorgt beispielsweise dafür, daß in Ausnahmefällen bestimmte Satzzeichen durch andere ersetzt werden, daß in häufig vorkommenden Buchstabenfolgen (z.B. Quellensiglen) an sinnvollen Stellen automatisch Trennmöglichkeiten für T<sub>E</sub>X eingefügt werden, aber auch dafür, daß der gesamte Wörterbuchtext mit den erforderlichen T<sub>E</sub>X-Befehlen versehen wird.

<sup>145</sup> Nur am Rande sei angemerkt, daß die Geschwindigkeit der Programmausführung, die bei sehr langen Texten eine Rolle spielen mag, hier vernachlässigbar ist: Die Bearbeitung durch das C-Programm ist nicht schneller als der Ablauf des KEDIT-Makros, wie ich mit Hilfe eines REXX-Programmes herausfinden konnte. Der Zeitaufwand für Programmabläufe sollte aber generell bei dem Einsatz der EDV nicht vernachlässigt werden: Da die unterschiedlichen Programme in der Regel immer wieder ablaufen, kann sich bei einem zu langsamen Ablauf doch ein merklicher Zeitverlust ergeben. In der Regel ist der Computer wie auch der Bearbeiter während dieses Prozesses für andere Tätigkeiten nicht einsetzbar, falls das Betriebssystem nicht mehrere Prozesse gleichzeitig bearbeiten kann, wie dies bei echtem Multitasking beispielsweise unter UNIX oder OS/2 der Fall ist.

EDV nicht vereinbar sind.<sup>146</sup> Vor allem: Ein meist erheblicher Teil des Wörterbuches liegt nicht-maschinenlesbar in gedruckter Form vor. Dies stellt verschiedene Anforderungen an die lexikographische Datenverarbeitung. Eine dieser Anforderungen ist oben bei der Beschreibung der Verweismöglichkeiten benannt worden.<sup>147</sup> Eine andere Möglichkeit besteht darin, den gedruckten Text maschinenlesbar zu machen und ihn in das Gesamtsystem zu integrieren. Dabei beginnen die Schwierigkeiten meist schon bei dem ersten Schritt: Wie kann aus der häufig sehr klein gedruckten und teilweise schon vergilbten Vorlage ein maschinenlesbarer Text hergestellt werden.<sup>148</sup> Ist dieses Problem gelöst, so stellt sich eine weit schwierigere Aufgabe: Das sogenannte Wörterbuch-Parsing.

Parsen bedeutet, eine maschinenlesbare Textdatei, die in sich nur durch Schriftartenkennzeichnungen uä. strukturiert ist, so aufzubereiten, daß unterschiedliche Textteile (z.B. Stichwort, Wortangaben, Bedeutungserklärungen, Belegtext, Datierung, Fundstellenangaben) voneinander getrennt und entsprechenden Feldern in Datenbanken zugewiesen werden können. In der linguistischen Terminologie heißt dies, eine implizite Struktur explizit zu machen, so daß am Ende eine vollständige Strukturbeschreibung des Textes, hier also des Wörterbuchmanuskriptes, vorliegt. Die interne formale Struktur des Wörterbuches, die nur durch bestimmte Kombinationen von Schriftartangaben und Textteilen erschlossen werden kann, wird in eine externe Struktur überführt. Wenn der Belegtext beispielsweise immer recte und der Erklärungsteil immer kursiv gedruckt wird, so kann die Umschaltung zwischen beiden Schriftarten dazu benutzt werden, um den Belegtext zu kennzeichnen: alles, was innerhalb der Steuerungszeichen für Beginn und Ende dieser Schriftart steht, muß Belegtext sein und kann dem entsprechenden Feld der Datenbank zugeordnet werden. Auch die Kombination bestimmter Satzzeichen mit Schriftartenkennzeichnungen kann Artikelpositionen eindeutig bestimmen. In der Praxis ist es allerdings sehr kompliziert, alle Informationen des nur implizit strukturierten Wörterbuchtextes so aufzuschlüsseln, daß alle möglichen Feldtypen erfaßt werden können. In dem relativ neuen Forschungszweig der Computerlinguistik wird für das Wörterbuchparsen daher zunächst die Erstellung einer Wörterbuchgrammatik, dh. eine vollständige Bestandsaufnahme der Artikelpositionen wie auch ihrer internen Strukturmerkmale gefordert. Das DRW hat als Beta-Tester eine Beta-Version<sup>149</sup> des von RALF HAUSER und ANGELIKA STORRER<sup>150</sup> entwickelten Programmes LEXPARSE testen können. Im Rahmen des Lexikographischen Kolloquiums von H. E. WIEGAND lernte ich einen seiner Studenten kennen, der eine Magisterarbeit über Datenbanken und Wörterbücher<sup>151</sup> schrieb; wir konnten ihn als wissenschaftliche Hilfskraft gewinnen und er war überwiegend damit beschäftigt, das LEXPARSE-Programm im DRW zu testen. Das Ergebnis war letztlich enttäuschend, was allerdings weder an dem Programm noch am Tester lag: Die internen Strukturen eines historischen (und historisch gewachsenen) Wörterbuchs

146 Dies bedeutet nicht, daß der Einsatz der EDV in der Lexikographie zum Verzicht auf essentielle lexikographische Angaben zwingen müßte. Aber bei der Möglichkeit, Wortartikel frei zu formulieren, wählt der Lexikograph schon um der Abwechslung willen – nicht aber aus zwingenden Gründen – unterschiedliche Formulierungen und vor allem unterschiedliche Plazierungen von Informationsteilen, wie dies bei einer strengen formalen Struktur kaum möglich ist.

147 Vgl. oben 6.5.2.

148 Vgl. dazu oben 5.4.

149 Die Beta-Version eines Programmes ist noch nicht offiziell freigegeben, weil es noch nicht korrekt läuft. Damit mögliche Fehler noch in dieser Phase erkannt werden, stellen sich Personen zur Verfügung, die genügend Erfahrung im Umgang mit derartigen Programmen und genügend schwierige Anforderungen an ein solches Programm haben, um es unter Praxisbedingungen testen zu können.

150 Vgl. HAUSER 1993.

151 Vgl. KAMMERER 1995.

sind, wie sich hier allzu deutlich gezeigt hat, in sich so wenig konsistent oder aber so kompliziert, daß eine korrekte Umsetzung dieser Strukturmerkmale in Feldkennzeichnungen nicht möglich ist. Als ein Beispiel möge die Genusangabe dienen: Im Regelfall genügen die Angaben *m., f. oder n.* Auch eine Kombination dieser Angaben ließe sich in wechselnder Reihenfolge hinreichend identifizieren. Anders ist es allerdings, wenn Sonderfälle in unterschiedlichster Art beschrieben werden müssen: *m., im Ae. manchmal auch f., afries. n.*, was besagt, daß ein Substantiv maskulin ist, daß es aber im Altenglischen auch Femininum und im Altfriesischen Neutrum sein kann.

In dem Maße, wie ein Wörterbuch unterschiedlich formulierte und angeordnete Positionen aufweist, wird eine formale Strukturierung schwerer, wenn nicht sogar unmöglich: eine Wörterbuchgrammatik im oben genannten Sinn, die auch für fach- und wörterbuchfremde Programmierer eines Parsers verständlich wäre, kann mit einem noch vertretbaren Zeitaufwand nicht formuliert werden. Die Konsequenz aus dieser Erfahrung konnte es dennoch nicht sein, den Versuch eines Imports fertiger Wortartikel des DRW in FAUST aufzugeben. Es mußte freilich ein Weg eingeschlagen werden, der für die besonderen Verhältnisse des DRW besser geeignet war als ein Parsing-Programm, das für eine Vielzahl von Wörterbuchtypen bestimmt war. Auch hier gilt, daß in einem solchen Fall die intime Kenntnis eines Wörterbuchs, verbunden mit einer gewissen Entscheidungskompetenz, die beste Voraussetzung dafür ist, ein solches Problem anzugehen. Denn dies bedeutet, daß auf die Formulierung einer expliziten Grammatik, die für fachfremde Personen verständlich ist, verzichtet werden kann. Auf Grund der Kenntnis des Wörterbuchs können die Zuordnungen notfalls heuristisch und ohne Überführung in eine explizite Bedingungsformulierung vorgenommen werden oder sogar Inkonsistenzen im Quelltext stillschweigend bereinigt werden.<sup>152</sup> Wenn dann noch eine Motivation hinzukommt, die zur Einhaltung bestimmter Zeitgrenzen zwingt, so sind dies gute Voraussetzungen für ein Gelingen eines solchen Vorhabens unter den Bedingungen der kontinuierlichen Produktion. Diese Motivation lag darin, daß ein Teil der für Band IX Heft 5/6 bestimmten Wortartikel bereits in FAUST geschrieben war. Der weit größere Teil freilich lag als ASCII-Datei vor. Innerhalb dieses (historisch gesehen: konventionell hergestellten) Teiles stand die Strecke der Wortableitungen mit *miß-*, die aus ca. 290 Wortartikeln bestand und innerhalb derer die Bearbeiterin eine sehr komplexe Verweisstruktur aufgebaut hatte. Die konventionelle Überprüfung dieser Verweisstruktur hätte nach Abschluß der Endredaktion mir obgelegen, so daß ich versuchte, mir diese undankbare Aufgabe durch die Übertragung der Wortartikel in FAUST zu erleichtern. Bei diesen Wortartikeln, die ohnehin noch der endredaktionellen Bearbeitung bedurften, bestand die Möglichkeit, dabei auf eine größere Einheitlichkeit der Formalien zu achten und dadurch einen Text zu gewinnen, der als Parsing-Vorlage geeignet war. Die Programmiermöglichkeit des Texteditors KEDIT konnte ich nutzen, um eine Reihe von Makros zu schreiben,<sup>153</sup> die eine so weitgehende Separierung der Wörterbuchpositionen ermöglichen, daß ein Import in FAUST denkbar erschien.

Es bedurfte dennoch vieler Experimente, um schließlich ein Ergebnis zu erreichen, bei dem

152 Daß dies letztlich doch eine implizite Wörterbuchgrammatik darstellt, ist selbstverständlich. Es fehlt aber der zeitraubende Transfervorgang vom Lexikographen zum Programmierer oder Computerlinguisten.

153 Es sind insgesamt 13 Makros, die wegen bestimmter Informationen, die teilweise erst innerhalb des Makroablaufes erzeugt werden, nicht zusammengefaßt werden können, sondern nacheinander ablaufen müssen. Sie umfassen etwa 1.800 Zeilen KEDIT-Befehle, was allerdings mehr gegen meine Programmierfähigkeiten spricht: Sicherlich wäre es möglich gewesen, vieles eleganter und einfacher zu programmieren, dies hätte aber wiederum einen erheblich höheren Zeitaufwand bedeutet und damit zu einer Verzögerung der Produktion geführt.

in den importierten Objekten nur noch wenig *händisch* verändert werden mußte. Eine Nachbearbeitung ist freilich in jedem Fall erforderlich, da die Objekte innerhalb der Datenbank durch Referenzen miteinander verbunden sind, die nur teilweise durch FAUST-interne Makroabläufe hergestellt werden können. Bei dieser Nachbearbeitung wie auch bei der redaktionellen Bearbeitung können die noch übriggebliebenen Fehler relativ leicht bereinigt werden.<sup>154</sup> Auf diesem Wege sind inzwischen einige hundert Wortartikel in FAUST importiert worden; bei jeder neuen importierten Wortstrecke können die KEDIT-Makros verbessert werden, so daß immer weniger Fehler korrigiert werden müssen.

## 7.2 Autarkie der Produktion

### 7.2.1 Von der Textdatei zum Druck

Die Herstellung einer Wörterbuchlieferung erfolgt in der Endphase konventionell. Durch den Druck von belichteten Filmen (anstelle von Bleisatz). Vor dieser Endphase gibt es allerdings noch Möglichkeiten, den Einsatz der EDV weiter auszubauen. Bislang wird der Textsatz zwar durch T<sub>E</sub>X hergestellt, aus historischen und ästhetischen Gründen aber nicht mit den Computer Modern Fonts, die DONALD E. KNUTH für sein Programm gestaltet hat. Vielmehr wird der Umbruch durch den Einsatz von besonderen Fonts so gerechnet, daß die Zeilenlänge derjenigen entspricht, die bei der Verwendung der Monotype Times Roman in der endgültigen Belichtung erzeugt wird. Dies bedeutet aber, daß wir darauf angewiesen sind, die Belichtung gerade bei dieser Firma vornehmen zu lassen, die diesen Service anbietet – eine Abhängigkeit, die spätestens dann problematisch wird, wenn man in besonderen Fällen Unterstützung benötigt und nicht erhält.<sup>155</sup> Die Konsequenz aus dieser Sachlage ist für mich eindeutig: Der Weg bis zur völlig autonomen Herstellung des Wörterbuchs muß so weit als möglich gegangen werden, um angesichts der immer komplizierter und vielfältiger werdenden Produktionskomponenten den Überblick und die Kontrolle selbst behalten zu können und nicht irgendwann einmal vor der Situation zu stehen, daß ein winziger Teil dieses Vorganges ausgefallen ist oder nicht mehr zur Verfügung steht und damit die Produktion angehalten werden muß. Also: Die Abhängigkeit von einer bestimmten Druckerei, die die Umwandlung der metrischen Fonts in Monotype Times Roman monopolisiert hat, muß aufgehoben werden. In der Praxis ist dies dadurch möglich, daß die auf dem Computermarkt zur Verfügung stehenden professionellen Schriftarten genutzt werden können. In Postscript, einer Druckerbeschreibungssprache, die auf sehr vielen Druckern und Lichtsatanlagen gelesen werden kann, gibt es die originalen Monotype-Fonts. Wie aber kann ich diese Fonts in T<sub>E</sub>X nutzbar machen? Es gibt ein Konvertierungsprogramm, das genau dies leisten soll: Die T<sub>E</sub>X-Ausgabedateien in PostScript-Ausgabedateien umzuwandeln. Freilich war dieses Programm für PC's mit dem Betriebssystem DOS noch nicht fertiggestellt; lediglich für OS/2 und UNIX bzw. LINUX gab es entsprechende Treiber. Nachdem auf meine Bitte das URZ sich bereiterklärt hatte, auch eine PC-Version des DVIPS-Treibers herzustellen, konnte mit dem Experimentieren begonnen werden. In dem Zeitpunkt, in dem ich

154 Es handelt sich dabei meist darum, daß bestimmte Zeichenfolgen innerhalb der Belegtexte auch in Fundstellenangaben vorkommen und dort zur Feldaufteilung genutzt werden. Da diese Zeichenfolgen nicht immer beim Parsen voneinander unterschieden werden können, treten in Einzelfällen Fehlinterpretationen auf. Ein Ausweg aus diesem Dilemma könnte es sein, die Fundstellenangaben als einheitlichen Block zu definieren (z.B. als *einen* String, der erst am Schluß des Parsings durch Wiedereinfügen von Leerzeichen in verschiedene Positionsangaben aufgelöst würde.

155 So stimmte bei dem Einsatz eines neuen Druckertreibers für T<sub>E</sub>X, der wegen einer neuen Version von T<sub>E</sub>X benötigt wurde, plötzlich die Skalierung eines Sonderzeichens nicht mehr. Der Versuch, dieses Problem bei der Druckerei lösen zu lassen, blieb ergebnislos.



diesen Aufsatz abschließen,<sup>156</sup> sind noch einige Probleme zu lösen – an welcher Stelle der Postscript-Fonts sind die Sonderzeichen versteckt, die das DRW benötigt, in welchen besonderen Fonts gibt es Zeichen, die nicht zum normalen Lieferumfang gehören: Die Liste ließe sich verlängern. Ich bin mir allerdings sicher, daß spätestens die nächste Lieferung des DRW auf diese Weise produziert werden wird, möglicherweise auch schon so, daß die Belichtung am Rechenzentrum zum Selbstkostenpreis erfolgen kann. Damit wäre ein Weg bis zu Ende gegangen worden: Der Weg vom Ersatz der Schreibmaschine durch den Computer bis zur völlig autarken Herstellung der Vorlage für den Offset-Druck.

## 7.2.2 Ausblicke

Welcher Weg aber ist mit Erreichen *eines* Zieles schon wirklich an seinem Ende angelangt? Auch die Ausgabe der EDV-Version des Wörterbuches im gedruckten Medium muß nicht das Endstadium darstellen. Denkbar ist es auch, eine maschinenlesbare Fassung des Wörterbuches mit der entsprechenden Software<sup>157</sup> neben oder anstelle des gedruckten Buches zu vertreiben. Und hier wiederum gibt es verschiedene Möglichkeiten wie die Pressung von CD-ROMs, aber auch die Verfügbarhaltung des Wörterbuchs in kommerziellen oder wissenschaftlichen Daten-netzen wie etwa INTERNET. Dies allerdings sind Fragen und Möglichkeiten, die erhebliche Probleme rechtlicher und technischer Natur mit sich bringen würden – vorrangig ist zunächst die Sicherung und der Ausbau des bisher erreichten Standards der Datenverarbeitung wie auch die Nutzung dieser Möglichkeiten nicht nur für eine Steigerung der Qualität des Wörterbuchs, sondern auch für eine Beschleunigung der Produktion.

## 8 Schlußbetrachtung

Bei dem Schreiben dieses Aufsatzes habe ich alle Schritte rekapitulieren müssen, die auf dem Weg zur Computerisierung des Deutschen Rechtswörterbuches erforderlich waren. Nutzen und Nachteil dieser Entwicklung stellten sich als Frage nochmals neu. Das Fazit ist nicht nur positiv: Die elektronische Datenverarbeitung hat die Wörterbucharbeit auf eine völlig neue Grundlage gestellt, hinter die nicht mehr zurückgegangen werden kann. Die Vielfalt von Informationsmöglichkeiten, die Bequemlichkeit der Texteingabe und -verarbeitung, der Wegfall der verschiedenen Korrekturgänge beim konventionellen Satz sind Fortschritte, die zu deutlich ins Auge fallen, als daß ein Zweifel an dem Nutzen der EDV für die Wörterbucharbeit möglich wäre. Es sind Fortschritte, die einerseits dem Qualitätsstandard des Wörterbuchs zugute kommen, die andererseits aber primär dem Verlag eine erhebliche finanzielle Entlastung bringen, ohne daß dies dem Wörterbuch anders als vielleicht durch eine gewisse Konstanz der Preisgestaltung oder durch den Wegfall von Druckkostenzuschüssen vergolten würde. Allein – der Preis, den wir innerhalb des Wörterbuchs für diesen Nutzen zahlen müssen, ist hoch. Der Arbeitsaufwand für das Erlernen des Umganges mit der Computertechnik und die Anstrengungen, die es erforderte, immer neue Programme auf ihren Nutzen für das DRW zu untersuchen und gegebenenfalls in das Gesamtsystem zu integrieren, waren erheblich. Dies alles war nicht im Rahmen der üblichen Arbeitszeiten zu leisten. Eine effektive Umstellung erfordert eine hohe persönliche Motivation, gerade weil eine Delegation dieser Arbeiten nicht möglich ist. Aus

<sup>156</sup> Mitte Oktober 1994.

<sup>157</sup> Wobei dies nicht unbedingt die zur Produktion verwendete Datenbank sein muß. Da alle Felder und Feldinhalte aus FAUST exportierbar sind, ließe sich auch für diesen Zweck der Einsatz anderer Programme denken. Der Dudenverlag benutzt beispielsweise für seine maschinenlesbaren Wörterbücher ein auf WINDOWS basierendes Retrievalsystem.

der ganz individuellen Erfahrung bei der Computerisierung des Deutschen Rechtswörterbuchs wage ich den Satz, daß die Verbindung zwischen der Leitung des Projektes und der Einführung der EDV die entscheidende Bedeutung für den Erfolg dieses Prozesses besaß. Bei einer Verteilung der Kompetenzen auf mehrere Personen wäre der Zeitablauf allein durch das Erfordernis des ständigen Informationsabgleiches erheblich verlängert worden. Darüberhinaus schützt nur die Kenntnis der Wörterbuchstruktur einerseits und die Kenntnis der in einem noch vertretbaren Zeitrahmen zu nutzenden Möglichkeiten der EDV davor, die Erwartungen an den Computereinsatz zu hoch anzusetzen oder aber die Struktur des Wörterbuchs unter dem Aspekt der EDV-Technik einzuengen. Der letzteren Gefahr, die in Gesprächen unter Lexikographen eine vielleicht zu große Rolle spielt, ist das DRW zumindest nicht erlegen. Die einzige Konzession an die Erfordernisse der Datenbanktechnik lag darin, daß an einer Stelle das Semikolon als Informationsrenner durch den sowieso sinnvolleren Punkt ersetzt werden mußte. Eine weitere Gefahr bei der Übertragung der EDV-Arbeiten an Computerspezialisten besteht wohl darin, daß sich unvermerkt deren Vorstellungen von der sinnvollen Gliederung der Wörterbucharbeit in das Projekt einschleichen können – bis am Schluß eine perfekte Datenverarbeitung mit allen nur denkbaren Nutzungsmöglichkeiten für ein Wörterbuch entwickelt worden ist, das es nur als Chimäre im Kopf der Programmierer gibt. Das Gesamtprojekt kann in diesem Fall den Händen der wissenschaftlichen Leitung entgleiten. Die dritte Gefahr schließlich besteht in der Verzögerung des Projektes oder dem Produktionsstillstand bis zu dem Zeitpunkt, in dem die Datenverarbeitung vollständig eingesetzt werden kann. Daß das DRW dieser Gefahr weitgehend ausweichen konnte, liegt meiner Ansicht nach darin, daß wir die EDV von Anfang an nur zur Erleichterung der *Produktion* einführen wollten. Als erstes wurde die Schreibmaschine bei der Erstellung der Druckvorlage durch die EDV ersetzt, dann der maschinelle Bleisatz, schließlich das Handwerkszeug auf dem Schreibtisch der Mitarbeiter.<sup>158</sup> Ein letztes Problem liegt freilich darin, daß die Arbeit mit einem solchen Werkzeug, das viele Möglichkeiten bietet, aber für seine Nutzung auch beherrscht werden will, nicht jedem Wissenschaftler sympathisch (und auch vielleicht: angemessen) ist. Ein Mittelweg zwischen den liebgewordenen Denkgewohnheiten (die sich verständlicherweise auch als individuelle Notwendigkeiten präsentieren lassen) und den nüchternen Anforderungen der Datenbanktechnik und -struktur an das wissenschaftliche Arbeiten läßt sich manchmal nicht leicht finden. Was der eine als Zwang empfinden mag, kann für den anderen eine begrüßenswerte Verpflichtung zu einem strukturierteren Denken darstellen. In den Diskussionen um die Auswirkungen der EDV im DRW wurde zweierlei deutlich: Wissenschaftliche Kreativität läßt sich durch ein solches Instrumentarium nicht einengen, und man sollte nicht irgendwann der Illusion erliegen, in den Strukturen, Feldern und Informationsmöglichkeiten dieser Datenbank das je erreichbare Wissen präsent zu haben. Die Vermittlung zwischen solchen Positionen gehört zu den Voraussetzungen eines wirkungsvollen und von den Betroffenen akzeptierten Einsatzes der elektronischen Datenverarbeitung in einem traditionsbeladenen Wörterbuch.

158 Dennoch kommt es zum gegenwärtigen Zeitpunkt zu einer Verzögerung der Publizierung des nächsten Doppelheftes (Band IX 5/6) um etwa zwei Monate: Der Import der bereits in ASCII-Dateien vorliegenden Wortartikel in FAUST war einerseits schneller möglich geworden als erwartet, hatte sich aber im konkreten Vollzug doch als schwierig erwiesen, so daß ich selbst den Zeitplan nur unter Verzicht auf die innerhalb FAUSTs mögliche Abgleichung der Verweisstruktur hätte einhalten können. Zugleich damit war es möglich geworden, bestimmte Informationen innerhalb der Datenbank zu nutzen, um die lebenden Kolummentitel so zu gestalten, daß nunmehr nicht nur die Lemmata genannt werden, sondern überdies noch die jeweilige Bedeutung des Wortartikels, mit der eine Seite beginnt oder endet: *Mordwerk – Morgen (III 1)*. Bis dies allerdings fehlerfrei lief, gab es immer wieder unvorhergesehene Schwierigkeiten zu überwinden, die erst jetzt den Abschluß des Umbruchs ermöglichen. Hier schien es mir gerechtfertigt zu sein, wegen der dadurch erreichbaren Qualitätsverbesserung diese Verzögerung in Kauf zu nehmen.

Schließlich: Die eigentliche lexikographische Arbeit wird durch den Einsatz der EDV zwar erleichtert und vereinfacht, aber das, was den erfahrenen Lexikographen ausmacht, wird dadurch nicht ersetzt. Und dies ist das im langjährigen Umgang mit Quellen erworbene Augenmerk, das Gespür für Unstimmigkeiten und Gefahren, das einen Klippen umschiffen läßt, an denen andere vielleicht gescheitert sind, es ist dies eine umfassende Textkenntnis und es sind geschichtliche Detailkenntnisse der Kultur-, Rechts- und Sachgeschichte, die – verbunden mit einer hohen hermeneutischen Sensibilität – erst den guten Lexikographen ausmachen. Für diesen Lexikographen, der zugleich Handwerker und Wissenschaftler ist, stellt die elektronische Datenverarbeitung ein hervorragendes Werkzeug dar – aber auch nicht mehr.

## 9 Literaturverzeichnis

- BAHR 1962 = JOACHIM BAHR, Zur Neubearbeitung des DEUTSCHEN WÖRTERBUCHS. Von der Exzerption zur Elekton. In: Zeitschrift für deutsche Wortforschung 18. 1962, 141 – 150.
- BLUMENTHAL/LEMNITZER/STORRER 1987 = ANDREAS BLUMENTHAL, LOTHAR LEMNITZER, ANGELIKA STORRER, Was ist eigentlich ein Verweis? Konzeptionelle Datenmodellierung als Voraussetzung computergestützter Verweisbehandlung. In: Das Wörterbuch. Artikel und Verweisstrukturen. Jahrbuch 1987 des Instituts für deutsche Sprache. Hrsg. GISELA HARRAS. 1988. 351 – 373.
- BUMA/EBEL 1963ff. = WYBREN JAN BUMA, WILHELM EBEL, Altfriesische Rechtsquellen. Texte und Übersetzungen. 1 – 8. Göttingen 1963ff.
- COWLISHAW 1988 = MIKE F. COWLISHAW, Die Programmiersprache REXX, [dt. Ausgabe von BALDUR KROHN]. München, Wien, London 1988.
- DICKEL/SPEER 1979 = GÜNTHER DICKEL und HEINO SPEER, DEUTSCHES RECHTSWÖRTERBUCH. Konzeption und lexikographische Praxis während acht Jahrzehnten (1897 – 1977), in: Praxis der Lexikographie: Berichte aus der Werkstatt, hrsg. von HELMUT HENNE, Tübingen 1979 (= Reihe germanistische Linguistik 22), 20 – 36.
- DRW = DEUTSCHES RECHTSWÖRTERBUCH (Wörterbuch der älteren deutschen Rechtssprache), herausgegeben (bis Bd. IV) von der Preußischen (später der Deutschen) Akademie der Wissenschaften und (ab Bd. V) der Heidelberger Akademie der Wissenschaften in Verbindung mit der Akademie der Wissenschaften der DDR (bis 1992), Bd. I Weimar (1914 – 1932), zuletzt erschienen Bd. IX Heft 5/6 Weimar 1994.
- DUDEN 1988 = DUDEN INFORMATIK. Mannheim, Wien, Zürich 1988.
- GOEBEL/LEMBERG/REICHMANN 1995 = ULRICH GOEBEL, INGRID LEMBERG, OSKAR REICHMANN, Versteckte lexikographische Informationen. Möglichkeiten ihrer Erschließung dargestellt am Beispiel des FRÜHNEUHOCHDEUTSCHEN WÖRTERBUCHES. Erscheint 1995.
- DWB = JACOB und WILHELM GRIMM, DEUTSCHES WÖRTERBUCH, Bd. I, Leipzig 1854, bis Bd. 33, 1971
- EBEL 1974 = FRIEDRICH EBEL, Über Legaldefinitionen. Rechtshistorische Studie zur Entwicklung der Gesetzgebungstechnik in Deutschland, insbesondere über das Verhältnis von Rechtsetzung und Rechtsdarstellung. Berlin 1974 (= Schriften zur Rechtsgeschichte 6).
- FAUST = FAUST. Handbuch Bd. I u. II, DORIS LAND Software-Entwicklung. Postfach 1126. D-8507 Oberasbach. 1991.
- FLINK 1991 = KLAUS FLINK (Hrsg.), Das Stadtrecht von Cleve. (Klever Archiv 11). Kleve 1991.
- FWB = FRÜHNEUHOCHDEUTSCHES WÖRTERBUCH. Hrsg. v. ROBERT R. ANDERSON, ULRICH GOEBEL. OSKAR REICHMANN. Band I bearb. v. OSKAR REICHMANN. Berlin-New York 1986ff.
- GULBINS/KAHRMANN 1993 = JÜRGEN GULBINS, CHRISTINE KAHRMANN, Mut zur Typographie. Ein Kurs für DTP und Textverarbeitung. Berlin, Heidelberg, New York... 1993.
- HAUSER 1993 = RALF HAUSER, LEXPARSE Version 0.951. A Dictionary Entry Parser. User's Manual. Tübingen 1993.
- KAMMERER 1995 = MATTHIAS KAMMERER, Bildschirmorientiertes Abfassen von Wörterbuchartikeln. Dargestellt am Beispiel des FRÜHNEUHOCHDEUTSCHEN WÖRTERBUCHES. – Typoskript. Magisterarbeit Heidelberg 1994. Druck in: Lexicographica Series Maior. Tübingen: Niemeyer (voraussichtlich 1995).
- KEES 1993 = PETER KEES, REXX in der Praxis. München-Wien 1993.
- KNUTH 1987 = DONALD E. KNUTH, The T<sub>E</sub>X book. Addison-Wesley 1987.

- LAUFS 1993 = ADOLF LAUFS. Das DEUTSCHE RECHTSWÖRTERBUCH. In: Akademie-Journal. Mitteilungsblatt der Konferenz der Deutschen Akademien der Wissenschaften. 2. 1993, 7 – 11.
- MIHM/ELMENTHALER 1990 = AREND MIHM, MICHAEL ELMENTHALER. Das Duisburger Stadtrecht 1518 Duisburg 1990.
- MIHM 1994 = MARGRET MIHM (Hrsg.). Die Protokolle des Duisburger Notgerichts 1537 – 1545 (Duisburger Geschichtsquellen 10). Duisburg 1994.
- MISGELD 1991 = WOLFGANG D. MISGELD. ORACLE für Profis. München-Wien 1991.
- REICHMANN 1990 a = OSKAR REICHMANN. Formen und Probleme der Datenerhebung I: Synchronische und diachronische historische Wörterbücher. In: Wörterbücher. Ein internationales Handbuch zur Lexikographie. Hrsg. v. FRANZ JOSEF HAUSMANN, OSKAR REICHMANN, HERBERT ERNST WIEGAND, LADISLAV ZGUSTA. II. 1990 (Handbücher zur Sprach- und Kommunikationswissenschaft 5.2) 1588 – 1611.
- REICHMANN 1990 b = OSKAR REICHMANN. Philologische Entscheidungen bei der Formulierung von Artikeln historischer Sprachstadienwörterbücher. In: Historical Lexicography of the German Language. Ed. by ULRICH GOEBEL and OSKAR REICHMANN in collaboration with PETER I. BARTA. I. Lewiston-Queenston-Lampeter 1990 (Studies in German Language and Literature I) 231 – 278.
- SPEER 1989 = HEINO SPEER. Das DEUTSCHE RECHTSWÖRTERBUCH. Historische Lexikographie einer Fachsprache. In: Lexicographica 5 (1989) 85 – 128.
- SPEER 1991 = HEINO SPEER. Das DEUTSCHE RECHTSWÖRTERBUCH: Vorstellung des Wörterbuchs und lexikographische Praxis am Beispiel „magdeburgisch“. In: Historical Lexicography of the German Language. Ed. by ULRICH GOEBEL and OSKAR REICHMANN in collaboration with PETER I. BARTA. II. Lewiston-Queenston-Lampeter 1991 (Studies in German Language and Literature VI) 675 – 711.
- SCHWARZ 1987 = NORBERT SCHWARZ, Einführung in T<sub>E</sub>X. Addison-Wesley 1987.
- STEER 1987 ff. = GEORG STEER u.a. (Hrsg.). Die Rechtssumme Bruder Bertholds. Eine deutsche abecedarische Bearbeitung der Summa Confessorum des Johannes von Freiburg. Synoptische Edition der Fassungen B, A und C. Iff. Tübingen 1987 ff.
- WALSDORFF 1987 = JOACHIM WALSDORFF, Kurze Einführung in WCSRIPT. Waterloo SCRIPT – Version 86.1 (87MAR31). Universitätsrechenzentrum Heidelberg. Im Neuenheimer Feld 293. D-6900 Heidelberg. 2. überarbeitete Auflage August 1987.
- WEGNER 1989 = IMMO WEGNER, Lexikographische Definition und Frame-Theorie im allgemeinen einsprachigen Wörterbuch. In: Wörterbücher. Ein internationales Handbuch zur Lexikographie. Hrsg. v. FRANZ JOSEF HAUSMANN, OSKAR REICHMANN, HERBERT ERNST WIEGAND, LADISLAV ZGUSTA. I. 1989 (Handbücher zur Sprach- und Kommunikationswissenschaft 5.1) 893 – 899.
- WONNEBERGER 1987 = REINHARD WONNEBERGER, L<sup>A</sup>T<sub>E</sub>X. 1987. (Addison-Wesley Kompaktführer).

*Dr. Heino Speer, Deutsches Rechtswörterbuch, Forschungsstelle der Heidelberger Akademie der Wissenschaften, Karlstr. 4, D-69117 Heidelberg, e-mail: Speer@drw.adw.uni-heidelberg.de*