

Elektronische Materialgrundlage und computergestützte Ausarbeitung eines historischen Belegwörterbuchs. Erfahrungen und Perspektiven am Beispiel des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS

- | | | | |
|-----|--|-------|---|
| 1 | Das Vorhaben und sein EDV-Konzept | 6 | Konzeptionelle Fragen der Publikation des Wörterbuchs für elektronische Nutzung |
| 2 | Das Textarchiv | 6.1 | Das Verhältnis von Wörterbuch und Wörterbuchbasis |
| 3 | Die Lemmatisierungskomponente | 6.2 | Benutzung des Wörterbuchs über Artikelstichwörter, über Register und über registerähnliche Hilfen |
| 4 | Das Belegarchiv | 6.2.1 | Register |
| 5 | Das Programmsystem für die Artikelarbeit | 6.2.2 | Volltextsuche und Suche in Feldern |
| 5.1 | Leistungsmerkmale des Artikelredigierprogramms | 7 | Resümee |
| 5.2 | Mehrwert der Artikelarbeitsdateien gegenüber den Artikeln im publizierten Wörterbuch | 8 | Literatur |

1 Das Vorhaben und sein EDV-Konzept

Das Vorhaben eines neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS ist im Gesamtzusammenhang der historischen Beleglexikographie des Deutschen zu sehen. Für das Althochdeutsche wie für das Frühneuhochdeutsche sind umfangreiche Belegwörterbücher im Erscheinen begriffen, die das Grimmsche Wörterbuch und seine ebenfalls in Bearbeitung befindliche Teilerneuerung (der Strecke A–F) als Epochewörterbücher für die genannten älteren Sprachstufen ergänzen.¹ Für das Mittelhochdeutsche in den zeitlichen Grenzen von etwa 1050 bis 1350 ist man dagegen bislang noch auf die Pionierarbeiten des 19. Jahrhunderts angewiesen, deren historische Verdienste ebenso unbestritten sind wie ihre Unzulänglichkeit angesichts dessen, was heute möglich ist. Dies sind das von Georg Friedrich Benecke (1762–1844) konzipierte und von Wilhelm Müller und Friedrich Zarncke in den Jahren 1854 bis 1866 in vier Bänden ausgearbeitete MITTELHOCHDEUTSCHE WÖRTERBUCH (= BMZ) und das – auf dem BMZ beruhende und ihn vor allem für das späte 14. und das 15. Jahrhundert reichlich ergänzende – dreibändige MITTELHOCHDEUTSCHE HANDWÖRTERBUCH von Matthias Lexer (= LEXER), das 1872 bis 1878 erschien. Im vollen Wortsinne Belegwörterbuch zu nennen ist der BMZ, seine Belegbasis ist aber im Verhältnis zum heutigen Stand der Quellenerschließung verhältnismäßig schmal, und seine Anlage als Stamm- bzw. Wortfamilienwörterbuch legte eine starke Konzentration der Belegsammlung und -darbietung auf die Stammwörter nahe. Lexer hat die Belegbasis gegenüber dem BMZ

¹ Vgl. im Literaturverzeichnis unter AWB, ¹DWB, ²DWB und FWB.

zwar erheblich erweitert, doch beschränkt er sich in großem Umfang auf Belegstellennachweise und verzichtet auf Belegzitate.

Die „zellen“, meinte Lexer 1878 im Vorwort zum dritten Band seines Werks (S. IV), „sind durch die beiden mhd. Wörterbücher gebaut und das weitere eintragen in dieselben wird eine verhältnismässig leichte arbeit der nachsammelnden sein“. Das auf Lexers „Zellen“ bezogene Nachsammeln geschah vor allem in Glossaren, Wortverzeichnissen und Registern zu Editionen und lexikographischen Untersuchungen mittelhochdeutscher Texte. Über 100 dieser Nachsammlungen sind kompiliert zu einem Gesamtverzeichnis, dem in Trier ausgearbeiteten ‚Findebuch zum mittelhochdeutschen Wortschatz‘ (1992). Das FINDEBUCH verweist in rund 35.000 Artikeln auf knapp 140.000 Glossarstichwörter, die ihrerseits in der Regel jeweils zu einer Vielzahl von Textbelegen führen.

Mit dem FINDEBUCH waren die Grenzen dessen, was auch dem kundigsten Benutzer an Eigenarbeit bei der Kombination verschiedener Wörterbücher und lexikographischer Hilfsmittel für das Mittelhochdeutsche zuzumuten ist, deutlich erreicht. Den gerade auch angesichts der sonst zu beobachtenden Fortschritte der historischen Beleglexikographie des Deutschen zunehmend als „unerträglich“ empfundenen „Zustände[n] auf dem Gebiet der mittelhochdeutschen Lexikographie“ (Stackmann 1990: V) war mit weiteren Anlagerungen an die alten Wörterbücher – sei es in der Form von Supplementen wie dem FINDEBUCH, sei es als Satellitenkranz von Spezialwörterbüchern – nicht mehr beizukommen, sie verlangten vielmehr zwingend einen Neuansatz in der Erfassung des mittelhochdeutschen Wortschatzes und der Beschreibung seines Gebrauchs, ein Wörterbuch also, das selbständig aus den Quellen zu erarbeiten ist.

Das neue Wörterbuch wird seit 1994 in zwei Arbeitsstellen an den Universitäten Göttingen und Trier mit Unterstützung der Deutschen Forschungsgemeinschaft und der Göttinger und Mainzer Akademien der Wissenschaften vorbereitet; zum 1.1.2000 sind die beiden Arbeitsstellen in das Akademieprogramm überführt worden. Die Vorbereitungsphase, die kurz vor dem Abschluß steht, gilt im wesentlichen der Materialbereitstellung, ferner der Diskussion und Erprobung konzeptioneller Grundsätze. Die Ausarbeitung des Werks, das auf insgesamt vier Bände zu je etwa 1000 Seiten berechnet ist, wird voraussichtlich 2001 beginnen und soll in einem Zeitraum von 20 Jahren abgeschlossen werden.

Als das jüngste der größeren Vorhaben zur historischen Beleglexikographie des Deutschen konnte das neue MITTELHOCHDEUTSCHE WÖRTERBUCH von vornherein ganz auf EDV-Basis gestellt werden. Zugute kamen den auf die EDV-Seite bezogenen Planungen die langjährigen Erfahrungen von Kurt Gärtner, dem Trierer Projektleiter, auf dem Gebiet des philologischen EDV-Einsatzes, die bis in die Mitte der 80er Jahre zurückreichenden computergestützten lexikographischen Vorarbeiten der Trierer Arbeitsstelle und nicht zuletzt der Beistand von Paul Sappler (Tübingen), der das für sein Vorhaben eines Wörterbuchs zu Gottfrieds ‚Tristan‘ entwickelte Programmsystem² zur Verfügung stellte und mit Rat und Tat bei der Anpassung und dem Ausbau für die Zwecke des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS behilflich war;³ damit war gewährleistet, daß im Falle des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS die Sprachlexikographie von den avancierten Methoden und Techniken der Textlexikographie profitieren kann.⁴

² Vgl. Sappler/Schneider-Lastin 1991.

³ Vgl. Recker/Sappler 1998.

⁴ Wie „Methoden und Techniken der Textlexikographie auf die Sprachlexikographie zu übertragen“ wären, diskutiert Sappler 1991 (Zitat S. 277).

Das EDV-Konzept des Vorhabens ist bereits mehrfach dargestellt worden unter Einschluß von Details der Arbeitsabläufe und ihrer technischen Realisierung.⁵ Hier genügt daher eine kurze einleitende Rekapitulation der Grundzüge, die zum Verständnis des Zusammenspiels der einzelnen Programmkomponenten und Materialzustände nötig ist; in Bezug auf diese sollen dann anschließend dem Thema des Bandes entsprechend die „Chancen und Perspektiven“, wie sie sich uns nach fünfjähriger Arbeit darstellen, erörtert werden.

Materialbereitstellung für ein historisches Belegwörterbuch heißt traditionell in erster Linie: Exzerption von Belegstellen aus den gedruckten Quellentexten, Verzettelung der Exzerpte, Ordnung der Belegzettel nach Artikelstichwörtern (Lemmatisierung).⁶ Die elektronische Materialbereitstellung für das mittelhochdeutsche Wörterbuch lehnt sich an das traditionelle Verfahren an, geht aber in einigen Punkten anders vor; sie erfordert zusätzliche Arbeitsgänge, spart dafür andere ein und modifiziert insgesamt die Abfolge.

Zusätzliche Arbeitsgänge entstehen daraus, daß die Materialgrundlage in elektronischer Form vorliegen und für die weitere elektronische Verarbeitung eingerichtet werden muß. Materialgrundlage sind im Falle des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS überwiegend Volltexte, nur ein kleinerer Teil des Belegarchivs wird im Zuge der Ausarbeitung der Artikel durch Einzelerfassung von Belegen aus weiteren Texten erhoben werden. Das elektronische Textarchiv ist die erste Komponente der elektronischen Wörterbuchbasis.

Die weiteren Arbeitsgänge, die der herkömmlichen Verzettelung und anschließenden Lemmatisierung der Belege entsprechen, können bei entsprechender Vorbereitung der elektronischen Quellentexte automatisch bzw. halbautomatisch erledigt werden. Dafür wird – neben einer Reihe von geeigneten Programmen⁷ – eine zweite Komponente benötigt, die die Lemmatisierungsinformation enthält, eine Art Thesaurus mit den Artikelstichwörtern und den ihnen zuzuordnenden Wortformen; diese Komponente heißt projektintern das Gerüst. Das Gerüst enthält eine auf der Grundlage der alten Wörterbücher und des FINDEBUCHS erarbeitete Stichwortliste mit insgesamt rund 80.000 Lemmata und die ihnen aus den bereits für das Belegarchiv bearbeiteten Texten zugewiesenen Wortformen.

Aus den Texten kann mit der Hilfe des Gerüsts eine lemmatisierte KWIC-Konkordanz (mit frei wählbarer Kontextlänge) erzeugt werden, der im herkömmlichen Verfahren der lemmatisierte Zettelkasten, also das eigentliche Belegarchiv entspricht. Wenn im folgenden in Anlehnung an die traditionelle Arbeitsweise vom Belegarchiv des mittelhochdeutschen Wörterbuchs die Rede ist, so gilt es stets, sich einen grundlegenden Unterschied vor Augen zu halten: Es handelt sich beim elektronischen Analogon eigentlich nicht um eine selbständige Komponente, sondern um einen Materialzustand, der auf der Grundlage von Texten und Gerüst bei Bedarf erzeugt wird und durch Änderungen in diesen beiden Komponenten beliebig beeinflußt werden kann. In der Phase der Materialbereitstellung dient die Ausgabe von Konkordanzen vor allem der Kontrolle der halbautomatischen Lemmatisierung. Erkannte Lemmatisierungsfehler werden durch Änderungen in den Texten (z.B. durch Einfügung von Homographenmarkierungen) und/oder im Gerüst (z.B. durch Umstellen von falsch zugeordneten Wortformen) behoben, ein neuer Konkordanzlauf stellt dann die be richtigte Lemmatisierung her.

⁵ Vgl. zuletzt Plate/Recker 2000 und Sappler 2000b.

⁶ Zur Exzerption für das FWB und DRW vgl. Lemberg 1996.

⁷ Die verwendeten Programme sind auf der Basis des leistungsstarken Tübinger Systems von Textverarbeitungsprogrammen (TUSTEP) entwickelt, das erweiterbare Module für alle Anforderungen philologischer Datenverarbeitung zur Verfügung stellt.

Mit der Erzeugung des lemmatisierten Belegarchivs, das als elektronisches Analogon des herkömmlichen Zettelkastens gelten kann, ist die Arbeit der Materialbereitstellung geleistet. Es wäre nun möglich, hier den elektronischen Kreislauf zu unterbrechen, das heißt die Belege zu einem Stichwort für die Weiterarbeit in herkömmlicher Weise auf Zetteln auszugeben. Damit würden aber Vorteile verschenkt, welche die elektronische Materialgrundlage bietet. Für die elektronisch gestützte Artikelarbeit wird daher ein weiterer Materialzustand erzeugt, die Artikelarbeitsdatei, und es werden in einem Artikelredigierprogramm Funktionen zur Verfügung gestellt, die es dem Bearbeiter erlauben, die herkömmlichen Arbeitsgänge der Artikelarbeit bis hin zum Satz des Artikels elektronisch durchzuführen.

Weil der gesamte Arbeitsablauf computergestützt ist und alle Materialzustände in elektronischer Form vorliegen, ist auch die elektronische Publikation der Materialien wie des Wörterbuchs selbst ohne großen technischen Mehraufwand möglich. Die in diesem Zusammenhang zu erörternden Fragen sind daher im Falle des neuen *MITTELHOCHDEUTSCHEN WÖRTERBUCHS* in erster Linie konzeptionell lexikographischer Natur.

Nach dieser kurzen Rekapitulation des EDV-Konzepts des Vorhabens sollen nun die einzelnen Komponenten und Materialzustände im Hinblick auf ihre „Chancen und Perspektiven“ einer näheren Betrachtung unterzogen werden.

2 Das Textarchiv

Das elektronische Textarchiv besteht gegenwärtig aus den Volltexten des rund 75 Nummern (Einzeltexte und Textsammlungen unterschiedlichen Umfangs) umfassenden Grundkorpus der Wörterbuchquellen,⁸ die Einwerbung von Mitteln für eine massive systematische Aufstockung zunächst um maschinenlesbare Versionen der rund 75 weiteren FINDEBUCH-Quellen, die noch nicht im Grundkorpus berücksichtigt sind,⁹ wird zur Zeit vorbereitet.¹⁰ Eine Reihe weiterer Texte, die zum Quellenbereich des neuen *MITTELHOCHDEUTSCHEN WÖRTERBUCHS* zählen, sind aus dem Fundus der Projektbeteiligten und durch Deposita von Fachkollegen in maschinenlesbarer Fassung für das Wörterbuch zugänglich.

Angesichts der großen Zahl der Quellentexte wie des langen Zeitraums, in dem sie stets zugriffsbereit verfügbar gehalten werden müssen, ist eine sorgfältige einheitliche Einrichtung der Texte und ein softwareunabhängiges Auszeichnungs- und Archivierungssystem nötig.

Bei der Einrichtung der elektronischen Texte wird „ausgabendiplomatisch“ verfahren. Daß diese Entscheidung nicht trivial ist, wird deutlich, wenn man sich vor Augen hält, daß das Quellenspektrum des Wörterbuchs die ganze Breite von verhältnismäßig stark normalisierten Editionstexten bis hin zu auch graphisch überlieferungsgetreuen Handschriftenabdrucken umfaßt. Die Alternative zum ausgabendiplomatischen Prinzip hätte darin bestanden, mindestens leichte Normalisierungen bei den graphisch überlieferungsnahen Editionen durchzuführen, um die Abbildung ungewöhnlicher und schwierig darzustellender, lexikographisch aber meist unerheblicher graphischer Differenzierungen zu ersparen. Genauere Prüfung dieser Alternative zeigte jedoch, daß es erstens immer zahlreiche

⁸ Bibliographie im Anhang zu Gärtner (2000) und auf der Homepage des Projekts, dort auch mit einem alphabetischen Index.

⁹ Bibliographie auf der Homepage des Projekts unter „Findebuchquellen“.

Einzelfälle gibt, in denen Unklarheit über die lexikographische Relevanz von Schreibvarianten besteht; die Aufstellung von Normalisierungsregeln und die Kontrolle ihrer Durchführung hätte also einen nicht unbeträchtlichen Aufwand erfordert. Zweitens aber, und dies ist der wichtigere Gesichtspunkt, wäre diese Normalisierung über den gesamten Text einer Quelle für Wörterbuchzwecke zum größten Teil verlorene Arbeit, denn nur ein Bruchteil der Belege eines Quellentextes wird später im Wörterbuch tatsächlich zitiert.

Neben diesem materiellen Aspekt hat die ausgabendiplomatische Einrichtung der elektronischen Wörterbuchquellen auch einen formalen: Er besteht zum einen in der Abbildung der Referenz (für die automatische Erzeugung der Fundstellenangabe bei den Belegtexten), zum anderen in der Auszeichnung von editorischem Beiwerk (wie Zwischenüberschriften des Herausgebers usw.), das nicht für das Belegarchiv ausgewertet werden soll. Beides zusammen läuft auf eine genaue Analyse und entsprechende sachliche Auszeichnung der formalen Struktur der Quellentexte hinaus.

Die über diese obligatorische Grundform der elektronischen Quellen des Textarchivs hinausgehenden weiteren Bearbeitungszustände enthalten zusätzliche Eintragungen, die der Steuerung der Lemmatisierung, der Auswahl oder dem Ausschluß von Belegen, ihrer Kommentierung usw. dienen. Sie brauchen hier im einzelnen nicht weiter beschrieben zu werden, dies ist bereits an anderer Stelle ausführlich geschehen.¹¹

Der beträchtliche Aufwand, der für die Einrichtung und Bereithaltung des Textarchivs getrieben wird, scheint auf den ersten Blick kein Gegenstück im herkömmlichen Verfahren der Materialbereitstellung zu haben und lädt daher besonders zu kritischen Nachfragen ein. Was läßt sich zu seiner Begründung anführen?

Zunächst ist darauf aufmerksam zu machen, daß es sich in Wirklichkeit entweder um die ökonomische Vorverlegung und konzentrierte Durchführung von Arbeiten handelt, die im herkömmlichen Verfahren in späteren Phasen der Wörterbucharbeit ausgeführt werden müssen, oder um die Vorbereitung der programmgesteuerten Erledigung solcher Arbeiten. Erspart wird zum Beispiel

- das mühsame und fehleranfällige Abschreiben von einzelnen Belegen, sei es bereits bei der Exzerption selbst, sei es bei der Abfassung bzw. Erfassung der Artikelentwürfe
- die zusätzliche Konsultation der gedruckten Quellenwerke während der Artikelarbeit, die herkömmlich häufig dennoch nötig war, zum Beispiel weil die Kontextmenge auf dem Belegzettel nicht genügte oder weil die Transkription Zweifel erweckte
- die nicht weniger fehleranfällige Notierung der Quellensigle und genauen Fundstelle
- die in der Regel als letzter Arbeitsgang nötige Belegrevision (Kontrolle der Belegzitate des fertigen Artikels an den Quellen)
- schließlich ein großer Teil des Aufwands, der vor oder zu Beginn der Artikelarbeit für die Sortierung der Belegzettel unter verschiedenen Gesichtspunkten getrieben werden mußte (Lemmatisierung; Gruppierung lemmatisierter Belege innerhalb eines Stichworts nach verschiedenen Gesichtspunkten der Quellentypologie oder nach Wortformen usw.).

¹⁰ Es handelt sich um ein deutsch-amerikanisches Kooperationsvorhaben, das an der Universität Trier in Zusammenarbeit mit dem Electronic Text Center der University of Virginia, Charlottesville, durchgeführt werden soll und die Einrichtung eines großen Digitalen mittelhochdeutschen Textarchivs für die Publikation im Internet und auf CD-ROM zum Ziel hat. Vorbereitende Arbeiten, die durch eine Anschubfinanzierung aus dem Forschungsfonds der Universität Trier unterstützt werden, wurden im April 2000 aufgenommen. Ein Antrag auf Drittmittelförderung ist eingereicht worden. Vgl. die Homepage des Vorhabens unter dem URL: <http://gaer27.uni-trier.de/MhdTA/welcome.htm>.

Ferner bietet die elektronische Auswertung der Wörterbuchquellen auch einen substantiell lexikographischen Vorteil, die Möglichkeit nämlich, Exzerption und Belegauswahl voneinander zu trennen und die letztere zu einem späteren Zeitpunkt vorzunehmen, wo sie mit größerer lexikographischer Einsicht durchgeführt werden kann. Herkömmlich muß schon wegen des Aufwands, der mit dem Abschreiben oder Kopieren von einzelnen Belegstellen und mit der weiteren Bearbeitung der Belegzettel (Sortierung usw.) verbunden ist, gleich bei der Exzerption, also in der Regel – wenn nicht nach vorhandenen Wortverzeichnissen exzerpiert wird – bei der fortlaufenden Lektüre einer Quelle, eine Auswahl getroffen werden: „Bei traditioneller Arbeit ist Auswahl ein praktischer Notbehelf und findet immer zu früh statt [...]“ (Sappeler 1991:281). Die Ergebnisse dieser frühzeitigen Auswahl können bei der späteren Artikelarbeit, wenn die Sichtweise stichwortbezogen ist und die Belegauswahl besser begründet vorgenommen werden könnte, nur noch ganz begrenzt durch Nachexzerption korrigiert werden.¹² Bei elektronischer ‚Verzettelung‘ der Quellen in der vorgestellten Weise können dem Artikelbearbeiter grundsätzlich alle Belege einer Quelle für ein Stichwort zugänglich gemacht werden. Daß dann im Bereich des hochfrequenten Wortschatzes Maßnahmen gegen nicht mehr überblickbare Überfülle von Belegen getroffen werden müssen, ist klar; näheres dazu unten unter 4. zum Belegarchiv.

Der Aufwand, der für die Bereitstellung der elektronischen Quellenbasis des Wörterbuchs getrieben werden muß und auf den ersten Blick unverhältnismäßig hoch erscheinen mag, ist also nicht allein durch die effektive Organisation der computergestützten Arbeitsabläufe, sondern vor allem lexikographisch gut begründet. Er entlastet die eigentliche Ausarbeitung des Wörterbuchs, weil ein großer Teil der ‚niederen‘, bei der Ausarbeitung der Artikel störenden und ablenkenden Arbeit bereits bei der vorbereitenden Materialbereitstellung erledigt wird; und er erlaubt es, die Belegbasis des Wörterbuchs lexikographisch zuverlässiger gesteuert einzurichten.

Wenngleich das elektronische Textarchiv genuin lexikographisch begründet, als Materialbasis für den Zweck der Ausarbeitung des Wörterbuchs eingerichtet ist, so ist nicht zu übersehen, daß es einen über die Wörterbuchzwecke hinausgehenden eigenen Wert für andere, auch wörterbuchferne linguistische oder literaturgeschichtliche Untersuchungsanliegen besitzt. Geeignet dafür macht es gerade die gewählte ‚ausgabendiplomatische‘ Einrichtung, weil sie einerseits die zugrundeliegenden Editionstexte ohne spezielle, an den Wörterbuchzweck gebundene Änderungen abbildet, andererseits aber bereits eine reiche sachliche Auszeichnung bietet. Das elektronische Quellenarchiv des Wörterbuchvorhabens dürfte sich daher gut eignen als Kern einer zukünftigen digitalen mittelhochdeutschen Bibliothek im WWW.

3 Die Lemmatisierungskomponente

Der Kern der Lemmatisierungskomponente ist das ‚Gerüst‘. Es enthält die allgemeinen, nicht einzelstellenbezogenen Lemmatisierungsinformationen, das heißt Lemmata und ihnen zuzuordnende Wortformen; homographe Wortformen, die mehreren Lemmata gemeinsam

¹¹ Plate/Recker (2000).

¹² Ausführlich zu diesem Problem und zu den im herkömmlichen, nicht-computergestützten Verfahren möglichen Gegenmaßnahmen Reichmann (1986), hier Abschnitt „6.2. Die Quellenbearbeitung“, S. 48–51 und Reichmann (1990).

sind, werden durch angehängte Markierungen unterschieden (und ihre Vorkommen in den Texten entsprechend ausgezeichnet). Die formale Struktur des Gerüsts ist an anderer Stelle bereits beschrieben und durch Abbildung veranschaulicht,¹³ hier soll es nur um seine lexikographische Funktion gehen.

Das Gerüst ist vor allem Lemmatisierungsinstrument, es dient also zunächst – im Zusammenspiel mit den Texten und gesteuert von Programmen, die mit Texten und dem Gerüst als Eingabe arbeiten – der Bereitstellung des lemmatisierten Belegarchivs. Als solches ist es dynamisch, d.h. es wächst mit jedem weiteren Text vor allem um die darin neu auftretenden Wortformen, in geringem Maße auch um neue Lemmata; letzteres dann, wenn Lemmata belegt werden, die noch nicht in den aus den Vorgängerwörterbüchern und dem FINDEBUCH kompilierten und vor der Bearbeitung der ersten Quellentexte in das Gerüst eingetragenen Lemmaliste verzeichnet sind. Grundsätzlich wäre es möglich gewesen, auch auf Lemmaebene mit einem leeren Gerüst zu beginnen und die Stichwörter sukzessive mit ihrer erstmaligen Belegung in den bearbeiteten Texten anzusetzen. Der Vorteil des gewählten Verfahrens, das sich im Falle des Mittelhochdeutschen die Existenz umfangreicher Vorgängerwörterbücher zunutze machen konnte, dürfte aber auf der Hand liegen: Die bei der Bearbeitung der Texte für das Belegarchiv sonst in jedem einzelnen Fall eines neu belegten Lemmas durchzuführende Arbeit der Festlegung der Gestalt des Lemmazzeichens und der häufig nötigen zusätzlichen unterscheidenden Angaben (grammatische Angaben, Homographenindizes) wird erspart. Die bereits vorhandene Lemmaliste ist die Grundlage für die Erzeugung automatischer Lemmatisierungsvorschläge für neue, noch nicht im Gerüst vorhandene Wortformen. Außerdem bietet die Gesamtlemmaliste aus den Vorgängerwörterbüchern eine ungefähre Zielvorgabe und damit einen Anhaltspunkt für die Qualifizierung des Arbeitsstandes der Materialbereitstellung, jedenfalls insoweit, wie die Lemmaebene betroffen ist; für weitergehende Fragen ist der jeweilige Stand des Belegarchivs einzubeziehen, s. dazu unten unter 4.

Weil die allgemeine Lemmatisierungsinformation in einer besonderen Komponente gespeichert wird, ist die Lemmatisierung leicht veränderbar. Dies betrifft sowohl Änderungen bei einzelnen Wortformen wie auch systematische Abwandlungen der äußeren Zugriffsstruktur überhaupt. Der erste Fall tritt regelhaft auf bei der Korrektur der halbautomatischen Lemmatisierung jedes Textes, der erstmals für die Aufnahme seiner Belegstellen in das Belegarchiv bearbeitet wird. Hier handelt es sich unter anderem darum, die automatische Einordnung neuer Wortformen ins Gerüst zu kontrollieren und gegebenenfalls falsch eingeordnete Wortformen umzustellen zu einem anderen Stichwort. Würde diese allgemeine Lemmatisierungsinformation statt im Gerüst (Zuordnung einer *Wortform zu einem Stichwort) in den Texten (Zuordnung bei jedem Vorkommen dieser Wortform) aufbewahrt, dann müßte die Korrektur bei allen Vorkommen einer Wortform in den ausgewerteten Texten ausgeführt werden; im Gerüst dagegen ist die Korrektur nur an einer einzigen Stelle nötig.

Ebenso leicht sind Änderungen auf Lemmaebene möglich. Das Bedürfnis kann aus verschiedenen Gründen auftreten, die Durchführung ist aber im herkömmlichen Zettelkastenverfahren mit erheblichem Aufwand verbunden, der die Revision einmal getroffener Entscheidungen über die Lemmatisierungsprinzipien erschwert. Im Falle des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS waren bzw. sind zum Beispiel verschiedene Lemmatisierungsgrundsätze, die der aus den Vorgängerwörterbüchern kompilierten Lemmaliste zugrundeliegen, nicht unumstritten; die Entscheidung darüber, ob sie letztlich beibehalten oder verworfen werden, kann sich allein an dem sachlich begründeten Dafürhalten

¹³ Plate/Recker (2000).

orientieren, weil der Arbeitsaufwand für die gegebenenfalls erwünschten Änderungen im Gerüst verhältnismäßig gering wäre. Erwogen wurde zum Beispiel, ob die trennbaren Partikelverben, ein Massenphänomen bereits im Mittelhochdeutschen, die in der Lemmaliste als artikelwertige Stichwörter an alphabetischer Stelle erscheinen, wie im BMZ unter den Simplexverben mitbehandelt oder wie im LEXER in Reihenartikeln zusammengefaßt werden sollten.¹⁴ Weitere Beispiele, bei denen die Änderung in Richtung auf Zusammenlegung zu größeren Einheiten diskutiert worden ist, betreffen die in der Lemmaliste bislang ebenfalls als eigene artikelwertige Stichwörter behandelten Ansätze der Adjektivadverbien¹⁵ und der *ge-präfigierten* Verben.¹⁶ Alle erwogenen Änderungen könnten mit verhältnismäßig geringem Aufwand jederzeit bis zum Beginn der Ausarbeitung durchgeführt werden, ohne daß Arbeiten in den bis dahin ausgewerteten Texten selbst nötig würden. Selbstverständlich sind Umstellungen aller Art ferner auch, unabhängig von der letztlich im Wörterbuch durchzuführenden Zugriffsstruktur, für beliebige andere Zwecke jederzeit möglich.

Eine weitere Benutzungsmöglichkeit des Gerüsts sei hier angedeutet (vgl. dazu ausführlicher unten unter 5.2.1): Es kann durch eine einfache automatische Umordnung in ein alphabetisches Register aller im Belegarchiv bezeugten Schreibformen verwandelt werden, in dem von jeder Schreibform der bearbeiteten Quellentexte auf das bzw. die Lemmata verwiesen wird, dem oder denen sie zuzuordnen sind.

Das Gerüst ist zusammen mit den zugehörigen Programmen ein wirkungsvolles Instrument für die halbautomatische Lemmatisierung mittelhochdeutscher Texte und kann auch außerhalb des Wörterbuchprojekts benutzt werden. Bedarf dazu dürfte vor allem auf Seiten der Herausgeber mittelhochdeutscher Texte bestehen, die im Zuge ihrer Editionsarbeiten Konkordanzen benötigen und/oder ihren Ausgaben Glossare oder Wörterbücher beigeben wollen. Lemmatisierte Konkordanzen sollten eigentlich zum selbstverständlichen Rüstzeug eines jeden Herausgebers gehören, ihre Anfertigung erschien aber noch vor nicht allzu langer Zeit auch bei computergestützter Arbeit so aufwendig, daß man häufig darauf verzichtete oder sich mit einfachen Wortformenindizes behalf. Selbst die publizierten Produkte bleiben noch bis in die jüngste Zeit hinein beträchtlich hinter dem zurück, was angesichts der heutigen Möglichkeiten als Standard für derartige Hilfsmittel gelten müßte.¹⁷

4 Das Belegarchiv

Das Belegarchiv wird aus den dafür vorbereiteten Quellentexten mit der Hilfe des Gerüsts automatisch erzeugt. Es handelt sich um eine dynamische Größe, die mit jedem weiteren in geeigneter Weise präparierten Text wächst. Die Ausgabemöglichkeiten sind vielfältig, es kann in Konkordanzform mit frei wählbarer Kontextlänge vollständig oder für beliebige Teilstrecken, auf der Grundlage des gesamten Textarchivs oder von Teilcorpora ausgeben und entweder auf Papier ausgedruckt oder mit verschiedenen Recherchemöglichkeiten elektronisch z.B. in einem Internetbrowser über die Homepage des Projekts zur Verfügung gestellt werden.¹⁸ Zur Zeit dient die Ausgabe des Belegarchivs vor allem der

¹⁴ Ausführliche Diskussion dieser Frage bei Plate (2000a).

¹⁵ Vgl. Baumgarte (2000).

¹⁶ Vgl. Tao (2000) und Wawer (2000).

¹⁷ Vgl. die Übersicht bei Gärtner/Kühn (1998).

Kontrolle der halbautomatischen Lemmatisierung, daneben wurde es auch bereits für die Anfertigung von Probeartikeln benutzt (vgl. dazu unten unter 5) und ferner für verschiedene Recherchen im Zusammenhang mit der Diskussion konzeptioneller Fragen.

Umfang und Zusammensetzung des Belegarchivs hängen ab vom Umfang der bearbeiteten Texte und vom Grad ihrer Auswertung für das Belegarchiv. Oben unter 2 wurde bereits darauf hingewiesen, daß bei der Eingabe von Volltexten grundsätzlich jeder Beleg eines Textes ins Belegarchiv aufgenommen werden kann. Dies ist ein Vorteil gegenüber dem herkömmlichen Verfahren, bei dem Exzerption und Belegauswahl weitestgehend zusammenfallen und der späteren Artikelbearbeitung stark voreignen, indem sie ihr in teils beträchtlichem Umfang Belege vorenthalten, obwohl in der Phase der Exzerption die Entscheidung über die Belegauswahl wesentlich schlechter begründet durchzuführen ist als bei der eigentlichen Artikelarbeit, wo das Material erstmals in stichwortbezogener Zusammenstellung in den Blick kommt. Andererseits würde eine unbeschränkte Aufnahme aller Belege auch des höchstfrequenten Wortschatzes (vor allem der Synsemantika) eine für den Artikelbearbeiter kaum noch überschaubare Belegfülle ergeben. Für das Belegarchiv des mittelhochdeutschen Wörterbuchs werden die Quellen daher in unterschiedlicher Dichte ausgewertet. Das Grundkorpus des Belegarchivs gliedert sich entsprechend der Exzerptionsdichte in drei Gruppen. Die erste Quellengruppe aus insgesamt 18 Texten und Textsammlungen¹⁹ wird vollständig für das Belegarchiv ausgewertet, das heißt jeder Beleg dieser Texte, einschließlich aller Belege für den höchstfrequenten Wortschatz, erscheint im Belegarchiv. Bei der zweiten, insgesamt 21 Texte umfassenden Quellengruppe²⁰ wird der höchstfrequente Wortschatz für das Belegarchiv ausgeblendet, die Belege für den übrigen Wortschatz gehen aber noch vollständig ins Belegarchiv ein. Nur die in vorhandenen Ausgabenglossaren, Wortverzeichnissen und Registern verzeichneten Belege werden schließlich in der dritten, 34 Texte zählenden Quellengruppe erhoben.²¹ Das gezielte Ausblenden von Belegen für hochfrequenten Wortschatz (zweite Quellengruppe) bzw. die gezielte Aufnahme von Glossarbelegen (dritte Quellengruppe) wird technisch durch spezielle Markierungen an den betreffenden Wortformenvorkommen bewirkt.

Das Grundkorpus wird nach unserer Erwartung und den ersten Erfahrungen an Probeartikeln die Forderung nach einem „die Homogenität sichernden, möglichst breiten zentralen Teil“ des gesamten Quellenkorpus (Reichmann 1990:1603, hier T3) erfüllen können. Für die weitere „Erfassung peripherer Wortschatzteile“ (ebd.; ergänze: und peripherer Gebrauchsweisen des zentralen Wortschatzes) wird eine größere Zahl zusätzlicher Quellen herangezogen werden müssen. Systematisch vorbereitet werden kann das zum einen für jene durch das FINDEBUCH erschlossenen Quellen, die nicht bereits im Grundkorpus vertreten sind²² und durch geeignete computergestützte Verfahren der Auswertung der beiden Vorgängerwörterbücher, die durch eine der Trierer Arbeitsstelle angeschlossene Arbeitsgruppe maschinenlesbar gemacht wurden und miteinander (und mit dem FINDEBUCH) verknüpft im WWW für elektronische Benutzung publiziert werden.²³

Einige Zahlen zum Stand des Belegarchivs im Juli 1999 mögen eine Vorstellung vom Umfang der Sammlung vermitteln. Insgesamt sind nach der Bearbeitung aller Texte der

¹⁸ Die Nutzungsmöglichkeit über die Projekthomepage bleibt zur Zeit noch den Arbeitsstellen vorbehalten, soll jedoch bald frei zugänglich gemacht werden.

¹⁹ Vgl. in der Bibliographie im Anhang zu Gärtner (2000) und auf der Homepage des mittelhochdeutschen Wörterbuchs die Nummern P1–P9 und A1–A9.

²⁰ Vgl. die Bibliographie [wie Fußnote 18] unter B1–B21.

²¹ Vgl. die Bibliographie [wie Fußnote 18] unter C1–C34.

²² Vgl. die Bibliographie auf der Homepage des Projekts unter ‚Findebuchquellen‘.

ersten, von acht Texten der zweiten und von neun Texten der dritten Quellengruppe des Grundkorpus 18.341 Lemmata mit zusammen rund 960.000 Stellen im Belegarchiv vertreten; die acht Texte der zweiten Quellengruppe enthalten für den höchstfrequenten Wortschatz weitere rund 350.000 Belege, die aber für das Belegarchiv ausgeblendet worden sind. Die weitere Bearbeitung der zweiten und dritten Quellengruppe wird die Zahl der im Belegarchiv ausgeblendeten Stellen (aus der zweiten Quellengruppe) stark erhöhen und insgesamt die Relation Lemma/Belege zugunsten der Lemmata verändern, d.h. es werden, vor allem durch die dritte Gruppe der gezielt nach Ausgabenglossaren bearbeiteten Quellen, verstärkt Belege für niederfrequenten Wortschatz hinzukommen.²⁴ Der niederfrequente Wortschatz macht in allen lexikalischen Corpora den weitaus größten Teil der Stichwörter und den geringsten Teil der Belege aus: Zur Zeit (September 1999) sind im Belegarchiv 7.156 Lemmata, das sind knapp 40 Prozent aller belegten Stichwörter, mit nur einem einzigen Beleg vertreten. Bei herkömmlicher Exzerption wird dieser periphere Wortschatz verhältnismäßig stark bevorzugt berücksichtigt. Die Feststellung Oskar Reichmanns, daß in der Belegsammlung des FWB der „allgemeinsprachliche Wortbestand“ und seine „eigentlichen, allgemeinsprachlichen und geschichtlich relativ konstanten“ Gebrauchsweisen verhältnismäßig zu gering belegt seien,²⁵ legt diese Schwäche der herkömmlichen Belegexzerption und Belegauswahl offen. Die breite Exzerption des häufiger belegten Kernwortschatzes, die bei elektronischer Auswertung der Quellen möglich ist, vergrößert für den Artikelbearbeiter in diesem Bereich die Menge des Überblickbaren und eröffnet daher die Möglichkeit, den Gebrauch des Kernwortschatzes im Wörterbuch angemessener als bislang zu beschreiben. Dies scheint uns entschieden zu den Chancen der computergestützten Lexikographie zu gehören.

Wie das Textarchiv, so ist auch das Belegarchiv über die Zwecke des Wörterbuchs hinaus für andere Anliegen benutzbar. Seine Bereitstellung im Internet für elektronische Nutzung mit verschiedenen Recherchemöglichkeiten wird zur Zeit vorbereitet, für die beiden Arbeitsstellen ist der Zugriff über einen internen Teil der Homepage des Projekts bereits realisiert. Bei herkömmlicher Arbeit waren Wörterbuch-Belegarchive für den weiteren Kreis der potentiell an solchen Sprachmaterialien Interessierten nur sehr schwer zugänglich; man mußte sich eben in die Arbeitsstelle selbst an die Zettelkästen begeben oder riskieren, mit umständlichen Anfragen an die Wörterbuchbearbeiter den Betrieb aufzuhalten. Die ungehinderte und bequeme Zugänglichkeit des Belegarchivs für jedermann ist ein Vorteil der elektronischen Materialbereitstellung.

5 Das Programmsystem für die Artikelarbeit

Die Erfahrungen mit der Artikelarbeit beschränken sich in der jetzigen Phase der Materialbereitstellung auf wenige größere Artikel, die zur Erprobung konzeptioneller Überlegungen verfaßt wurden.²⁶ Aus ihnen ergibt sich jedoch schon eine Reihe von Gesichtspunkten hinsichtlich der Chancen und Perspektiven dieser Phase der computergestützten Wörterbucharbeit, die im folgenden benannt seien.

²³ Vgl. dazu Burch/Fournier/Gärtner 1998. – Selbstverständlich dienen die Vorgängerwörterbücher dabei nur als Wegweiser zu Belegmaterial, dieses muß dann aus den Quellen erhoben werden.

²⁴ Die Relation Lemma : Belege im Belegarchiv beträgt bei den Texten der ersten Quellengruppe 1:43, in der zweiten Quellengruppe 1:21, in der dritten 1:10.

5.1 Leistungsmerkmale des Artikelredigierprogramms

Es dürfte grundsätzlich Einigkeit in der Erwartung bestehen, daß die eingangs erwähnte Option, mit Beginn der Phase der Ausarbeitung des Wörterbuchs zum traditionellen Verfahren der Artikelarbeit überzugehen, die eigentlichen Chancen der elektronischen Materialbereitstellung verspielen würde. Andererseits ist es offensichtlich, daß es mit der einfachen Ausgabe der Belege in ein Textverarbeitungsprogramm nicht getan ist, sondern daß ein eigenes System von Programmen für die Artikelarbeit benötigt wird, für dessen Realisierung philologische Einsicht in die erforderlichen Leistungsmerkmale ebenso wie anspruchsvolle Programmierkenntnisse benötigt werden. Die Ermutigung, diese Herausforderung anzunehmen, kam vor allem aus der überzeugenden Demonstration eines wiederum von Paul Sappler (Tübingen) entwickelten Prototypen, der bereits 1988 verfügbar war.²⁷ Er wurde anlässlich der Abfassung von Probeartikeln in den Arbeitsstellen des mittelhochdeutschen Wörterbuchs vom Programmator für die Bedürfnisse des Projekts eingerichtet und bietet das Vorbild für die Entwicklung eines umfassenderen Programmsystems für die Artikelarbeit.

Die Leistungsmerkmale des Artikelredigierprogramms lassen sich in drei Hauptgruppen einteilen. Die erste Gruppe unterstützt vorbereitende Sortierläufe, denen bei herkömmlicher Arbeit das Ordnen der Belegzettel für ein Stichwort nach bestimmten äußeren Kriterien entspricht, z.B. alphabetisch nach Wortformen, nach Quellen in chronologischer Reihenfolge oder einem beliebigen anderen quellentypologischen Merkmal, innerhalb der Belege einer Quelle aufsteigend nach Vorkommen usw. Ferner gehört dazu die Unterstützung einer inhaltlichen Sortierung nach ausdrucksseitig abfragbaren Merkmalen der Textumgebung; wie bei der Artikelarbeit Kontextabfragen im elektronischen Belegmaterial zu einem Stichwort sinnvoll eingesetzt werden können, zeigt an einem Beispiel Sappler (1999:280).

Die zweite Gruppe betrifft die eigentlichen Editierungsfunktionen. Dabei handelt es sich um Unterstützungen

- für das Umordnen der Belege nach Gebrauchstypen
- für die Einfügung von Gliederungsmarken und entsprechenden Gliederungskommataren (Bedeutungserläuterung usw.)
- für die Verdopplung von Belegen, die mehr als einmal im Artikel zitiert werden sollen
- für das Ausblenden von Belegen, die nicht im Artikel erscheinen sollen, oder des Belegzitats, wenn nur die Stelle angegeben werden soll
- für die mit dem Belegzitieren im engeren Sinne zusammenhängenden Arbeiten wie Festlegung des Belegschnitts, (gekennzeichnete) Kürzung oder Ergänzung, einzelstellenbezogene Kommentierung einschließlich der Möglichkeit, alle diese Eingriffe wieder aufzuheben und den originalen Wortlaut automatisch zu restituiieren
- schließlich für das Verweisen innerhalb des Artikels und auf andere Artikel.

Bei der dritten Gruppe von Funktionen des Artikelredigierprogramms handelt es sich um durch Prüfprogramme unterstützte Sicherungsfunktionen, die z.B. Datenteile gegen unerlaubte Aktionen des Artikelautors schützen (vor allem die Fundstellenangabe, die unantastbar ist, aber auch die Belege selbst, die zwar in jeder beliebigen Weise geändert oder

²⁵ Lexikographische Einleitung (1986), S. 160 (zu den Gründen vgl. dort S. 48–51).

²⁶ Sie sind zum Teil publiziert in den Beiträgen des von Gärtner/Grubmüller (2000) herausgegebenen Tagungsbandes.

²⁷ Vgl. die anschauliche Darstellung bei Sappler/Schneider-Lastin 1991.

auch mit einer Markierung für Ausblendung versehen, nie aber aus der Artikelarbeitsdatei völlig gelöscht werden dürfen) und verschiedene weitere formale Konsistenzmerkmale geltend machen.

Die Erfahrungen mit dem bereits realisierten Leistungsumfang des Artikelredigierprogramms sind so positiv, daß eine Rückkehr zur herkömmlichen Arbeit mit Belegzetteln oder dem Ausformulieren der Artikel in einem Textverarbeitungsprogramm für uns nicht mehr in Frage kommt. Als Hauptvorteil erscheint uns, daß die Aufmerksamkeit in höherem Maße der eigentlichen philologischen Arbeit im Belegmaterial zugewendet werden kann, während die lästigen Schreib-, Kontroll- und Sortierarbeiten weitgehend der Maschine übertragen sind. Ein nicht zu unterschätzender Vorteil ist es auch, daß mit der Hilfe eines anschließenden Satzprogramms jederzeit eine Vorschau auf den gesetzten Artikel möglich ist.

Abschließend seien zwei Eigenschaften der elektronischen Artikeltexte hervorgehoben, die für die Funktionsweise eines Programmsystems mit dem beschriebenen Leistungsumfang unerlässlich sind, daneben aber auch der späteren Benutzung des elektronischen Wörterbuchs (vgl. dazu unten unter 6) zugute kommen: Zum einen die Tatsache, daß die Belegzitate mit dem elektronischen Belegarchiv und damit den ihm zugrundeliegenden elektronischen Quellentexten verknüpft sind, zum anderen die Strukturierung des Artikeltextes in Felder (dies ist vor allem für die Prüfprogramme nötig). Ersteres ermöglicht es, direkt aus der elektronischen Fassung des Wörterbuchs heraus die für ein Belegzitat angegebene Fundstelle in einem Quellentext anzusprechen (und so das Belegzitat in seinem ursprünglichen Gebrauchsverband zu prüfen, vgl. unter 6.1), die zweite Eigenschaft erlaubt das auf Felder beschränkte Suchen im elektronischen Wörterbuchtext (vgl. unter 6.2.2).

5.2 Mehrwert der Artikelarbeitsdateien gegenüber dem Artikel im publizierten Wörterbuch

Vor allem bei hochfrequentem Wortschatz und großen Artikeln haben die Artikelarbeitsdateien einen gewissen Mehrwert gegenüber dem fertiggestellten Artikel, der im publizierten Wörterbuch erscheint. Dies ergibt sich daraus, daß die Belegauswahl zu einem großen Teil erst während der Artikelarbeit vorgenommen wird und nicht bereits bei der Auswertung (Exzerption) der Quellen für das Belegarchiv. Bei der Artikelarbeit kann sich die Auswahl bereits auf eine Ordnung des Belegmaterials nach Gebrauchstypen („Bedeutungen“ usw.) stützen. Im Ergebnis wird daher bei höherfrequentem Wortschatz für bestimmte Artikelpositionen immer eine größere Menge des Belegmaterials gesichtet und grob geordnet worden sein als im publizierten Wörterbuchartikel zitiert wird. Für Spezialuntersuchungen, feinere Durcharbeitung usw. steht in den Artikeldateien dann also bereits weiteres vorsortiertes Material zur Verfügung. Ähnliches gilt für bestimmte Materialordnungen, die auf dem Weg zur endgültigen Artikelgliederung erprobt, bei der Weiterarbeit dann aber als im Artikelzusammenhang nicht aussagekräftig verworfen worden sind; für andere Untersuchungszwecke können sie jedoch durchaus ihren Wert haben.

6 Konzeptionelle Fragen der Publikation des Wörterbuchs für elektronische Nutzung

Als erstes historisches Belegwörterbuch des Deutschen wird das neue MITTELHOCHDEUTSCHE WÖRTERBUCH vollständig computergestützt erarbeitet. Quellen, Belegarchiv und Artikelarbeit bis hin zum Satz der Artikel sind Zustände und Prozesse eines geschlossenen elektronischen Kreislaufs. Daher besteht selbstverständlich auch die Möglichkeit der Publikation des Wörterbuchs selbst in den ‚neuen Medien‘ WWW und CD-ROM; ja man könnte die elektronische Publikation sogar als das Gegebene, die Ausgabe im Druck dagegen als auffälligen sekundären Medienwechsel und als zusätzlichen Schritt des lexikographischen Prozesses auffassen.

Die Möglichkeiten der neuen elektronischen Publikationsformen werden aktuell auch unter Wörterbuchmachern engagiert diskutiert, wobei eine begrüßenswerte Verlagerung der Aufmerksamkeit von eher unspezifischen, gelegentlich geradezu trivialen Gesichtspunkten hin zu substantiell lexikographischen Fragen zu beobachten ist. Die ‚neuen Medien‘ verstärken also in erfreulicher Weise die Forderung nach konzeptioneller Reflexion und Explizitheit, der die Wörterbuchmacher bereits durch die seit den 70er Jahren auch im deutschen Sprachbereich blühende theoretische Wörterbuchforschung und Wörterbuchkritik ausgesetzt sind und deren erste reiche Ernte in Oskar Reichmanns ‚Lexikographischer Einleitung‘ (1986) zum FRÜHNEUHOCHDEUTSCHEN WÖRTERBUCH gehalten wurde.

Zwei konzeptionelle Probleme der mit den neuen Medien verbundenen „Chancen und Perspektiven“ historischer Belegwörterbücher seien hier diskutiert. Sie betreffen das Verhältnis des ausgearbeiteten Wörterbuchs zu seiner Basis (Quellen und Belegarchiv) und das Verhältnis der Benutzung des Wörterbuchs über Artikelstichwörter zur Benutzung über andere, registerartige Zugriffshilfen.

6.1 Das Verhältnis von Wörterbuch und Wörterbuchbasis

Die Frage nach dem Verhältnis von Wörterbuch und Wörterbuchbasis ergibt sich bei elektronischer Publikation dadurch, daß die Möglichkeit der Verknüpfung sowohl mit dem lemmatisierten Belegarchiv wie auch mit den Quellentexten selbst besteht; beide können dann direkt aus einem Wortartikel heraus angesprochen werden. Im Falle des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS ist diese Verknüpfung in der Datenbasis durch die Art der Materialbereitstellung und die Anlage des Programmsystems für die Artikelarbeit (vgl. dazu oben unter 5.1) von vornherein gegeben, sie muß also nicht nachträglich hergestellt werden.

Das Belegarchiv enthält in der Regel mehr Belege als im Wörterbuch-Artikel zitiert werden, es bietet also für Spezialuntersuchungen zusätzliches Belegmaterial und es erlaubt – vor allem bei Unterstützung durch entsprechende Editorfunktionen – die Arbeit mit ihm. Das ist für die Wörterbuchbenutzung unter bestimmten speziellen Fragestellungen unbedingt ein Zugewinn der elektronischen Publikation, und insgesamt möglicherweise ein Anreiz zu selbständigerem, kritischem Umgang mit dem Wörterbuch. Um falschen Erwartungen vorzubeugen, ist aber daran zu erinnern, daß das elektronische Belegarchiv die Form einer lemmatisierten KWIC-Konkordanz hat, die noch keine weiteren lexikographischen Strukturierungen aufweist. „Das Lesen in einer Konkordanz kann für den Philologen durchaus anregend sein; man nimmt in Kauf, daß es kein Gespräch mit einem dar-

stellenden Lexikographen ist und man nicht auf Beobachtungen und Wissen anderer aufbaut, sondern selber ganz von unten anfängt“ (Sappler 1991:278). Der substantielle Unterschied zwischen Rohmaterialien und Wörterbuch darf im Interesse der Benutzer also nicht verwischt werden. Der Regelfall der Konsultation verlangt einen ausgearbeiteten Artikel, der für den kundigen Benutzer des betreffenden Werks aus sich heraus verständlich ist; die Bereitstellung mehr oder weniger unbearbeiteter Materialien kann hier nur ein zusätzliches Angebot an diejenigen sein, die bereit und in der Lage sind, ihre Fragestellung ohne lexikographische Hilfe zu verfolgen.

Die Verknüpfung des Wörterbuchs mit den Quellentexten kann theoretisch dem Zugriff in beiderlei Richtung dienen: Zum einen von den im Wörterbuch zitierten Belegen zu den jeweiligen Stellen in den Quellentexten; zum anderen von einer Textstelle (Wortform) zum betreffenden Wörterbuchartikel. Einleuchtend ist vor allem der Weg vom Belegzitat zum Quellentext, der die Prüfung einer Stelle in ihrem ursprünglichen Gebrauchszusammenhang erlaubt.

Auch die umgekehrte Richtung sieht auf den ersten Blick vielversprechend aus, bei näherem Hinsehen erweist sie sich jedoch als problematisch. Reizvoll ist aus der Sicht der Wörterbuchmacher die Vorstellung, daß man auf diese Weise die Schwelle zur Benutzung des großen Belegwörterbuchs absenken und bereits während der Ausarbeitung des Wörterbuchs eine größere Benutzergemeinde aufbauen könnte; dies wäre besonders im Falle des Mittelhochdeutschen zu begrüßen, weil leider immer wieder festgestellt werden muß, daß nur wenige der potentiellen Benutzer den Umstieg vom TASCHENLEXER auf die großen Wörterbücher bewältigen.²⁸ Doch könnte eine solche Verknüpfung von Quellentexten mit dem Wörterbuch bei den Benutzern leicht Erwartungen wecken, die in der Regel enttäuscht werden würden. Denn da nur ein geringer Teil der Stellen jeder Quelle im Sprachwörterbuch zitiert wird, und da ferner die lexikographische Erläuterung dort in aller Regel nicht einzelnen Textstellen, sondern Typen des Gebrauchs einer Wortschatzeinheit gilt, leistete diese Verknüpfungsrichtung meist nicht mehr als die Hinführung zu dem einschlägigen Artikel des großen Wörterbuchs. Diese Hilfe kann erwünscht sein von Benutzern, denen die Lemmatisierungs- und Normalisierungsprinzipien des betreffenden Wörterbuchs nicht vertraut sind, doch ist sie gerade bei elektronischer Publikation des Wörterbuchs rationeller durch ein Schreibformenregister und gegebenenfalls durch weitere Hilfs-Zugriffsstrukturen des Wörterbuchs zu realisieren (vgl. dazu unten unter 6.2.1). Wer darüber hinaus einzelstellenbezogene Verständnishilfen sucht, der konsultiert auch in Zukunft besser die dafür gedachten textbezogenen Hilfsmittel wie Stellenkommentare, Ausgabenglossare oder Textwörterbücher.

6.2 Benutzung des Wörterbuchs über Artikelstichwörter, über Register und über registerähnliche Hilfen

Wenn es um die Chancen und Perspektiven der neuen elektronischen Publikationsformen geht, findet unter Lexikographen der Gesichtspunkt erweiterter Zugriffsmöglichkeiten auf das Wörterbuch gegenwärtig besondere Aufmerksamkeit. Das gedruckte Werk, so die Überlegung, erlaubt in der Regel nur eine einzige Zugriffsmöglichkeit, jene über die Arti-

²⁸ Zum Zusammenhang von BMZ, LEXER und TASCHENLEXER vgl. Plate (1997). Das neuerdings mit dem TASCHENLEXER konkurrierende „Kleine mittelhochdeutsche Wörterbuch“ von Beate Hennig ist von vornherein nicht auf die Hinführung zur Benutzung der großen Wörterbücher angelegt und erschwert sie sogar; vgl. dazu Plate (2000b).

kelstichwörter. Die elektronische Fassung gewährt dagegen grundsätzlich freien Zugang über beliebig definierbare Merkmale des Wörterbuchtexts (Volltextsuche) bzw. in bestimmten Textteilen der sachlich strukturierten Wörterbuchdaten (Suche in Feldern, z.B. in bestimmten Artikelpositionen oder sachlich markierten Textteilen wie Belegtext, Erläuterungstext, Quellensiglen usw.).

6.2.1 Register

Die systematische Beschäftigung mit der Einrichtung zusätzlicher Zugriffsstrukturen für Wörterbücher ist verhältnismäßig jungen Datums, sie stammt aber noch aus der Zeit vor dem Siegeszug der neuen elektronischen Publikationsformen und bedurfte nicht der Inspiration durch deren technische Möglichkeiten, denn sie konnte sich an dem herkömmlichen Hilfsmittel orientieren, das im gedruckten (und lange zuvor schon im geschriebenen) Buch zusätzlich zur Hauptzugriffsstruktur (wie sie z.B. ein Inhaltsverzeichnis darstellt) weitere Zugriffsmöglichkeiten anbietet, dem Register also.

Im Bereich der historischen Lexikographie des Deutschen hat wohl erstmals Oskar Reichmann in der ‚Lexikographischen Einleitung‘ (1986) des FWB nachdrücklich auf die Registermöglichkeiten hingewiesen,²⁹ nicht weniger als 15 verschiedene Register zur Strecke *a* bis *ausgang* wurden 1995 vollständig oder in Ausschnitten vorgelegt und ausführlich diskutiert in einer von Reichmann gemeinsam mit Ulrich Goebel und Ingrid Lemberg verfaßten Monographie, deren einprägsamer (Haupt-)Titel „Versteckte lexikographische Information“ inzwischen zum Schlagwort für eine bislang unerschlossene Wörterbuch-Dimension geworden ist. Er bezeichnet das „bekannte Faktum, daß Wörterbücher in aller Regel nur über eine einzige Zugriffsstruktur verfügen und daß damit alle Daten, die sich dieser Struktur entziehen, versteckt bleiben, so wichtig sie für andere lexikographische, linguistische, philologische, wortgeschichtliche sowie für kulturhistorische usw. Fragestellungen und außerdem zur Kontrolle der eigenen lexikographischen Tätigkeit auch sein mögen“ (Goebel/Lemberg/Reichmann 1995: VII). Geeignete Register können solche „versteckte Information“ zugänglich machen. Welche der zahlreichen möglichen Register realisiert werden sollten und können, hängt von den antizipierten Benutzungsanliegen, von der Ergiebigkeit des zugrundeliegenden Wörterbuchs hinsichtlich bestimmter Informationstypen und nicht zuletzt auch von dem je nach Registerart ganz unterschiedlichen Aufwand ab, der für die Gewinnung der Registerdaten und ihre weitere Bearbeitung zu einem tauglichen Hilfsmittel für die jeweilige Fragestellung nötig ist.

Nicht bei allen Registern handelt es sich tatsächlich um eigenständige weitere Zugriffsstrukturen. Das von Reichmann in der ‚Lexikographischen Einleitung‘ zum FWB (1986:160) erwogene Schreibformenregister z.B., das alle im zitierten Belegmaterial erscheinenden Schreibformen enthält und zu jeder das betreffende Artikelstichwort bzw. (bei homographen Schreibformen) die möglichen Lemmatisierungen angibt, ist eine bloße Hilfsstruktur für den Zugriff über Artikelstichwörter, die sich von der herkömmlichen Realisierung in Form von Verweisartikeln nur durch ihren Umfang und die Zusammenfassung in einem eigenen Alphabet unterscheidet. Im Falle des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS läßt sich ein solches Register automatisch aus dem Gerüst erzeugen (vgl. oben unter 3).

²⁹ Im Abschnitt „21. Register zur Erschließung von Datentypen des Wörterbuchs“, S. 158–163.

Eine weitere mögliche Hilfsstruktur, die bei Reichmann (1986) und bei Goebel/Lemberg/Reichmann (1995) nicht besprochen wird, könnte dazu dienen, Abweichungen von den aus gegenwärtssprachlichen Wörterbüchern gewohnten Lemmatisierungsprinzipien aufzufangen. In größerem Umfang treten sie in der historischen Lexikographie des Deutschen besonders im Bereich der Wortbildungen auf, speziell in der Tradition der mittelhochdeutschen Lexikographie bietet z.B. die Behandlung der trennbaren Partikelverben Anstoß.³⁰ Die Diskussion solcher strittigen Lemmatisierungsprinzipien kann sich ganz auf das sachlich Gebotene konzentrieren, wenn gewährleistet ist, daß abweichende Nachschlagegewohnheiten durch entsprechende Zugriffshilfen unterstützt werden. Dies gilt nicht ausschließlich, aber in besonderer Weise für die elektronische Wörterbuchfassung, weil bei ihrer Benutzung der Aufwand für das ‚Nachschlagen‘ generell stark reduziert ist.

Nicht um solche Hilfsstrukturen für den Zugriff über Artikelstichwörter, sondern um wirklich eigenständige andere Zugriffsmöglichkeiten geht es, wenn bei Goebel/Lemberg/Reichmann (1995) und in ihrer Nachfolge von der Erschließung „versteckter lexikographischer Information“ die Rede ist. Diese Idee, deren Realisierung, wie schon gesagt, nicht an die technischen Möglichkeiten der neuen Medien gebunden ist, wird gleichwohl in der gegenwärtigen Diskussion über die elektronische Benutzung von Wörterbüchern stark betont. Der Grund dafür sind bestimmte Vorteile, die elektronische Suchsysteme gegenüber dem herkömmlichen gedruckten Register bieten. Sie werden von Sappler (2000a) prägnant wie folgt charakterisiert: „Wo bisher ein – vom Schöpfer des Registers mit Überlegung formulierter – Registereintrag im Register aufgesucht wurde und wo anschließend den Referenzen in die so erschlossenen Grunddaten nachgegangen wurde, kann jetzt unmittelbar in den Grunddaten sichtbar gemacht werden, wo bestimmte, auch komplexe Abfragebedingungen erfüllt sind. Der Benutzer dieses „Registers“ hat mehr zur Verfügung (nämlich die gesamten Grunddaten), kann genauer zielen (durch Kombination der Bedingungen) und kommt schneller ans Ziel (mit elektronischer Geschwindigkeit).“ Andererseits ist nicht zu übersehen: „Meist aber, wenn auch nicht notwendig, bekommt er weniger Führung und Vorarbeit als durch ein traditionelles systematisches Register. Trotz guten Ansätzen scheint diese zweite Art des Umgangs mit Registern [d.h. der Umgang mit elektronischen Suchsystemen] methodisch erst am Anfang zu stehen“ (ebd.).

Es bedarf heute keiner prophetischen Gaben mehr, um zu erkennen, daß in sehr absehbarer Zeit die Publikation von Nachschlagewerken in elektronischer Form, mindestens zusätzlich zum Druck, der Regelfall sein wird. Daher müssen sich die lexikographischen Langzeitprojekte bereits jetzt der Frage stellen, ob und gegebenenfalls wie den im Vorstehenden angedeuteten neuen Möglichkeiten der Wörterbuchbenutzung konzeptionell, also in der Anlage des Wörterbuchs selbst, Rechnung zu tragen ist. Zwei Bemerkungen allgemeiner Art sollen dazu hier abschließend gemacht werden.

Zunächst gilt es festzuhalten, daß die Zugriffsstruktur alphabetisch-semasiologischer Wörterbücher, die herkömmlich in aller Regel als einzige realisiert ist, nicht eine beliebige unter mehreren und prinzipiell mit den Register-Zugriffsstrukturen gleichrangig ist, sondern diejenige, die dem usuellen Gebrauch des Wörterbuchs gemäß seinem genuinen Zweck (im Sinne von Wiegand 1998: 303ff.) dient; die von Goebel/Lemberg/Reichmann 1995 diskutierten Registerstrukturen dagegen bedienen, wie es in der oben einleitend zitierten Charakterisierung zutreffend heißt, „andere [...] Fragestellungen“ als jene, auf die die sprachwissenschaftliche Darstellungsform „(alphabetisch-semasiologisches) Wörterbuch“ zielt. Genaugenommen ist es daher auch unzutreffend, hinsichtlich des Registerziels von der Er-

³⁰ Vgl. Plate (2000).

schließung „versteckter lexikographischer Information“ zu sprechen: Insofern die gemeinte Information eine lexikographische ist, steht sie im jeweiligen Artikelzusammenhang an der richtigen Stelle; wenn Register bestimmte Informationstypen aus dem Artikelzusammenhang isolieren und einen anderen Zusammenhang herstellen, ist in aller Regel keine lexikographische Information intendiert.³¹ Ein Beispiel mag dies illustrieren. Die Angabe, daß ein bestimmtes Verb der Objektsprache in einer bestimmten Gebrauchsweise mit Genitiv oder Akkusativ oder Präpositionalobjekt konstruiert werden kann, ist im Zusammenhang des Wörterbuch-Artikels zu diesem Verb eine genuin lexikographische Information. Der entsprechende Eintrag in einem Register der Verbkonstruktionen unter einer Rubrik, die sämtliche im Wörterbuch dokumentierten Fälle der Konstruktion mit Genitiv als Objektskasus verzeichnete, wäre dagegen eine grammatische Information (die z. B. im Zusammenhang einer Untersuchung der Kasusfunktionen wertvoll sein könnte). Das gedachte Register der Verbkonstruktionen wäre ebensowenig als Mittel der Erschließung „versteckter lexikographischer Information“ zu bezeichnen, wie man umgekehrt das Wortregister zum Syntaxteil einer Grammatik ein Hilfsmittel für die Erschließung „versteckter syntaktischer Information“ nennen würde. Der beträchtliche Wert, den Wörterbuch-Register haben können, liegt also gerade darin, daß sie es erlauben, das Wörterbuch nicht nur als Wörterbuch zu benutzen, sondern auch zu anderen Zwecken, z.B. als Materialsammlung für grammatische Untersuchungen. Es ist vor allem eine Frage der für die Ausarbeitung des Wörterbuchs zur Verfügung stehenden Ressourcen, ob eine Wörterbuch-Arbeitsstelle es sich leisten kann, auch solche ‚uneigentlichen‘ Benutzungsweisen des Wörterbuchs zu unterstützen. Wenn sich die Wörterbuchmacher – wie Goebel/Lemberg/Reichmann (1995: 260) im Falle des FWB – dagegen entscheiden müssen, können sie es mit gutem lexikographischen Gewissen tun; ihrem Wörterbuch wird deswegen als Wörterbuch nichts Wesentliches fehlen.³²

6.2.2 Volltextsuche und Suche in Feldern

Das eben Ausgeführte gilt unabhängig davon, ob das Wörterbuch im Druck oder elektronisch benutzt wird. Die folgende Bemerkung betrifft die speziellen Möglichkeiten elektronischer Publikation. Hier trifft die Idee erweiterter Zugriffsmöglichkeiten bei Lexikographen auf besonders fruchtbaren Boden, weil sie zum Teil ohne den abschreck-

³¹ Eine Ausnahme bildet das Register des Erläuterungswortschatzes, wenn es zu einem onomasiologischen Umkehrwörterbuch ausgearbeitet wird. Dieser Typ von Wörterbuchregistern ist methodisch besonders schwierig, wie die Forschungsdiskussion zeigt (zuletzt Plate 1992 und Goebel/Lemberg/Reichmann 1995, 1–127). Entsprechend selten ist es bislang realisiert worden, im Bereich der historischen Lexikographie des Deutschen vgl. vor allem den ‚Neuhochdeutschen Index zum mittelhochdeutschen Wortschatz‘ von Koller/Wegstein/Wolf (1990, dazu Plate 1992) und die umfangreiche Probe eines onomasiologischen Umkehrwörterbuchs zum FWB bei Goebel/Lemberg/Reichmann (1995:53–127). Diese Probe, der anspruchsvollste Versuch in dieser Gattung überhaupt, bezieht sich auf die Strecke **a** bis **ausgang** des FWB und bietet für diese Strecke des Ausgangswörterbuchs nur die Artikel **a, A** bis **Ausgang** des Umkehrwörterbuchs. Welchen Umfang das vollständige Umkehrwörterbuch zum vollständigen FWB hätte, läßt sich daran ermessen. Eine größere Strecke hätte vermutlich manche methodischen Probleme noch deutlicher gemacht, zum Beispiel das der onomasiologischen Verweise (vgl. dazu Goebel/Lemberg/Reichmann 1995:35).

³² Der Gesichtspunkt des Wertes der Register als Mittel der lexikographischen Selbstkontrolle ist dabei unberührt.

kenden Aufwand eigener registrographischer Anstrengungen der Wörterbuchmacher realisiert werden kann, nämlich in der Form der Volltextsuche oder der Suche in Feldern. Für die Beurteilung dieser beiden Zugriffsweisen ist es nützlich, sich den Unterschied zu einer durch Register unterstützten Wörterbuchabfrage vor Augen zu halten. Volltextsuche ist in einem elektronischen Wörterbuchtext immer möglich, sie liefert jedoch erstens in der Regel zu einem großen Teil unspezifische Daten (während beim Registermachen bereits die Rohdaten einen spezifischen Filter passiert haben, den der Textauszeichnung fürs Register), zweitens sind diese Daten unstrukturiert (für die Strukturierung sorgt der zweite Schritt des Registermachens, die Formulierung der Registereinträge, ihre Hierarchisierung, die Einfügung von Verweisen usw.). Entsprechend gering ist der Nutzen einer Volltextsuche für anspruchsvollere Fragestellungen. Auch die Suche in Feldern liefert unstrukturierte Daten, doch sind in diesem Fall die Rohdaten bereits in gewisser Weise gefiltert, weil sie nur in bestimmten Teilen des Textes, den Feldern eben, erhoben wurden. Wenn das elektronische Wörterbuch mehr bieten soll als die Möglichkeit einer bloßen Volltextsuche, andererseits aber die – an sich wünschenswerte – Ausarbeitung von eigenen Zugriffsstrukturen (Registers) für bestimmte Datentypen nicht geleistet werden kann, dann sehen sich die Wörterbuchmacher also auf die Unterstützung von gezielten Abfragen durch die Strukturierung der Artikeltexte in *Felder* verwiesen.

Verhältnismäßig aufwendig kann selbst eine einfache Feldstrukturierung (z.B. Stichwort, Formteil, grammatische Angabe, Erläuterungstext, Belegtext) sein, wenn ein im Druck erschienenes Wörterbuch nachträglich digitalisiert und für die elektronische Publikation aufbereitet wird, weil automatisch nur das ausgezeichnet werden kann, was im zugrundeliegenden gedruckten Werk durch explizite Angaben oder durch typographisch eindeutige Merkmale abgesetzt ist (generell dazu vgl. Dummer/Michaelis/Schlaefer [1998]). Beispiele für Retrodigitalisierungsvorhaben mit unterschiedlich großem Ehrgeiz – und entsprechend geringerem oder höheren Zeit- und Kostenaufwand – hinsichtlich der inhaltlichen Strukturierung sind der elektronische Verbund der vorhandenen mittelhochdeutschen Wörterbücher (BMZ, LEXER, FINDEBUCH)³³ und das Projekt der Retrodigitalisierung der älteren Bände des DRW.³⁴

Weniger aufwendig ist die inhaltliche Strukturierung der Daten und damit die Vorbereitung für gezielte Recherchen in der elektronischen Fassung bei Wörterbüchern, die sich eines speziellen Programmsystems für die Unterstützung der Artikelarbeit bedienen, wie es das DRW seit Band IX (erschienen 1992–1996) tut³⁵ und zukünftig das neue Mittelhochdeutsche Wörterbuch (vgl. dazu oben unter 5). Denn schon die Programme, die verschiedene Aspekte der Artikelarbeit unterstützen, verlangen eine formal eindeutige Absetzung von bestimmten Elementen des Artikeltextes (z.B. des Stichworts, der Belegzitate, der Fundstellenangaben, des lexikographischen Kommentartextes), und die Absetzung beliebiger weiterer Datentypen (z.B. der verschiedenen Typen lexikographischen Kommentars wie: Angabe von Übersetzungsäquivalenten, von Bedeutungsverwandten oder Gegensatzwörtern, von räumlichen, zeitlichen oder textsortenspezifischen Gebrauchsbeschränkungen, von Konstruktionsangaben usw.) kann durch entsprechende formale Vorgaben in der Artikelarbeitsdatei unterstützt oder gefordert werden. Die Grenzen solcher vorbereitenden

³³ Zu den Problemen der strukturellen Auszeichnung für gezielte Recherchen vgl. Fournier (2000) und den Beitrag von Burch/Fournier in diesem Band.

³⁴ Vgl. dazu Speer (1998:15) mit ausdrücklicher Absetzung von dem bei den mittelhochdeutschen Wörterbüchern gewählten Digitalisierungsverfahren.

³⁵ Vgl. das Vorwort von Heino Speer zu Bd. IX.

Kennzeichnung bei der Artikelarbeit für spätere gezielte Feldabfragen im Wörterbuchtext wären dort überschritten, wo sie ein nachgeordnetes Anliegen zur Hauptsache machte und die Artikelarbeit selbst behinderte, indem sie z.B. die nötige Formulierungsfreiheit der Artikelverfasser durch schematische Vorgaben beschränkte oder ihre Aufmerksamkeit vom Artikelzusammenhang abzöge und stattdessen auf mögliche Registerzusammenhänge ablenkte.

7 Resümee

Der Beitrag erörtert Chancen und Perspektiven der computergestützten Erarbeitung eines historischen Belegwörterbuchs, wie sie sich den Verfassern nach fünfjähriger Vorbereitung für das neue *MITTELHOCHDEUTSCHE WÖRTERBUCH* darstellen. Die Ausarbeitung des Wörterbuchs wird voraussichtlich 2001 beginnen, innerhalb von 20 Jahren soll das in der Druckfassung auf vier Bände zu je ca. 1000 Seiten berechnete Werk abgeschlossen sein und dann an die Stelle der Vorgängerwörterbücher des 19. Jahrhunderts für den Quellenzeitraum 1050 bis 1350 treten können. Als jüngstes der größeren Vorhaben der historischen Beleglexikographie des Deutschen konnte das neue *MITTELHOCHDEUTSCHE WÖRTERBUCH* von vornherein vollständig auf EDV-Basis gestellt werden. Die Materialsammlung beruht auf einem für das Wörterbuch eingerichteten umfangreichen elektronischen Textarchiv mit Volltexten aller Quellen des Grundkorpus und auf einer aus den Vorgängerwörterbüchern kompliierten Lemmakandidatenliste, die rund 80.000 Artikelstichwörter umfaßt. Mit einem Programmsystem für die halbautomatische Lemmatisierung wird nach entsprechender Vorbereitung der Quellentexte aus diesen und einer eigenständigen Lemmatisierungskomponente, die aus der um die Wortformen der bearbeiteten Texte erweiterten Lemmatische besteht, das lemmatisierte Belegarchiv erzeugt. Auf seiner Grundlage und mit der Hilfe eines weiteren Systems von Programmen, die die Artikelarbeit von einer ersten vorbereitenden Sortierung der Belege bis hin zum fertig gesetzten Artikel unterstützen, wird das Wörterbuch ausgearbeitet werden. Die vollständig elektronisch gestützte Erarbeitung des Wörterbuchs erlaubt es, neben der als Lieferungswerk erscheinenden Druckfassung das Wörterbuch und seine Materialbasis mit überschaubarem zusätzlichem Aufwand auch für elektronische Nutzung mit weiteren Recherchemöglichkeiten zu publizieren.

Die im vorliegenden Beitrag am Beispiel des neuen *MITTELHOCHDEUTSCHEN WÖRTERBUCHS* erörterten Chancen der computergestützten Lexikographie liegen in der Möglichkeit, mit Rechnerhilfe die herkömmlichen Arbeitsgänge und ihre Abfolge in lexikographisch vorteilhafter Weise zu reorganisieren und zu rationalisieren. Ein großer Teil des herkömmlich mit dem Abschreiben, Korrigieren und Lemmatisieren einzelner Belegstellen verbundenen Arbeitsaufwands kann eingespart werden bei einmaliger Bereitstellung sorgfältig korrigierter und für die weitere Verarbeitung präparierter elektronischer Volltexte der Wörterbuchquellen und dem Einsatz eines halbautomatischen Lemmatisierungsverfahrens. Dies gestattet es insbesondere, die Belegauswahl, die herkömmlich bereits bei der Exzerption vorgenommen werden muß, zum größten Teil dem Artikelverfasser zu überlassen, der sie lexikographisch begründet vornehmen kann. Die Ausarbeitung der Artikel wird unterstützt von einem Programmsystem, das die mechanischen Sortier-, Schreib- und Kontrollarbeiten zum größten Teil automatisch erledigt, so daß der Artikelverfasser sich ungestörter der Interpretation des Belegmaterials und dem Formulieren des lexikographi-

schen Befundes widmen kann. Zugleich sind in der Artikeldatei durch ihre obligatorische formale Anlage bestimmte Mindestinformationen über die Struktur des Wörterbuchtextes mitkodiert, die später gezielte Recherchen in der elektronischen Fassung des Wörterbuchs erlauben. Im weiteren Sinne zu den Chancen der computergestützten Lexikographie zu zählen ist die Tatsache, daß im Zuge der Vorbereitung und Ausarbeitung des Wörterbuchs Materialien und Hilfsmittel entstehen, die zusätzlich zum Wörterbuch oder sogar unabhängig von ihm für andere linguistische und philologische Anliegen von Interesse sein können. Im Falle des neuen *MITTELHOCHDEUTSCHEN WÖRTERBUCHS* sind dies vor allem das elektronische Textarchiv, das elektronische Belegarchiv und schließlich ein wirkungsvolles Instrument für die halbautomatische Herstellung lemmatisierter Indices und Konkordanzen zu beliebigen mittelhochdeutschen Texten.

Die diskutierten Perspektiven sind solche, die sich für die Wörterbuchbenutzung aus der Möglichkeit der elektronischen Publikation des Wörterbuchs ergeben und Fragen der Wörterbuchkonzeption aufwerfen. Diese betreffen zum einen das Verhältnis zwischen dem ausgearbeitetem Wörterbuch und den ihm zugrundeliegenden Materialien (vor allem Text- und Belegarchiv), wenn diese gemeinsam mit dem Wörterbuch elektronisch publiziert werden und die Möglichkeit besteht, sie direkt aus einem Wörterbuchartikel heraus anzusprechen. Zum anderen stellt sich die Frage, in welchem Umfang das Wörterbuch neben seiner eigentlichen Aufgabe auch nicht-usuellen Benutzungsanliegen, die Register oder registerähnliche Zugriffsmöglichkeiten verlangen, entgegenkommen soll.

8 Literatur

(a) Wörterbücher (gedruckte)

AWB = *ALTHOCHDEUTSCHES WÖRTERBUCH*. Auf Grund der von Elias von Steinmeyer hinterlassenen Sammlungen im Auftrage der sächsischen Akademie der Wissenschaften zu Leipzig. Hgg. Elisabeth Karg-Gasterstädt, Theodor Frings. Bisher 4 Bde. Berlin: Akademie-Verlag 1968ff.

BMZ = *MITTELHOCHDEUTSCHES WÖRTERBUCH*. Mit Benutzung des Nachlasses von Georg Friedrich Benecke ausgearbeitet von Wilhelm Müller und Friedrich Zarncke. Nachdruck der Ausg. Leipzig 1854–1866 mit einem Vorwort und einem zusammengefaßten Quellenverzeichnis von Eberhard Nellmann sowie einem alphabetischen Index von Erwin Koller, Werner Wegstein und Norbert Richard Wolf. 5 Bde. Stuttgart: Hirzel 1990.

DRW = *DEUTSCHES RECHTSWÖRTERBUCH*. Wörterbuch der älteren deutschen Rechtssprache. Hg. Heidelberger Akademie der Wissenschaften. Bisher 10 Bde. Weimar: 1930ff.

¹**DWB** = *DEUTSCHES WÖRTERBUCH* von Jacob Grimm und Wilhelm Grimm. 16 Bd. in 32 Bdn. und Quellenverzeichnis. Leipzig 1854–1971.

²**DWB** = *DEUTSCHES WÖRTERBUCH* von Jacob Grimm und Wilhelm Grimm. Neubearbeitung hg. von der Akademie der Wissenschaften der DDR in Zusammenarbeit mit der Akademie der Wissenschaften zu Göttingen. Bisher Bd. 1, 2 (Lfg. 1–3), 6, 7 und 8 (Lfg. 1–7). Leipzig: Hirzel 1983ff.

FINDEBUCH = *FINDEBUCH ZUM MITTELHOCHDEUTSCHEN WORTSCHATZ*. Mit einem rückläufigen Index. Hg. Kurt Gärtner, Christoph Gerhardt, Jürgen Jaehrling, Ralf Plate, Walter Röll, Erika Timm (Datenverarbeitung: Gerhard Hanrieder). Stuttgart: Hirzel 1992.

FWB = *FRÜHNEUHOCHDEUTSCHES WÖRTERBUCH*. Hg. Ulrich Goebel, Oskar Reichmann. Begründet von Robert R. Anderson, Ulrich Goebel und Oskar Reichmann. Bisher Bd. 1, Bd. 2, Bd. 3 (Lfg. 1–3), Bd. 4 (Lfg. 1+2) und Bd. 8 (Lfg. 1). Berlin: de Gruyter 1986ff.

Koller, Erwin/Wegstein, Werner/Wolf, Norbert Richard (1990): *Neuhochdeutscher Index zum mittelhochdeutschen Wortschatz*. Stuttgart: Hirzel.

LEXER = Mittelhochdeutsches Handwörterbuch von Matthias Lexer. Zugleich als Supplement und alphabetischer Index zum Mittelhochdeutschen Wörterbuche von Benecke-Müller-Zarncke. Nachdruck der Ausg. Leipzig 1872–1878 mit einer Einleitung von Kurt Gärtner. 3 Bd. Stuttgart: Hirzel 1992.

Stackmann, Karl (1990), Wörterbuch zur Göttinger Frauenlob-Ausgabe. Unter Mitarbeit von Jens Haustein. Göttingen: Vandenhoeck & Ruprecht (=Abhandlungen der Akademie der Wissenschaften in Göttingen, Philologisch-Historische Klasse; Folge 3, Nr. 186).

TASCHENLEXER = Matthias Lexer (1992): MITTELHOCHDEUTSCHES TASCHENWÖRTERBUCH. Mit den Nachträgen von Ulrich Pretzel. 38., unveränderte Aufl. Stuttgart: Hirzel. – Vgl. auch: Matthias Lexer: Mittelhochdeutsches Taschenwörterbuch in der Ausgabe letzter Hand. 2. Nachdruck der 3. Aufl. von 1885 mit einem Vorwort von Erwin Koller, Werner Wegstein und Norbert Richard Wolf und einem biographischen Abriß von Horst Brunner. Stuttgart: Hirzel 1992.

(b) Forschungsliteratur (gedruckte)

Baumgarthe, Susanne (2000): Vorstellung des Probeartikels *schœne*. – In: Gärtner, Kurt/Grubmüller, Klaus (2000).

Burch, Thomas/Fournier, Johannes/Gärtner, Kurt (1998): Mittelhochdeutsche Wörterbücher auf CD-ROM und im Internet. Der Einsatz von SGML in der Retrodigitalisierung lexikographischer Standardwerke. – In: Akademie-Journal. Mitteilungsblatt der Konferenz der deutschen Akademien der Wissenschaften 1998/2, S. 17–24.

Burch, Thomas/Fournier, Johannes (2000): Zur Anwendung der TEI-Richtlinien auf die Retrodigitalisierung mittelhochdeutscher Wörterbücher (in diesem Band)

Dummer, Sven/Michaelis, Frank/Schlaefer, Michael (1988): Zur Digitalisierung historischer Wörterbücher. – In: Lexikos 8 (Afrilex-Reeks/Series 8:1998), 194–222.

Fournier, Johannes (2000): Digitale Dialektik. Chancen und Probleme mittelhochdeutscher Wörterbücher in elektronischer Form. – In: Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. 85–108. Hg. v. Herbert Ernst Wiegand. – Tübingen: Niemeyer.

Gärtner, Kurt (2000): Quellenauswahl, Arbeitsverfahren, Zeitplanung und Kooperation. – In: Gärtner, Kurt/Grubmüller, Klaus (2000).

– und Grubmüller, Klaus (Hgg.) (2000): Ein neues Mittelhochdeutsches Wörterbuch. Prinzipien, Probeartikel, Diskussion. (= Nachrichten der Akademie der Wissenschaften in Göttingen. I Philologisch-historische Klasse. Jahrgang 2000, Nr. 1).

Gärtner, Kurt/Kühn, Peter (1998): Indices und Konkordanzen zu historischen Texten des Deutschen. Bestandsaufnahme, Typen, Herstellungsprobleme, Benutzungsmöglichkeiten. – In: Werner Besch u.a. (Hgg.) Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. 1. Halbband. (Handbücher zur Sprach- und Kommunikationswissenschaft 2). 2., vollst. neubearb. u. erw. Aufl. Berlin/New York, 715–742.

Goebel, Ulrich/Lemberg, Ingrid/Reichmann, Oskar (1995): Versteckte lexikographische Information. Möglichkeiten ihrer Erschließung dargestellt am Beispiel des Frühneuhochdeutschen Wörterbuchs. – Tübingen: Niemeyer (= Lexicographica, Series Maior 65).

Lemberg, Ingrid (1996): Die Belegexzertion zu historischen Wörterbüchern am Beispiel des Frühneuhochdeutschen Wörterbuchs und des Deutschen Rechtswörterbuchs. – In: Herbert Ernst Wiegand (Hg.): Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger Lexikographischen Kolloquium, 83–102. Tübingen: Niemeyer.

Plate, Ralf (1992): Onomasiologische Umkehrlexikographie auf dem Prüfstand. Anlässlich des Erscheinens von: Erwin Koller/Werner Wegstein/Norbert Richard Wolf: Neuuhochdeutscher Index zum mittelhochdeutschen Wortschatz. – In: Zeitschrift für Dialektologie und Linguistik 59, 312–329.

– (1997): Votum für ein kleines Belegwörterbuch zur mittelhochdeutschen Klassik. Zugleich ein Beitrag zur Kritik der beiden großen mittelhochdeutschen Wörterbücher und des TASCHENLEXER. In: Energeia 23 (Tokyo), 57–97.

- (2000a): Zum Lemmestatus und Buchungsort der trennbaren Partikelverben im neuen Mittelhochdeutschen Wörterbuch. – In: Gärtner, Kurt/Grubmüller, Klaus (Hgg.): Ein neues MITTELHOCHDEUTSCHES WÖRTERBUCH. Prinzipien, Probeartikel, Diskussion.
 - (2000b): „Erster schneller Zugriff“ oder Fehlgriff? Zum alten und zu einem neuen mittelhochdeutschen Wörterbuch für Studierende. – In: Zeitschrift für Dialektologie und Linguistik 67 [im Druck].
 - und Recker, Ute (2000): EDV für Wörterbuchzwecke und neue lexikographische Arbeitsweisen. Erfahrungen beim Aufbau des elektronischen Text- und Belegarchivs für das MITTELHOCHDEUTSCHE WÖRTERBUCH [im Druck].
- Recker, Ute/Sappler, Paul (1998): Aufbau des maschinenlesbaren Text- und Belegarchivs für das MITTELHOCHDEUTSCHE WÖRTERBUCH. – In: Rudolf Große (Hg.): Bedeutungserfassung und Bedeutungsbeschreibung in historischen und dialektologischen Wörterbüchern. Beiträge zu einer Arbeitstagung der deutschsprachigen Wörterbücher, Projekte an Akademien und Universitäten vom 7. bis 9. März 1996 anlässlich des 150jährigen Jubiläums der Sächsischen Akademie der Wissenschaften zu Leipzig. (=Abhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig. Philologisch-historische Klasse. Bd. 75, H. 1). 249–253. Stuttgart/Leipzig.
- Reichmann, Oskar (1986): Lexikographische Einleitung. – In: FWB [s. oben unter (a)]. Bd. 1, 10–164.
- (1990): Formen und Probleme der Datenerhebung I: Synchronische und diachronische historische Wörterbücher. – In: Franz Josef Hausmann u.a. (Hgg.): Wörterbücher: Ein internationales Handbuch zur Lexikographie. 1588–1611. Berlin/New York.
- Sappler, Paul (1991): Strukturierungs- und Auswahlhilfen bei Autorwörterbuch und Sprachwörterbuch. – In: Eijirō Iwasaki (Hg.): Begegnung mit dem „Fremden“: Grenzen – Traditionen – Vergleiche. Akten des VII. Internationalen Germanisten-Kongresses, Tokyo 1990. Bd. 4, 277–281. München.
- (2000a): Probleme literarhistorischer und inhaltlicher Erschließung durch Register. (Im Druck.)
 - (2000b): Prinzipien des EDV-Konzepts. – In: Gärtner, Kurt/Grubmüller, Klaus (2000).
- Sappler, Paul/Schneider-Lastin, Wolfram (1991): Ein Wörterbuch zu Gottfrieds ‚Tristan‘. – In: Kurt Gärtner, Paul Sappler, Michael Trauth (Hgg.): Maschinelle Verarbeitung altdeutscher Texte IV. Beiträge zum Vierten Internationalen Symposium, Trier 28. Februar bis 2. März 1988. 19–28. Tübingen.
- Speer, Heino (1996): Vorwort. – In: DRW (s.o. unter [a]). Bd. IX, III–VII.
- (1998): Ein Wörterbuch, die elektronische Datenverarbeitung und ihre Folgen. – In: Akademiejournal 2/98. Mitteilungsblatt der Konferenz der Deutschen Akademien der Wissenschaften. Mainz. 11–16.
- Tao, Jingning (2000): Vorstellung des Probeartikels *nēmen* (B). – In: Gärtner, Kurt/Grubmüller, Klaus (2000).
- Wawer, Annne (2000): Vorstellung des Probeartikels *nēmen* (A). – In: Gärtner, Kurt/Grubmüller, Klaus (2000).
- Wiegand, Herbert Ernst (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband. Berlin/New York: de Gruyter.

(c) Wörterbücher und Literatur im WWW

- Homepage des neuen MITTELHOCHDEUTSCHEN WÖRTERBUCHS im WWW. April 2000. <http://gaer27.uni-trier.de/MhdWB>.
- Burch, Thomas/Fournier, Johannes/Gärtner, Kurt (1999): Mittelhochdeutsche Wörterbücher im Verband [miteineinander verknüpfte digitale Versionen von BMZ, LEXER und Findebuch; diese vgl. oben unter (a)]. April 2000. <http://gaer27.uni-trier.de/MWV-online/MWV-online.html>.
- Lemberg, Ingrid (1998): DRW-digital. Neue Wege zur versteckten lexikographischen Information. Vortrag gehalten auf dem 7th International Symposium on Lexicography, Kopenhagen, April 1998. April 2000. <http://www.rzuser.uni-heidelberg.de/~q63/kopenhagen.html>.

- Plate, Ralf (1997): [wie oben unter (b)]. April 2000. <http://gaer27.uni-trier.de/MhdWB/plate97.pdf>.
- Speer, Heino (1999): Digitale Version des DRW [dieses vgl. oben unter (a)]. April 2000. <http://www.uni-heidelberg.de/institute/sonst/adw/drw>.
- (1998): [wie oben unter (b)]. April 2000. http://www.rzuser.uni-heidelberg.de/~cd2/drw/publikat/speer_akadjourn.htm.

*Ralf Plate, Trier
Ute Recker, Trier*

