

Ein Wörterbuch, die elektronische Datenverarbeitung und die Folgen

[Gedruckt: Akademie-Journal 2/98 S. 11-16]

Heino Speer

Deutsches Rechtswörterbuch

Heidelberger Akademie der Wissenschaften

Ein Wörterbuch ...

In den Geisteswissenschaften gibt es viele Forschungsvorhaben, die für die Anwendung der elektronischen Datenverarbeitung prädestiniert sind. Dies gilt in verschiedenem Maße für die unterschiedlichen Wissenschaftsgebiete, aber auch für die unterschiedlichen Stadien der Projektbearbeitung. Wo immer auch das Gedächtnis der Wissenschaftler einer Ergänzung durch das Gedächtnis der Maschine bedarf, dort kann die EDV sinnvoll eingesetzt werden. Ob es um die in der ganzen Welt verstreuten Bruchstücke von antiken Inschriften geht, die durch eine Computersimulation auf dem Bildschirm zu kompletten Inschriften zusammengefügt und damit erstmals korrekt gelesen werden sollen [FN. 1], oder aber um die Erschließung großer Textmengen unter bestimmten Kriterien - überall dort kann das menschliche Gedächtnis durch den Computer unterstützt werden. Bei neuen Forschungsvorhaben beginnt der Einsatz der EDV bereits bei dem Sammeln der Quellenbasis und führt über die verschiedenen Bearbeitungsschritte bis hin zum fertigen Layout der Veröffentlichung. Was aber ist mit den nicht allzu seltenen geisteswissenschaftlichen Projekten, die unter ganz anderen technischen Voraussetzungen angetreten sind und eine mehr oder weniger lange Geschichte hinter sich haben, in denen der Computer noch keine Rolle spielen konnte?

Ich will versuchen, die Schwierigkeiten, die Herausforderung und den Gewinn durch den Einsatz der EDV in einem klassischen Langzeitprojekt innerhalb der deutschen Geisteswissenschaft zu schildern. Seit fünfundzwanzig Jahren leite ich die Forschungsstelle des Deutschen Rechtswörterbuchs [FN. 2] bei der Heidelberger Akademie der Wissenschaften und bin auch für die Computerisierung dieses Wörterbuchs seit 1986 verantwortlich und habe sie in weiten Teilen alleine durchgeführt.

Über das Deutsche Rechtswörterbuch wurde bereits im ersten Heft des Akademie-Journals von dem Vorsitzenden seiner Kommission, Professor Adolf Laufs, berichtet. Hier sei daher nur wiederholt, daß das Deutsche Rechtswörterbuch (DRW) ein historisches Wörterbuch der älteren deutschen (besser gesagt: westgermanischen) Rechtsprache vom Beginn der schriftlichen Überlieferung in der Spätantike bis in das 18. und frühe 19. Jahrhundert ist. Das Wörterbuch wurde 1896/97 von der damaligen Preußischen Akademie der Wissenschaften in Berlin begründet, aber von Beginn an in der einzigen Arbeitsstelle in Heidelberg bearbeitet. Durch die politische Geschichte dieses Jahrhunderts gelangte es in den Fünfzigerjahren in die Obhut der Heidelberger Akademie. Im Verlauf der Jahrzehnte wurden bis heute in neun Bänden und den ersten beiden Doppelheften von Band X die rechtlich relevanten Wörter des Frankolatinischen, Langobardischen, Altenglischen, Altsächsischen, Altfriesischen, Althochdeutschen, Mittelniederländischen, Mittelniederdeutschen, Mittelhochdeutschen und Frühneuhochdeutschen bis zum Neuhochdeutschen in 77.816 alphabetisch geordneten Wortartikeln von *Aachenfahrt* bis zum Artikel *Pfaffenfürst* dargestellt. Das Gesamtwerk wird sechzehn Bände umfassen und zwischen den Jahren 2030 und 2036 abgeschlossen sein. Um die typische - und wichtige - Frage vorweg zu beantworten: Durch den Einsatz der Elektronischen Datenverarbeitung - unterstützt durch andere Straffungsmaßnahmen - konnte die Bearbeitungszeit je Band von elf (Band VI und VII) über sieben (Band VIII) auf fünf Jahre (Band IX

und voraussichtlich die weiteren Bände) herabgesetzt werden. Seit 1986 nimmt die elektronische Datenverarbeitung im lexikographischen Alltag des DRW einen immer höheren Stellenwert ein. Der erste Schritt in diese Richtung hatte Folgen, die heute nicht nur die wissenschaftliche Arbeit am Wörterbuch prägen, sondern auch gravierende Konsequenzen für das Verhältnis des Wörterbuchs zum Verlag und zu anderen Wörterbüchern haben. Es steht damit überdies in einer wissenschaftlichen Öffentlichkeit, die sich durch die Kommunität der Wissenschaftler im Internet erheblich verbreitern wird.

die elektronische Datenverarbeitung ...

Ich will hier nicht näher auf die Entwicklung der EDV-Anwendung im DRW eingehen. Der gegenwärtige Stand der Nutzung der EDV muß aber jeweils kurz angedeutet werden, um die Probleme und Chancen deutlich zu machen, die sich jetzt und für die nähere Zukunft aus der computergestützten Herstellung eines Wörterbuchs ergeben. Dabei lassen sich verschiedene Stufen sowohl bei der Erarbeitung eines Wörterbuchs als auch bei dem Einsatz der EDV unterscheiden: Dies sind die Grundlagen der Arbeit, die Erstellung der Wortartikel, die Publikation des Wörterbuchs und schließlich die Nutzbarmachung des in solchen Forschungsstellen versammelten Wissens. Hier freilich genügt die Bemerkung, daß das DRW in einer Datenbank innerhalb eines Netzwerkes erstellt wird und der Verlag Dateien erhält, in denen das vollständige Layout jeder Seite ohne die Möglichkeit oder Notwendigkeit einer weiteren Bearbeitung gespeichert ist. Freilich ist diese Lösung aus den personellen Ressourcen des DRW erwachsen und auf die besonderen Bedürfnisse des DRW zugeschnitten. Auf andere Wörterbücher kann sie nicht ohne weiteres übertragen werden. Auf die Konsequenzen aus den beiden letzten Sätzen werde ich am Schluß noch einmal eingehen.

Die Quellenbasis und die EDV

Die Herstellung eines historischen Wörterbuchs beruht traditionell auf der Auswertung seiner Belege. Das Zettelarchiv des DRW besteht aus über 2,2 Millionen Belegzetteln, die in jahrzehntelanger Sammeltätigkeit nicht nur in Deutschland, sondern auch in den übrigen mitteleuropäischen Ländern aus Texteditionen, Urkundenwerken aller Art und teilweise auch aus nur handschriftlich vorliegenden Texten exzerpiert wurden. Exzerption heißt hier: Herausschreiben zumindest des Stichwortes, der Quellenbezeichnung und der genauen Fundstelle, manchmal auch der Datierung, und manchmal eines Textausschnittes auf einen Zettel in einem bestimmten Format, der in der Heidelberger Forschungsstelle alphabetisch in Pappkästen eingeordnet wurde. Die Gesamtzahl der Kästen beträgt etwa 1.350. Diese Materialbasis wird in bislang ungeahnter Weise durch die elektronische Datenverarbeitung erweitert.

Jeder Belegtext, der für einen Wortartikel in die Datenbank eingegeben wird, enthält außer dem Stichwort noch eine Vielzahl anderer Wörter, die für die Zitierung in anderen Wortartikeln in Frage kommen. Diese treffend und erfreulich unwissenschaftlich als „Inzuchtbelege“ bezeichneten Mehrfachverwendungen eines Beleges werden in der allgemeinsprachlichen Lexikographie manchmal nicht gern gesehen, weil sie die Repräsentativität der Belegauswahl zu verzerren drohen. In einem Wörterbuch der Rechtssprache bestehen diese Bedenken nicht: Eine Legaldefinition des Diebstahls beispielsweise kann nicht nur, sondern muß sogar in vielen anderen Wortartikeln zitiert werden.

Die Erweiterung der Quellenbasis beschränkt sich freilich nicht auf Belegtexte aus dem Wörterbuch selbst. Eine bislang nicht näher bestimmbare Anzahl von historischen Texten, die für das DRW und andere Wörterbücher als Quelle in Frage kommen, ist entweder für Verlagspublikationen oder für private oder

sonstige Forschungsarbeiten maschinenlesbar gemacht worden. Diese Texte werden - sofern bekannt und verfügbar - in einen besonderen Teil der Datenbank des DRW integriert, so daß jedes Wort dieser Texte in einer alphabetischen Liste abrufbar ist.

Maschinenlesbare Texte stellen, ob sie nun als Volltexte oder als Belegtexte erfaßt worden sind, nur einen Teil der elektronisch verfügbaren Quellenbasis dar. Der andere Teil besteht aus Graphikdateien, genauer gesagt: Faksimiles von Quellentexten. Eine Reihe von Quellen des DRW - und dies ist bei anderen Wörterbüchern ebenso - sind frühe Drucke des 16. bis 18. Jahrhunderts, die heute nicht mehr ohne weiteres aus den klimatisierten Räumen der großen Bibliotheken ausgeliehen werden. Eine effektive Arbeit mit Quellenwerken, aus denen alle paar Tage ein Belegtext entnommen werden muß, ist aber nur dann möglich, wenn das betreffende Werk in unmittelbarer Nähe des lexikographischen Arbeitsplatzes verfügbar ist. Und auch das schon bedeutet einen Zeitaufwand, der sich in der täglichen Arbeit summiert. Der Lexikograph muß eine Quellensigle verifizieren, die oft vor etwa einhundert Jahren auf den Archivzettel geschrieben worden ist, er muß die Signatur dieser Quelle herausfinden, das Buch an seinem Standort suchen, gegebenenfalls bestellen, muß es an seinem Schreibtisch aufschlagen und die Fundstelle suchen sowie schließlich den Belegtext abschreiben.

Nicht unbedingt hierfür war das zusätzliche Modul unserer Datenbank gedacht - das FAUST-Bildarchiv. Aber es ist hervorragend geeignet, die lexikographische Arbeit in diesem Bereich zu erleichtern. Mit einem normalen Scanner, der Bilder in Graphikdateien umwandelt, können natürlich auch Buchseiten verarbeitet werden. Diese Buchseiten, sind sie einmal gescannt worden, stehen jedem Lexikographen, der sie irgendwann einmal benötigt, auf Tastendruck zur Verfügung, und er kann den gewünschten Text von dem Faksimile auf seinem Bildschirm abschreiben. *Abschreiben* - es ist wichtig, sich dies zu vergegenwärtigen, denn bei den alten Schriften versagen Zeichenerkennungsprogramme (OCR) häufig, so daß aus diesen Faksimiles kaum maschinenlesbare Texte hergestellt werden können. In der Datenbank des DRW waren im Februar 1998 rund 3.800 Faksimiles (von Einzel- oder Doppelseiten) integriert, die meist während der täglichen Arbeit gescannt wurden und so einen Thesaurus bilden, der die Arbeit der zukünftigen Lexikographen beschleunigen wird. Derartige Faksimiledateien können auch von Mikrofilmen hergestellt werden. Voraussetzung hierfür ist ein Mikrofilmscanner, den die Heidelberger Akademie der Wissenschaften für alle Forschungsstellen erworben hat und der in die EDV des DRW integriert ist. In nächster Zukunft sollen etwa 20.000 (Doppel-) Seiten in unsere Datenbank importiert werden. Wichtigste Quellenwerke der Rechtsgeschichte stehen dann ohne den Griff zu dem wertvollen Original in Sekundenschnelle für die Auswertung zur Verfügung.

Lexikographischer Arbeitsplatz: Von der Schreibmaschine zum Computer

An diesem Beispiel wird besonders deutlich, in welchem Maße sich der Arbeitsplatz eines Lexikographen in den letzten Jahren verändert hat. Zu Beginn meiner Tätigkeit in den Siebzigerjahren wurden alle Belege zu einem Wortartikel mit der Hand oder Schreibmaschine auf DIN-A4-Blätter geschrieben. Wurde die Artikelstruktur verändert, so mußten diese Blätter auseinandergeschnitten und die Belege auf andere Blätter aufgeklebt werden. Überprüfungen der Beleginterpretation oder auch nur der Genauigkeit der Abschrift erforderten den Einblick in die Quelle und die Korrektur im Manuskript. Aus dem vielfach veränderten und korrigierten Manuskript wurde die Vorlage für den Setzer mit den Auszeichnungen der Schriftarten hergestellt; der Setzer wiederum stellte die Druckfahnen her, die in langwierigen Korrekturvorgängen in der Forschungsstelle bis zum Imprimatur bearbeitet wurden. Jeder einzelne dieser Arbeitsschritte führte häufig zu Fehlern, die bis zum Druck unentdeckt bleiben konnten und heute noch für ein lexikographietypisches Lebensgefühl sorgen, das dem des Reiters über den Bodensee vergleichbar ist. Keine Veränderung der Arbeitstechnik kann alle Fehler verhindern. Aber allein dadurch, daß die Texte nur einmal abgeschrieben werden müssen, verringert sich die Fehlerquote bei diesem Arbeitsgang erheblich und steigt andererseits die Chance, dennoch vorhandene Fehler zu entdecken. Hierzu hilft auch

die Möglichkeit, die Quellentexte als Faksimile auf den Bildschirm zu holen. Die Autopsie der Quelle bzw. der Textedition ist immer der sicherste Weg der Vergewisserung, der aber unter dem Druck des Manuskriptabschlusses nicht immer gewählt werden kann, wenn dazu erst ein Buch bestellt und eingesehen werden muß. Der Tastendruck, mit dem die entsprechende Buchseite auf den Bildschirm geholt wird, steigert sicherlich die Qualität des fertigen Wörterbuchs.

Auch in anderer Weise erleichtert die moderne Technik die Herstellung des Wörterbuchs. Da alle Teile der Wortartikel in Erfassungsmasken eingegeben werden, die bereits eine Vorformatierung der notwendig oder fakultativ dazugehörigen Informationen enthalten, wird der Lexikograph von dem Nachdenken über triviale Anforderungen seiner Arbeit wie beispielsweise die Frage nach dem richtigen Schriftgrad einer bestimmten Informationseinheit entlastet und kann sich den schwierigen inhaltlichen und interpretatorischen Problemen seiner Wörter ungeteilt widmen. Hinzu kommt eine permanente Verfügbarkeit vieler Informationen in der Datenbank, die in der bisherigen Wörterbucharbeit häufig das nur durch langjährige Tätigkeit erwerbbar Spezialwissen über bestimmte Quellen, Zitierweisen, interne Regeln usw. darstellte, das von Anfängern mühsam von den älteren Mitarbeitern erfragt und dann gelernt werden mußte. Bei den Artikelstrukturen macht sich ebenso die Vernetzung aller Informationen bemerkbar. Ein Doppelklick mit der Maus auf ein Wort im Feld für Bedeutungserklärungen zeigt sofort alle Artikel, in denen dieses Wort bereits für Erklärungstexte verwendet wurde. Das Wort *Diebstahl* beispielsweise wird in Erklärungstexten zu den Wörtern *Kirchraub*, *Landschaden*, *Mauserei*, *Milchstehlen*, *Mühdiebe*, *Mundraubung*, *Nachtdiebe*, *Nascheinung*, *Notraub* und *Obstdieberei* (neben 18 weiteren Wörtern) benutzt.

Herstellung des Wörterbuchs

Die Herstellung des gedruckten Werkes hat sich von der Druckerei und dem Verlag auf die Forschungsstelle verlagert. Dies ist eine Verlagerung von hochwertiger Arbeit auf die Forschungsstellen, die durch die moderne Technik zwar ermöglicht worden ist, aber den Wissenschaftlern Zusatzkenntnisse auf den verschiedensten Gebieten der Druckgestaltung und Drucktechnik abverlangt. Der Weg, wie nach allen redaktionellen Arbeiten das endgültige Layout erzeugt und der Wörterbuchtext mit allen Formatierungen dem Verlag übermittelt wird, braucht hier nur kurz gestreift zu werden: Aus dem Retrievalsystem FAUST heraus werden alle Wortartikel für eine Wörterbuchlieferung in eine Textdatei exportiert. Diese Datei wird mit relativ komplizierten, auf das Layout des DRW hin programmierten Makros so weiterverarbeitet, daß eine Eingabedatei für das Textsatzsystem TeX entsteht. Die Bearbeitung mit TeX erzeugt wiederum eine Datei, die den Wörterbuchtext im zweispaltigen Satz mit dem endgültigen Zeilenfall, lebenden Kolumnentiteln, einer fortlaufenden Spalten- und Bogenzählung sowie gegebenenfalls den typographischen Markierungen für einen neuen Band oder den Beginn eines neuen Buchstabens enthält (diese für mich sehr schwierige Layoutformatierung konnte ich nicht selbst vornehmen, sie ist für das DRW entwickelt worden). Diese Datei wird dann durch ein Konvertierungsprogramm so umgewandelt, daß jede Druck- oder Lichtsatzanlage mit den weitverbreiteten PostScript-Schriften einen korrekten Ausdruck oder eine korrekte Belichtung erzeugen kann. Der Verlag erhält diese Datei; seine Funktion in diesem Stadium beschränkt sich darauf, die Filme für den Druck zu erzeugen und die Herstellung und Auslieferung der Druckexemplare vorzunehmen.

Hypertextualität des Wörterbuchs

Dies ist der Weg zur Publikation eines *Druckwerkes*. Neun Bände Deutsches Rechtswörterbuch, die in Halbleder gebunden vor mir stehen, lassen zumindest dem Bücherfreund deutlich werden, daß dies eine unverzichtbare Publikationsform bleiben muß. Es wird immer genügend viele Benutzungssituationen geben, in denen der Griff zum gedruckten Werk der Nutzung der EDV vorzuziehen ist (falls sie überhaupt genutzt werden kann). Und dennoch: Nachschlagewerke haben per se eine Infrastruktur, die erst im Zeitalter der Hypertextualisierung in allen Einzelheiten und im gesamten Umfang realisiert werden kann.

Schon in den frühesten gedruckten Wörterbüchern wird von einem Wortartikel auf den anderen verwiesen, rudimentär zwar und deutlich noch als zaghafter Versuch der Vernetzung von Wissen. Und in den späteren Wörterbüchern wird das Verweissystem zu einer ständigen Crux der Lexikographen. Denn so sinnvoll und nützlich die Verweise sind - niemand kann dafür garantieren, daß Verweise erfüllt werden, daß sie nicht durch eine Veränderung der Artikelstruktur ins Leere gehen und damit auch den Leser ins Leere laufen lassen. Dieser Gefahr unterliegen Verweise in einer Datenbank oder einer vergleichbaren Struktur wesentlich seltener als im gedruckten Wörterbuch. Hier zeigt sich der größte Unterschied zwischen einem Druckwerk und einer elektronischen Publikation: Der gedruckte Text ist nicht mehr veränderbar. Die Bleistiftanmerkungen in meinem Handexemplar des DRW, das ich von meinen Vorgängern übernommen habe, sprechen hier eine deutliche Sprache: Was hier handschriftlich verzeichnet ist, bleibt Arkanwissen der Forschungsstelle, wird nicht den Erwerbern des Druckwerkes mitgeteilt, oder wenn, dann höchstens als Nachtragsbändchen nach dem Abschluß des Werks, in dem jeweils nachgeschlagen werden müßte. Aber auch wenn der Text des Wörterbuchs maschinenlesbar vorliegt, wird ja beileibe nicht jede Veränderung an jeder relevanten Stelle zu einer vergleichbaren Veränderung führen. Ist der Text eine zeilenorientierte Datei, dann ist dies ebenso unmöglich wie in dem sequentiell geordneten Wörterbuch. Erst eine Datenbank bietet die Möglichkeit, daß jede Änderung in einem Objekt sofort in jedem anderen Objekt wirksam wird, auf das sich das veränderte Objekt bezieht. Um die abstrakte Ebene zu verlassen: Im ersten Band des DRW steht der Artikel *Arglist*, auf den in den späteren Bänden bei *List*, *Mut* und *Neufund* Bezug genommen wird. Der Bearbeiter des ersten Bandes hat dies nicht gewußt; der Bearbeiter des neunten Bandes kennt den Artikel *Arglist* des ersten Bandes und weiß um diesen Zusammenhang. Für das Druckwerk kann er auf den ersten Band verweisen. Dann ist ab dem neunten Band die Informationsvernetzung in das davor liegende Alphabet vorhanden, aber nicht für den Benutzer des ersten Bandes in das folgende Alphabet. Der Nutzer einer Datenbank allerdings hat jederzeit Zugriff auf die jeweils aktuellsten und umfassendsten Informationen und weiß, daß eine wichtige Information für den Artikel unter den Buchstaben L, M oder N zu finden ist.

Der Zugriff auf die in einem Wörterbuch vorhandenen Informationen ist das Ziel der Lexikographie. Wie können möglichst viele der Informationen eines Wörterbuchs dem Benutzer zumindest als Nachschlageangebot vorgelegt werden, wenn es inzwischen deutlich geworden ist, wieviel an Information in einem gedruckten Wörterbuch verborgen ist und verborgen bleibt? Die Lexikographie hat ihre eigenen Methoden hierzu entwickelt, hat Register verschiedenster Art erstellt und so Zusatzfragen ermöglicht, die der Benutzer selbst ohne die Durcharbeitung des gesamten Wörterbuchtexes garnicht stellen konnte.

Ein Wörterbuch allerdings, das die Möglichkeiten der EDV nutzt, hat hier ganz andere Formen der Informationsvermittlung zur Verfügung. Sie beruhen darauf, daß in einer Datenbank Informationen vernetzt werden können, daß Indices unterschiedlichster Art automatisch erstellt werden und daß Abfragen in unterschiedlichsten Kombinationen das gesamte Wörterbuch in Sekundenschnelle durchsuchen können. Von einem Wortartikel wird man nicht nur auf einen sinnverwandten Artikel verwiesen, sondern bekommt ihn sofort gezeigt und kann von diesem aus wiederum weiterverwiesen werden - ein potentiell unendliches Verweissystem, das freilich auch einen Benutzer fordert, der mehr wissen will als die Bedeutung *eines* Wortes (hier zeigt sich auch die sachlich-enzyklopädische Seite des DRW, das eben nicht nur Wortbedeutungen erklärt). Von einem Belegzitat kann man sofort alle verfügbaren Informationen zu der Textquelle erhalten, kann unter Umständen einen größeren Textausschnitt als elektronischen Volltext sehen oder sich die Faksimileseite, auf der dieser Text zu finden ist, auf seinem häuslichen Drucker ausdrucken. Die lexikographischen Entscheidungen der Wörterbuchschreiber werden überprüfbarer, der Benutzer wird stärker in die Interpretation der Belege miteinbezogen, weil er sie selbst überprüfen kann.

Natürlich ist dies nicht ohne größere Anstrengungen auf der Seite der Lexikographen möglich. Die Datenbank muß konzipiert und erstellt werden, sie muß mit Daten gefüllt werden, Faksimiles müssen gescannt und elektronisch lesbare Texte von Verlagen oder Editoren eingeworben oder selbst hergestellt werden. Diese Arbeit muß neben der alltäglichen Wörterbucharbeit geleistet werden, sonst wird die Akzeptanz eines Langfristunternehmens durch die forschungspolitischen Entscheidungsträger sehr schnell in Frage gestellt. Aber sie führt auch im Endeffekt zu einer Beschleunigung der Wörterbucharbeit selbst

und zu einem Informationsinstrumentarium, wie es in dieser Komplexität einerseits und in dieser leichten und intuitiven Bedienbarkeit andererseits bislang nicht möglich war. Die letzte Konsequenz hieraus ist, daß neben dem unverzichtbaren Druckwerk das Wörterbuch im Internet als Datenbank präsent sein muß, um allen potentiellen Benutzern in allen Teilen der Welt Informationen zu vermitteln, die sonst auf die wenigen Besucher der Forschungsinstitute und der großen Bibliotheken beschränkt wären.

Die Voraussetzung hierfür ist die Maschinenlesbarkeit des gesamten bislang bearbeiteten Wörterbuchs. Es gibt einige Projekte, die sich mit diesen Fragen beschäftigen und Wörterbücher wie das Mittelhochdeutsche Wörterbuch von Matthias Lexer in einer digitalen Fassung auf CD-ROM oder im Internet zur Verfügung stellen. Andere derartige Projekte sind geplant oder werden hinsichtlich des richtigen Weges einer Digitalisierung diskutiert. Auch das DRW hat derartige Pläne entwickelt und bemüht sich um deren Förderung mit öffentlichen Geldmitteln, zum Teil hat es auf diesem Gebiet auch schon Erfolge erzielt (die letzten drei Bände sind ohne Drittmittel innerhalb der Forschungsstelle aus Textdateien oder über den gedruckten Text maschinenlesbar gemacht und in die Datenbank integriert worden). Die von uns favorisierte Lösung unterscheidet sich allerdings von den Vorhaben, die bislang geplant oder durchgeführt worden sind. Es sind dies entweder die Erzeugung einer Textdatei durch zeichengenaues doppeltes Abschreiben (meist in Niedriglohnländern), bei der die beiden unabhängig voneinander erzeugten Dateien automatisch miteinander verglichen werden. Die andere Möglichkeit stellt die Nutzung eines OCR-Programmes dar. Meiner Kenntnis nach kämpfen diese Vorhaben insbesondere mit zwei Schwierigkeiten: Einmal sind die Druckvorlagen häufig so problematisch (Kleindruck und ein zu geringer Kontrast zwischen Druckzeichen und Druckhintergrund), daß die Fehlerkorrektur zu viel Zeit benötigen würde. Die andere Schwierigkeit haben beide Verfahren gemeinsam: Durch das Abschreiben oder maschinelle Einlesen entsteht jeweils eine Datei, in der allein typographische Auszeichnungen dazu dienen können, strukturelle Merkmale des Wortartikels zu definieren. Bei einem über mehrere Jahrzehnte hinweg entstandenen Wörterbuch sind die Verwendungsweisen der Typographie aber in der Regel so unterschiedlich, daß eindeutige Ergebnisse nicht zu erwarten sind.

Meine Vorgehensweise unterscheidet sich insofern davon, als zwar auf den Einsatz eines OCR-Programms verzichtet wird (die Fehlerquote war einerseits zu hoch und andererseits die Strukturierung des Ergebnisses so rudimentär, daß in mühsamsten Nacharbeitungsprozessen ein sogenanntes Wörterbuchparsing durchgeführt werden mußte), andererseits aber das Abschreiben des Wörterbuchtextes nicht in Niedriglohnländer verlegt wird, sondern in der Forschungsstelle selbst erfolgen soll. Entscheidend ist hierbei freilich, daß durch dieses Abschreiben keine Textdatei entstehen soll, sondern Objekte innerhalb der Datenbank - kurz gesagt, wissenschaftliche Hilfskräfte sollen das DRW von Band I bis Band VI in die bestehenden Erfassungsmasken unserer Datenbank eintragen und dabei unter Anleitung durch einen Lexikographen gleich die erforderlichen Vernetzungen innerhalb der Datenbank und damit innerhalb der Artikelstrukturen vornehmen. Als Ergebnis entsteht eine effiziente Datenbank als Nachschlageinstrument, in der keine Nachbearbeitung über ein Wörterbuchparsing oder die Erstellung einer auf typographischen Merkmalen beruhenden Wörterbuchgrammatik erforderlich ist.

Diese Datenbank kann entweder in das HTML-Format exportiert und damit ins Internet gestellt werden (dies ist für die 1.342 Wortartikel des Buchstaben *O* bereits erfolgt). Oder aber die Datenbank selbst wird auf einem WWW-Server zur Verfügung gestellt. Die Firma Land arbeitet bereits an diesem Projekt; nähere Angaben finden sich in dem oben genannten Aufsatz über das DRW und das Internet.

Wenn aber ein Wörterbuch seinen Datenbestand im Internet anbietet, so hat dies Folgen, die nicht nur auf dieses eine Wörterbuch beschränkt sein dürfen. Davon soll im letzten Abschnitt die Rede sein.

... und die Folgen

Die Möglichkeiten, die sich für Wörterbücher durch die Nutzung der EDV - und hier konkret: des Internet

- bieten, sind vermutlich bislang weder in ihrer Tragweite überall erkannt, geschweige denn realisiert worden. Auch das DRW steht erst am Anfang einer sinnvollen Nutzung der EDV. Der weitere Weg wird dadurch bestimmt sein müssen, daß die auf das eigene Wörterbuch zugeschnittene EDV-Lösung sozusagen gemeinschaftsfähig gemacht werden muß. Dies hat folgende Gründe: Bislang stellten Wörterbücher Insellösungen dar, sichtbar gemacht durch die Bandreihen in den Bibliotheken an den unterschiedlichsten Standorten. Wer nachschlagen wollte, tat dies in einem Wörterbuch des Althochdeutschen oder des Frühneuhochdeutschen, des Altfrisischen oder der Rechtssprache, je nach dem konkreten Nachschlageinteresse. Daß aber zu der besonderen Frage weitere Informationen auch in einem der nicht nachgeschlagenen Wörterbücher zu finden wären, blieb dem Benutzer in den meisten Fällen verborgen. Unser Nachschlageverhalten wird sich entscheidend durch die Vernetzungen innerhalb des Internet ändern. Wer einmal im Internet Informationen gesucht hat, weiß zwar, wie schwierig und zeitraubend dies sein kann, er weiß aber auch, zu welchen überraschenden Informationen es dabei kommen kann.

Und hier setzt die digitale Zukunft der Wörterbücher ein und an. Als Fernziel sollte ein Informationssystem zur deutschen Sprache entstehen, in dem sowohl die Grammatik und Rechtschreibung der Gegenwartssprache abgefragt werden können, als auch Informationen zu jeder historischen Sprachstufe des Deutschen und der verwandten Sprachen. Mag dies auch utopisch anmuten - die Realisierung dieser Vision setzt wenig mehr als die Bereitschaft aller einschlägigen Forschungsunternehmungen einschließlich ihrer Trägerorganisationen und der beteiligten Verlage zu einer Kooperation unter Hintanstellung der eigenen Bedenken voraus. Noch sind wenige Wörterbücher digitalisiert worden, und es bestehen auch sehr unterschiedliche Vorstellungen darüber, wie dies im Einzelfall zu geschehen habe. Aber je mehr Forschungseinrichtungen erkennen, daß die elektronische Datenverarbeitung zu einer Verbesserung der Wörterbuchqualität *und* zu einer Beschleunigung der Arbeit führen kann, desto mehr Wörterbücher werden den Weg ins Internet finden. Dann aber bedarf es vereinbarter Standards und Normen, die es ermöglichen, eine einheitliche Informationsstruktur einzurichten, die gleichermaßen für die deutsche Sprachwissenschaft wie für die große Zahl von Menschen, die sich für ihre Sprache unter den verschiedensten Aspekten interessieren, zur Verfügung steht.

Dies beginnt bei der Frage, wie elektronische Texte für eine große Benutzerzahl bereitgestellt werden können. Welche Kodierungen werden für die vielfältigen Sonderzeichen und Diacritica benutzt, wie kann man für unterschiedliche Systeme Wahlmöglichkeiten in der Darstellung der Zeichen anbieten? Wer verwaltet welche digitalen Texte, wer sorgt für die Regelung aller rechtlichen Fragen, die damit verbunden sind? Rechtsfragen sind häufig finanzielle Fragen, also: Wer entwickelt ein Abrechnungssystem, das es den Verlagen und Forschungseinrichtungen ermöglicht, ihre Leistungen gegen eine angemessene Gebühr zur Verfügung zu stellen? Und ist es denkbar, die Möglichkeiten des Internet so zu nutzen, daß auch kleinere Wörterbuchunternehmungen, denen die personellen Ressourcen für eine eigenständige EDV-Lösung fehlen, von den an anderer Stelle entwickelten Lösungen profitieren können? Wenn es großen Verlagsgruppen gelingt, über das Internet die unterschiedlichsten Redaktionen in einem Redaktionssystem zusammenzubinden und sie damit von dem Verwaltungsaufwand für proprietäre EDV-Lösungen zu befreien, so könnte dies auch für Wörterbücher gelten - wenn die Entscheidungsträger genügend Kooperationswillen über Länder- und Akademiegrenzen hinweg und verbunden mit einer betriebswirtschaftlichen Denkweise aufbringen können.

Die technischen und organisatorischen Probleme sind bereits gelöst oder lösbar, wenn auch noch nicht für den Bereich der deutschen Sprachwissenschaft. Wie kann es gelingen, die vielen unterschiedlichen Wissenschafts-, Wirtschafts- und Politikbereiche zu einer Zusammenarbeit zu motivieren, die die Isolation der wissenschaftlichen Teilbereiche aufbricht, und im Interesse der an unserer Sprache interessierten Menschen ein wirklich umfassendes Informationssystem zu demjenigen geistigen Bereich zu schaffen, der das Verbindende zwischen uns allen ist?

Die deutschen Akademien könnten hierin auf Grund ihrer Struktur, die die verschiedensten Wissensbereiche unter sich vereinigt, eine wichtige und führende Rolle übernehmen, die für alle beteiligten Forschungsvorhaben über die Nutzung verteilter Ressourcen (an Datenbeständen und Kompetenz gleichermaßen) zu einer Beschleunigung und Qualitätssteigerung führen würde.

Post scriptum: Dies war der Stand Anfang Mai 1998. Jetzt, Mitte Oktober desselben Jahres, haben Professor Rainer Kuhlen (Konstanzer Hypertextsystem), Professor Oskar Reichmann (Frühneuhochdeutsches Wörterbuch) und ich bei der DFG einen Förderantrag gestellt, durch den ein internetfähiges und datenbankbasiertes Lexikographisches Produktions- und Informationssystem (LPI) zunächst geschaffen und anhand der beiden beteiligten Wörterbücher erprobt werden soll. Gespräche mit Verlagen über die Integration verschiedener großer geisteswissenschaftlicher Nachschlagewerke und ihre kommerzielle Nutzung in einem solchen System lassen ein grundsätzliches Interesse hieran erkennen.

FN. 1: <http://www.uni-heidelberg.de/institute/sonst/adw/edh>

FN. 2: <http://drw-www.adw.uni-heidelberg.de/drw/>

FN. 3: Vgl. hierzu Heino Speer, DRW to FAUST. Ein Wörterbuch zwischen Tradition und Fortschritt. In: *Lexicographica* 10 (1994) 171-213, und: Ingrid Lemberg, Sybille Petzold, Heino Speer, Der Weg des Deutschen Rechtswörterbuchs in das Internet. In: *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*. In: *Lexicographica. Series Maior* 84 (1998) 262-284.