

Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie

Angelika Storrer

Erscheint in: Lemberg, Ingrid; Schröder, Bernhard; Storrer, Angelika (Hgg.): Chancen und Perspektiven computerunterstützter Lexikographie. Tübingen, S. 87-104

1. Wörterbücher der Zukunft – alte und neue Visionen

„Alles Menschenwerk ist unvollkommen. Zu den unvollkommensten Erzeugnissen des Menschen gehören aber unstrittig die Wörterbücher“. So beginnt ein Aufsatz, der 1910 unter dem Titel „Wörterbücher der Zukunft“ in der germanisch-romanischen Monatszeitschrift erschienen ist (Tiktin (1910, 243)). Der Autor kritisiert darin sehr überzeugend die Unzulänglichkeiten der seinerzeit verfügbaren Wörterbücher und entwirft Konzepte zu deren Verbesserung: Bilder und Illustrationen sollen die Bedeutungserläuterungen anschaulicher und verständlicher machen. Sinngemäß Zusammengehöriges soll auch räumlich beieinander stehen, z.B. sollen bei Nahrungsmitteln alle regional verschiedenen Bezeichnungen aufgeführt werden, die Teile eines Ganzen sollen unter der Bezeichnung für das Ganze zu finden sein. Jede lexikalische Einheit soll bei all den Wortgruppen verzeichnet werden, zu denen sie ihrer Bedeutung nach gehört, sodass man sowohl vom Wort zum Begriff als auch vom Begriff zum Wort gelangen kann. Das Ansinnen, ein kleiner Kreis von Philologen könne für alle Wortschatzbereiche adäquate Bedeutungserläuterungen erstellen, hält Tiktin für ein „Ding der Unmöglichkeit“ und empfiehlt deshalb, „das segensreiche Prinzip der Arbeitsteilung“ auch für die Lexikographie nutzbar zu machen und fachbezogenen Wortschatz von Lexikographen beschreiben zu lassen, die sich im jeweiligen Fachgebiet auskennen (Tiktin (1910, 248)).

Das von Tiktin konzipierte Wörterbuch der Zukunft gibt es auch im Jahr 2000 noch nicht. Die inzwischen verfügbare Computertechnik bietet allerdings Anlass, bislang als unrealistisch verworfene Konzeptionen für Wörterbücher der Zukunft neu zu prüfen und erneut über die Zukunft der Wörterbücher nachzudenken. Insbesondere die als „Hypertext“ bezeichnete digitale Schreib- und Lesetechnologie macht es nun möglich, lexikographische Desiderate und Konzepte zu realisieren, die 1910 nach Tiktins eigener Einschätzung noch einer fernen Zukunft vorbehalten waren. In Hypertexten können Text-, Bild-, Ton- und Videodateien zur anschaulichen Vermittlung lexikalischen Wissens genutzt und die Wörterbuchartikel durch computerisierte Verweise, sog. „Links“, verknüpft werden. Lexikalische Daten können so modelliert werden, dass in Abhängigkeit von Nutzerinteressen und Nutzungssituationen die jeweils relevanten lexikographischen Angaben und Verweise herausgegriffen und in ästhetisch ansprechender Weise am Bildschirm dargestellt werden. Das Alphabet als Zugriffsstruktur im gedruckten Medium wird ersetzt durch Suchwerkzeuge, mit denen sich solche Datenbanken nach individuell wählbaren Kriterien durchsuchen lassen. Das World Wide Web (WWW), die Hypertext-Plattform im Internet, macht ein solches digitales Wörterbuch nicht nur weltweit zugänglich, sondern unterstützt auch die rasche und unkomplizierte Aktualisierung und Erweiterung der lexikographischen Daten. Die Verbindung von Information und Kommunikation im WWW erleichtert zudem die dezentrale Organisation des lexikographischen Arbeitsprozesses durch verteilt arbeitende Spezialisten.

Nicht zuletzt die Aufsätze in diesem Sammelband zeigen, dass Lexikographie und Metalexikographie beginnen, sich mit den Chancen des Hypertext-Konzepts auseinanderzusetzen.¹ Dennoch sind die derzeit verfügbaren elektronischen Wörterbücher noch stark von den Traditionen der Strukturierung und Gestaltung geprägt, die sich am gedruckten Medium herausgebildet haben. Das Potenzial des neuen Mediums wird erst ansatzweise genutzt. Dies gilt sowohl für die derzeit im WWW verfügbaren Wörterbücher² als auch für die auf CD-ROM vermarkteten Produkte³.

Der Grund dafür liegt darin, dass es sich bei vielen elektronischen Wörterbüchern um Parallelpublikationen zu gedruckten Wörterbüchern handelt und dass es immer noch das Print-Medium ist, an dem die Verlagslexikographie vornehmlich verdient (vgl. Klosa (in diesem Band)). Eine mediengerechte Hypertextualisierung eines Wörterbuchs ist aufwändig und muss sich rechnen; sie kann also unter kommerziellen Randbedingungen nur schrittweise erfolgen. Aufgabe der metalexikographischen Forschung muss es trotzdem sein, Zielvorstellungen für digitale Wörterbücher zu entwickeln, bei denen die Möglichkeiten und Chancen des neuen Mediums optimal genutzt werden. Solche Zielvorstellungen und deren prototypische Implementierung ermöglichen es kommerziellen und wissenschaftlichen Wörterbuchprojekten, die sukzessive Hypertextualisierung ihrer Bestände zu planen bzw. den Aufbau neuer Datenbestände mit den vorhandenen Ressourcen und Rahmenbedingungen möglichst optimal zu gestalten.

Mit den in Abschnitt 3 aufgestellten sieben Thesen möchte ich eine Diskussion über solche Zielvorstellungen anstoßen. Die Thesen greifen Aspekte auf, die in der Literatur zu elektronischen Wörterbüchern bereits diskutiert werden. Die Bündelung und thesenartige Zuspitzung dieser Aspekte erscheint mir aber wichtig zu sein, um künftige Diskussionen zu diesem Thema zu strukturieren, sodass sich die kontroversen Punkte auf Dauer von den weniger kontroversen unterscheiden lassen. Um die Thesen auch für Leser plausibel zu machen, die mit dem Hypertext-Konzept bislang nicht oder nur wenig vertraut sind, werde ich im folgenden Abschnitt kurz die lexikographisch relevanten Facetten des Konzepts erläutern und Hinweise auf weiterführende Literatur geben.

2. Was macht das Hypertextkonzept für die Lexikographie interessant?

Nachschlagewerke und speziell Wörterbücher werden gerne als Vergleichsobjekte herangezogen, wenn es darum geht, die für Hypertext typische nicht-lineare Organisationsform zu erläutern. Das Wörterverzeichnis von Wörterbüchern besteht aus einer Abfolge von Wörterbuchartikeln, die in beliebiger Abfolge und partiell rezipiert werden können. Entsprechend besteht auch ein Hypertext aus einer Menge von Modulen (auch „Knoten“ oder „informationelle Einheiten“ genannt), die in einer vom Benutzer selbst gewählten Abfolge am Bildschirm angezeigt werden. So wie die Wörterbuchartikel durch Verweise miteinander verknüpft sind, so sind auch die Module im Hypertext durch sog. Hyperlinks (im Weiteren „Links“ genannt) miteinander verbunden.

Diese Analogie von Wörterbuch und Hypertext hat sicher dazu beigetragen, dass schon sehr früh Wörterbücher in Hypertexte überführt wurden – das 1988 hypertextualisierte OED kann dabei bis heute als Leitprojekt gelten (vgl. Raymond/Tompa (1988)). Bei allen

¹ Vgl. Lemberg (in diesem Band), Müller/Schmidt (in diesem Band), Petelenz (in diesem Band), Richter (in diesem Band), Hass-Zumkehr (in diesem Band)).

² Vgl. Storrer/Freese (1996), Lemberg (in diesem Band).

³ Vgl. Feldweg (1997), Storrer (1998).

Gemeinsamkeiten dürfen jedoch die zentralen Unterschiede zwischen Print- und Hypertext-Wörterbüchern nicht übersehen werden:

- Module in Hypertext können im Gegensatz zu den statischen und stabilen Wörterbuchartikeln im Buch bei jeder Nachschlagehandlung neu zusammengesetzt werden, und zwar unter Berücksichtigung von situationsabhängigen Parametern wie der Muttersprache des Nutzers, seines Sprachlevels, dem Typ der Benutzungssituation.⁴
- Ebenso können Links nicht nur, wie die gedruckten Verweise, vom Autor fest verdrahtet sein, sondern können dynamisch nach Programmanweisung generiert werden und dabei in Abhängigkeit von situationsspezifischen Parametern ihre Eigenschaften verändern.
- Hypertextsysteme enthalten Such- und Recherchewerkzeuge, mit denen sich große Datenmengen nach selbst gewählten Kriterien im Nu auf eine Weise durchsuchen lassen, die im gedruckten Medium undenkbar oder zumindest sehr zeitaufwändig wäre.
- In Hypertexte können mediale Objekte wie Ton-, Bewegtbild- und Videodateien eingebunden werden, was das Spektrum der Beschreibungs- und Darstellungsmöglichkeiten für den Wörterbuchnutzer erheblich verbreitert.

Es sind im Wesentlichen drei Merkmale, die Hypertexte von anderen Medien, speziell vom gedruckten Buch, unterscheiden: Hypertexte sind computerverwaltete Texte, Hypertexte erlauben die Mehrfachkodierung von Daten in verschiedenen Symbolsystemen und deren Übermittlung auf mehreren Sinneskanälen, Hypertexte sind nicht-linear organisiert. Im Folgenden möchte ich diese drei Merkmale kurz erläutern und zeigen, was das Hypertext-Konzept für künftige lexikographische Produkte interessant macht.

2.1. Verwaltung durch ein Hypertextsystem

Sowohl zur Produktion als auch zur Rezeption von Hypertexten wird Software benötigt, die man als Hypertextsysteme bezeichnet. Erst Hypertextsysteme stellen die Werkzeuge bereit, um im Netzwerk der Module und Links zu navigieren, Module zu durchsuchen, zu verändern, ggf. vorhandene Tondateien abzuhören und Videosequenzen abzuspielen. Sie ermöglichen es, Module nach den individuellen Vorgaben des Nutzers zusammenzustellen und am Bildschirm anzuzeigen sowie Daten in Abhängigkeit von bestimmten Nutzungssituationen zu filtern. Die Bestimmung von Hypertext als Text, der sich nicht ohne Wertverlust auf Papier ausdrucken lässt, findet sich deshalb zurecht in verschiedensten Hypertext-Definitionen wieder⁵. Das Merkmal ist notwendig, um Hypertexte vom sog. „Textdesign“ im gedruckten Medium einerseits, vom computerverwalteten E-Text andererseits abzugrenzen:

- Als *Textdesign* bezeichnet man nach Blum/Bucher (1998) nicht-lineare Gestaltungsformen in gedruckten Zeitungen und Zeitschriften, die inzwischen auch zunehmend in andere Printpublikationen, v.a. Hand- und Lehrbüchern, Einzug halten. Textdesign verfolgt wie Hypertext das Ziel, die partielle und selektive Lektüre zu erleichtern, indem Informationen in kleinere Module zerlegt und auf den Seiten zu Clustern zusammengefügt werden. Dennoch bleiben Textdesign-Produkte stabile, physisch greifbare und begrenzbare Ganzheiten, in denen der Textdesigner über die Zusammenstellung der Module auf den Seiten entscheidet. Die zentrale Idee von Hypertext besteht dagegen genau darin, dass der

⁴ Vgl. z.B. Petelenz (in diesem Band), Richter (in diesem Band), Thielen/Breidt et al. (1998) und die These 4 in Abschnitt 3.

⁵ Z.B. Nelson (1972), Slatin (1991, 56).

Nutzer selbst bestimmt, welche Module er sich in welcher Reihenfolge und Anordnung auf dem Bildschirm anzeigen lassen möchte, wobei – wie oben bereits erwähnt – die Module und Links überhaupt nur „virtuell“ existieren können, d.h., erst bei Bedarf vom System aus kleineren Textbauteilen zusammengesetzt werden.

- Als *E-Texte (elektronische Texte)*⁶ bezeichnet man Texte, die in das World Wide Web eingebunden sind, ohne dass sie die für Hypertext charakteristische Organisationsform aufweisen. *E-Texte* sind häufig Parallel- oder Vorversionen von Print-Publikationen, die v.a. die schnelle und unkomplizierte Publikationsmöglichkeit des Internet nutzen; sie können ohne Wertverlust ausgedruckt und auf Papier gelesen werden.

Im Gegensatz zu E-Texten lassen sich Hypertexte nur mit Hilfe eines digitalen Lesegerätes rezipieren. Hierzu dienen bislang vornehmlich Computerbildschirme, bei denen – beim aktuellen Stand der Technik – erhebliche Abstriche an den Lesekomfort gemacht werden müssen. Allerdings zeigt die wachsende Akzeptanz von sog. „eBooks“ und der vorhersehbare Erfolg des internetfähigen Handys, dass man sich nicht voreilig der von Computerskeptikern oft vorgebrachten Auffassung anschließen sollte, Hypertext würde sich wegen des fehlenden Lesekomforts nicht durchsetzen. Erstens hat Bolter (1991, S. 4) aus medienhistorischer Perspektive zurecht darauf hingewiesen, dass Nutzungskomfort nicht das einzige Kriterium für die erfolgreiche Einführung neuer Medium darstellt, dass auch Bücher lange überwiegend auf Lesepulte gelegt und im Stehen rezipiert wurden, dass sich also auch das Medium „Buch“ nicht deshalb durchgesetzt hat, weil man es bequem auf dem Sofa lesen konnte. Außerdem ist absehbar, dass hinlänglich kostengünstige und gut transportable, digitale Lesegeräte mit papieräquivalenter Auflösung über kurz oder lang zu erschwinglichen Preisen auf dem Markt verfügbar sein werden. Es bedarf keiner besonderen prophetischen Gabe, um vorherzusagen, dass diese Entwicklungen den Bereich der Nachschlagewerke nachhaltig verändern werden.

Ein Nebeneffekt der Computerverwaltung sollte gerade für die lexikographische Anwendung nicht unterschätzt werden: Weil Druckraum teuer ist, konnte im Printwörterbuch immer nur ein sehr kleiner Teil potenziell verfügbarer Belege zu einem Lemma im zugehörigen Artikel berücksichtigt werden. Speicherplatz ist im Computer hingegen kein vergleichbar teurer Faktor; gerade bei Textdateien kann also ruhig eine Angabe mehr gemacht, ein Beispiel mehr gegeben werden (vgl. dazu auch These 3 in Abschnitt 3). Insgesamt können lexikographische Angaben verständlicher formuliert werden: Techniken zur Abkürzung und Textverdichtung, die in der Printlexikographie sinnvoll waren, um möglichst viel Daten auf möglichst knappem Raum unterzubringen⁷, sind im digitalen Medium überflüssig. Sie müssen sogar explizit rückgängig gemacht werden, um die lexikographischen Beschreibungen in „intelligente“ benutzeradaptive Wörterbuchsysteme zu integrieren (vgl. Breidt (1998), Thielen/Breidt et al. (1998)). Die Anforderung, kurz und prägnant abgefasste Wörterbuchartikel zu schreiben, verschiebt sich dadurch hin zu der Notwendigkeit, im Hypertextsystem Such- und Navigationswerkzeuge bereitzustellen, mit denen man aus einer großen Menge von Angaben diejenigen herausfiltern kann, die in einer aktuellen Benutzungssituation benötigt werden. Während ein kundiger Wörterbuchbenutzer bislang vor allem lernen musste, verdichtete Wörterbuchtexte korrekt zu entschlüsseln und souverän die vorhandenen inneren und äußeren Zugriffsstrukturen zu nutzen, muss der Nutzer digitaler Wörterbücher die Funktionalität seiner Such- und Filterwerkzeuge beherrschen lernen.

⁶ Die Bezeichnung E-Text habe ich übernommen von Dieter E. Zimmers Artikelserie „Die digitale Bibliothek“, vgl. <http://www4.zeit.de/nacht/digbib/index.html>.

⁷ Zur Textverdichtung im gedruckten Wörterbuch vgl. Wiegand (1987, 37ff), Wiegand (1996).

2.2. Mehrfachkodiertheit und Synästhetisierung

Hypertextsysteme können verschiedene mediale Objekte (Text-, Bild-, Audio- und Videodateien) verwalten, die in den Modulen kombiniert und durch Links verknüpft werden können.⁸ Es kann also auf dem visuellen und dem auditiven Kanal kommuniziert werden und dies unter Verwendung unterschiedlicher Symbolsysteme. Die im gedruckten Medium dominante Schrift lässt sich also nicht nur um Bilder und Grafiken anreichern, sondern auch um Ton- und Videodokumente. Hypertext-Gestaltung heißt, sich bewusst für ein- oder mehrkanalige Informationsvermittlung, für Schrift, Bild, Ton oder Video, zu entscheiden und aus den verschiedenen Elementen ein Ensemble zu flechten, das auf die Rezeption am Bildschirm und deshalb auf eine ganzheitliche Wahrnehmung als Bild hin ausgelegt ist (vgl. Schmitz (1997)). Diese Verflechtung von Schrift, Bild, Ton und Bewegung hat Freisler (1994, 31) als den Synästhetisierungsaspekt von Hypertext bezeichnet.

Der Synästhetisierungsaspekt erlaubt es, Wörterbücher nicht nur zweckmäßig, sondern auch ästhetisch ansprechend zu gestalten. Von lexikographischem Interesse sind aber vor allem die neuen Optionen zur mehrkanaligen Informationsvermittlung.⁹ Eine erste, vom Nutzwert aber bereits sehr wertvolle Anreicherung rein textbasierter Wörterbücher besteht darin, phonetische Umschriften um Audiodateien mit vertonten Ausspracheangaben zu ergänzen. Dies bietet sich vor allem für bilinguale und Lerner-Wörterbücher an und kommt dabei den Nutzern entgegen, die phonetische Umschriften nicht oder nur schwer lesen können. Vertonte Ausspracheangaben sind jedoch erst ein erster Schritt hin zu Wörterbüchern, die – ganz im Sinne der in Tiktin 1910 erhobenen Forderung nach besserer Anschaulichkeit – Text-, Bild-, Ton- und Videoelemente nutzen, um die Verständlichkeit und Anschaulichkeit von Bedeutungserläuterungen zu verbessern. Die MS-Bookshelf (1995) hat in ihren Wörterbucheintrag zu **Flamenco** beispielsweise ein Video integriert, das nicht nur die für den Tanz typischen Bewegungen, sondern auch den Rhythmus der dazugehörigen Musik vermittelt. Hypertext-Enzyklopädien wie MICROSOFTS ENCARTA vermitteln einen ersten Eindruck, wie Bilder und Videos, aber auch sukzessiv aufgebaute und durch gesprochene Sprache erläuterte Informationsgrafiken, sog. Animationen, zur Erläuterung komplizierter Sachverhalte und Abläufe eingesetzt werden können. Hier eröffnen sich v.a. für die Fachsprachen-Lexikographie interessante Perspektiven.

Ein Online-Wörterbuch, an dem sich der Wert der Mehrfachkodiertheit sehr eindrücklich zeigen lässt, ist das FACHGEBÄRDENLEXIKON PSYCHOLOGIE, das am Zentrum für Gebärdensprache der Universität Hamburg entwickelt wurde.¹⁰ In den Artikeln dieses Wörterbuchs sind schriftliche Bedeutungserläuterungen verknüpft mit Abbildungen von Handformen sowie mit Videos, auf denen die Gebärden vorgeführt werden. Jeder Artikel enthält nicht nur Links zu verwandten Fachbegriffen, sondern verknüpft auch Gebärden, die durch formähnliche Gesten ausgedrückt werden. Der Nutzer kann wahlweise alphabetisch oder über einen Sachindex einsteigen, er kann aber auch nach der Form einer bestimmten Gebärde suchen. Eine schön gemachte Umsetzung von Tiktins Anregung für das „Wörterbuch der Zukunft“, dass „bei jedem Tiernamen gesagt würde, wie vom menschlichen Ohre der Schrei des betreffenden Tieres aufgefaßt wird, durch welche Lautfolge die menschliche Sprache ihn wiederzugeben pflegt“ (Tiktin 1910, 253), ist das in Abbildung 1 gezeigte

⁸ Da inzwischen fast alle Hypertextsysteme verschiedene mediale Objekte verwalten, verzichte ich auf die früher übliche terminologische Unterscheidung zwischen rein textbasiertem „Hypertext“ und multimedialem „Hypermedia“.

⁹ Vgl. dazu Storrer (1998), Petelenz (in diesem Band); zum Einsatz in der wissenschaftlichen Lexikographie: Lemberg (in diesem Band), Richter (in diesem Band), Schröder (1997).

¹⁰ Vgl. <http://www.sign-lang.uni-hamburg.de/Projekte/PsychLex.html>.

multimediale Tiersprachenlexikon SOUNDS OF THE WORLD'S ANIMALS.¹¹ In ihm ist verzeichnet, wie Tierlaute in verschiedenen Sprachen der Welt verschriftet werden – wobei Proben der Originallaute als Audiodateien abgehört werden können.

2.3. Nicht-lineare Organisationsform

Die Grundidee der nicht-linearen Textorganisation lässt sich folgendermaßen skizzieren¹² : Der Autor eines Hypertextes verteilt seine Daten auf Module, die durch Links miteinander verknüpft sind. Metaphorisch gesprochen entsteht ein Wegenetz, mit den Links als Wegverbindungen zwischen den Modulen als den Orten, an denen Daten gespeichert sind. Die Verweisverfolgung geschieht durch das Aktivieren von Linkanzeigern, die als Schaltflächen, sensitive Wörter oder sensitive Graphiken gestaltet sein können. Ein Mausklick auf einen Linkanzeiger in einem Modul A führt dazu, dass ein damit verbundenes Modul B angezeigt wird.

Die nicht-lineare Organisationsform unterscheidet Hypertexte einerseits von zeitlich-linearen Audio- und Videodokumenten; andererseits von gedruckten Dokumenten, deren Struktur durch Teil-Ganzes- und Vorgänger-Nachfolger-Beziehungen zwischen den Dokumententeilen dominiert wird. Die Vorteile für lexikographischen Anwendungen liegen auf der Hand und wurden bereits erwähnt: Während ein gedrucktes Wörterbuch eine sequenzielle Anordnung der Textsegmente im Buch erzwingt¹³; erlauben es Hypertextsysteme, verschiedene gleichwertige Zugriffsmöglichkeiten auf die Daten anzubieten. Dies kommt der selektiven und problembezogenen Textrezeption entgegen, wie sie für Wörterbuchbenutzungssituationen typisch ist. Wörterbücher, ob gedruckt oder digital, werden meist ausschnittsweise rezipiert; welcher Ausschnitt selektiert wird, ist abhängig von der Benutzungssituation und dem dabei verfolgten Zweck. Von einem Wörterbuchartikel zu einem Lemma werden oft nur ganz spezielle Angaben benötigt, z.B. die Ausspracheangabe, die Bedeutungserläuterung oder eine grammatische Angabe. Aber auch die nicht-usuelle Verwendung eines Wörterbuchs als Lesebuch zielt nicht auf vollständige Textrezeption ab, wie bereits die Empfehlung Jacob Grimms zeigt: „Leser jedes Standes und Alters sollen auf den unabsehbaren Strecken der Sprache nach Bienenweise nur in die Kräuter und Blumen sich niederlassen, zu denen ihr Hang sie führt und die ihnen behagen“ (Grimm (1854, XII). Diesen selektiven und punktuellen Rezeptionsformen kommt die nicht-lineare Organisationsform von Hypertext optimal entgegen.

Ein weiterer Aspekt der nicht-linearen Organisationsform liegt darin, dass Textpassagen, die in der Printwelt in verschiedenen Büchern publiziert waren, nun über Links verknüpft werden können. Ein Beispiel hierfür sind digitale Textverbünde auf CD-ROM mit mehreren Nachschlagewerken wie Wörterbuch, Zitatsammlung, Chronologie und Multimedia-Atlas durch Links verknüpft sind. Die Links und die ergänzenden Suchfunktionen beschleunigen das gezielte Nachschlagen und fördern das themenorientierte Herumstöbern nach Informationen über die Grenzen einzelner Nachschlagewerke hinweg. Ein Beispiel aus der wissenschaftlichen Lexikographie ist der VERBUND MITTELHOCHDEUTSCHER WÖRTERBÜCHER, der Nachschlagewerke zum Mittelhochdeutschen verknüpft, die häufig im selben Arbeitskontext benötigt werden und auch explizit aufeinander bezogen sind¹⁴. Speziell zum amerikanischen Englisch gibt es im World Wide Web kostenlos verfügbare Angebote, in

¹¹ Cathy Ball: SOUNDS OF THE WORLD'S ANIMALS, vgl. <http://www.georgetown.edu/cball/animals/animals.html>.

¹² Zur Diskussion dieses Merkmals vgl. (Kuhlen (1991, 27ff), Freisler (1994, 21f), Storrer (2000).

¹³ Bei Wörterbüchern haben sich dabei verschiedene „Makrostrukturen“ genannte Anordnungsformen herausgebildet (vgl. Wiegand (1989), Wiegand (1998).

¹⁴ Vgl. <http://gaer27.uni-trier.de/MWV-online/MWV-online.html> und Burch/Fournier (in diesem Band).

denen verschiedene Nachschlagewerke miteinander verknüpft sind; sehr nützlich gerade für Nicht-Muttersprachler sind der WORDSMITH EDUCATIONAL DICTIONARY-THESAURUS¹⁵ und der MERRIAM-WEBSTER ONLINE¹⁶, die beide jeweils ein alphabetisches Wörterbuch und einen Thesaurus integrieren.

3. Sieben Thesen zur Nutzung des Hypertext-Konzepts in der Lexikographie

Die Chancen des Hypertext-Konzepts für die Lexikographie werden zwar zunehmend erkannt, die Diskussion darüber, wie die Mehrwerte des neuen Mediums optimal genutzt werden können, hat aber erst begonnen. Die nachfolgenden Thesen sollen dazu dienen, diese Diskussion zu strukturieren, die einzelnen, vom Medienwandel betroffenen Veränderungen herauszuarbeiten und die kontroversen Punkte von den weniger kontroversen unterscheiden zu helfen¹⁷. Die Thesen beziehen sich auf ein ideales digitales Wörterbuch der Zukunft, das im Hinblick auf den technischen Stand im Jahr 2000 neu konzipiert werden kann und dabei keinen speziellen ökonomischen und infrastrukturellen Beschränkungen unterliegt. Es ist mir bewusst, dass sich die meisten existierenden Wörterbuchprojekte nicht in einer solchen idealen Ausgangslage befinden, dass gerade die Verlagslexikographie Zwängen unterliegt, die die Verwirklichung solcher Zielvorstellungen nur schrittweise ermöglicht. Ich halte es dennoch für eine wichtige Aufgabe der Wörterbuchforschung, Ideen und Konzepte für digitale Wörterbücher der Zukunft zu entwickeln und die qualitative Verbesserung existierender Produkte durch Entwicklung von Forschungsprototypen und durch computergestützte Benutzungsforschung voranzutreiben.

Im vorigen Abschnitt wurde bereits gezeigt, dass die zunächst naheliegende Analogie von Wörterbuch und Hypertext einen zentralen Mehrwert des digitalen Mediums außer Acht lässt: Die Art und Weise, wie Daten in einem Datenbank- oder Hypertextsystem strukturiert sind, muss nicht der Art und Weise entsprechen, wie diese Daten dem Benutzer am digitalen Lesegerät (dem Bildschirm, dem Handy, dem eBook) präsentiert werden. Die Zielsetzung bei der Informationsmodellierung besteht vielmehr gerade darin, Daten so zu strukturieren, dass aus ein und demselben Datenpool für verschiedene Anwendungszwecke und Nutzungskontexte die jeweils relevanten Informationen herausgegriffen und in geeigneter Weise präsentiert werden können.¹⁸ Für die Informationsmodellierung gibt es bei digitalen Wörterbüchern zwei strategische Varianten:

1. Die Modellierung orientiert sich vornehmlich an den Bauteilen und Strukturen eines gedruckten Wörterbuchs. Die zentralen Bauteile sind dann lexikographische Angaben, die grundlegenden Strukturen sind Mikro-, Makro-, Artikel- und Verweisstrukturen, wie sie in

¹⁵ <http://www.wordsmyth.net/>.

¹⁶ <http://www.m-w.com/home.htm>.

¹⁷ Eine Vorfassung dieser Thesen habe ich mit den Teilnehmern des Workshops „SGML/XML-Einsatz in der Lexikographie“ diskutiert, der am 21.9.1999 an der Heidelberger Akademie der Wissenschaften stattgefunden hat. Die Anregungen aus dieser Diskussion haben zu einer Erweiterung von damals fünf auf die nun erörterten sieben Thesen geführt und mich überhaupt dazu ermutigt, einen derart programmatischen Text zu publizieren.

¹⁸ Vgl. hierzu auch Müller/Schmidt (in diesem Band). In der Datenbanktheorie spricht man in diesem Zusammenhang von Datenunabhängigkeit (vgl. Büchel/Schröder (in diesem Band)). Andere Zielsetzungen bei der Datenmodellierung, die Überprüfbarkeit von Integrität und Konsistenz, die Vermeidung von Updateanomalien und die Redundanzfreiheit, spielen im hier behandelten Zusammenhang eine untergeordnete Rolle.

der Metalexikographie beschrieben sind.¹⁹ Ein Beispiel für eine solche Modellierung ist die Document Type Definition, wie sie von der Text Encoding Initiative (TEI) für die Auszeichnung von Print-Wörterbüchern vorgeschlagen wurde.²⁰

2. Die Modellierung orientiert sich vornehmlich an den Einheiten und Strukturen des Wörterbuchgegenstands, in der Sprachlexikographie also v.a. an lexikalischen Einheiten und an den Relationen zwischen diesen. Die zentralen Bauteile sind dann linguistische Einheiten wie Grapheme, Morpheme und Lexeme; grundlegende Strukturen lassen sich beschreiben durch syntagmatische und paradigmatische Relationen sowie durch Type-Token-Beziehungen zwischen linguistischen Einheiten auf der Äußerungsebene und deren Korrelaten auf der Ebene des Lexikons. Ein Beispiel für eine solche Modellierung ist die lexikalische Datenbank WORDNET, die v.a. semantische Relationen zwischen Konzepten und Lexemen berücksichtigt (vgl. Fellbaum (1998)); in der deutschen „Schwester“ GERMANET sind aber auch syntagmatische Beziehungen und Wortbildungsregularitäten erfasst (vgl. Kunze/Wagner (in diesem Band); Hamp/Feldweg (1997)).

Die erste Variante liegt nahe, wenn ein gedrucktes Wörterbuch möglichst rasch ins digitale Medium überführt werden soll. Sie ist auch adäquat, wenn bei Projekten der sog. retrospektiven Digitalisierung vornehmlich angestrebt wird, eine digitale Kopie eines gedruckten Dokuments herzustellen, bei der die Formeigenschaften der gedruckten Vorlage gewahrt bleiben (vgl. Burch/Fornier (in diesem Band)). Wenn man jedoch die in Abschnitt 2 erläuterten Mehrwerte nutzen möchte, die Hypertext dem gedruckten Buch gegenüber aufweist, dann sollte man sich für die zweite Option entscheiden. Meine erste These lautet also:

These 1. Die konzeptuelle Datenmodellierung für digitale Wörterbücher sollte sich vornehmlich an linguistischen Einheiten und Strukturen ausrichten und nicht an den Strukturen eines Printwörterbuchs.²¹

Diese These möchte ich folgendermaßen begründen: Es wurde immer wieder beklagt, dass die alphabetische Anordnung der lexikalischen Einheiten im Wörterbuch viele wortschatzinterne Bezüge verdeckt, dass semantisch Zusammengehöriges räumlich auseinandergerissen wird.²² Diesen Nachteil musste man im gedruckten Buch durch Verweis- und Mikrostruktur sowie durch die Erschließung „versteckter“ lexikographischer Informationen in Registern kompensieren (vgl. Goebel/Lemberg/Reichmann (1995)).

Das digitale Medium bietet nun die Möglichkeit, die lexikographische Informationsmodellierung sehr eng an den Strukturen des Lexikons auszurichten, die ja in der modernen Linguistik relativ gut erforscht sind.²³ Dies erfüllt nicht nur die immer wieder gestellte Forderung, lexikologische Forschungsergebnisse stärker als bisher in der praktischen

¹⁹ Diese Strukturen sind sehr präzise erforscht und beschrieben, vgl. Wiegand (1989), Wiegand (1991), Bergenholtz/Tarp et al. (1999).

²⁰ Abrufbar unter <http://www.uic.edu/orgs/tei/p3/doc/p3di.txt>; vgl. dazu auch Büchel/Schröder (in diesem Band), Müller/Schmidt (in diesem Band) und Burch/Fournier (in diesem Band).

²¹ Eine sinngemäß formulierte These hat Andreas Blumenthal bereits 1987 in einem Workshop zu „Prinzipien des Entwurfs lexikalischer Datenbanken“ zur Diskussion gestellt; vgl. auch Blumenthal/Lemnitzer/Storrer (1988).

²² Vgl. Gloning/Welter (in diesem Band) und die Diskussion um integrierte Wörterbücher in dem sehr lesenswerten Sammelband „Nachdenken über Wörterbücher“ (Henne (1977, 47f), Wiegand (1977, 102f), Drosdowski (1977, 126f)). Dem Beitrag von Helmut Henne verdanke ich übrigens auch den Hinweis auf den Aufsatz von H. Tiktin.

²³ Im Sinne des in Lang (1983) beschriebenen Verhältnisses zwischen Lexikon und Wörterbuch.

Lexikographie zu berücksichtigen, sondern bringt drei weitere Vorteile mit sich. Erstens müssen bei entsprechender Modellierung viele Eigenschaften lexikalischer Einheiten, speziell in den Bereichen Flexion und Wortbildung, nicht mehr bei jedem Lemma einzeln aufgeführt werden, sondern könnten bei Bedarf über generelle Regeln abgeleitet werden. Zweitens lassen sich auf einem linguistisch motivierten Datenmodell flexiblere Sichten und Filter definieren, die sich auf den Informationsbedarf in bestimmten Nutzungskontexten einstellen (vgl. These 4). Drittens erfordert die Beschreibung linguistischer Merkmale und Strukturen für Datenbank- oder Hypertextsoftware ein höheres Maß an Formalisierung als die Erfassung entsprechender Merkmale im gedruckten Wörterbuch.²⁴ Bei umsichtiger Modellierung kann man deshalb von vornherein eine lexikalische Datenbank aufbauen, aus der Informationen für menschliche Benutzer einerseits, für Anwendungen der maschinellen Sprachanalyse andererseits herausgegriffen und in jeweils adäquater Form präsentiert werden, ganz im Sinne des in Breidt (1998) beschriebenen Wörterbuchs „für Mensch und Maschine“.

Auf der Basis einer linguistisch motivierten Informationsmodellierung kann auch das Nachschlagen aus einem Textverarbeitungsprogramm oder einem Internet-Browser heraus komfortabler und flexibler gestaltet werden. Ein einfaches Beispiel: Es ist für flektierende Sprachen wünschenswert, per Mausklick von einer flektierten Wortform im Text zu dem zugehörigen Eintrag im Wörterbuch zu gelangen. Bislang wird eine solche Funktion von keinem verfügbaren elektronischen Wörterbuch des Deutschen angeboten, obwohl die dafür benötigten sprachtechnologischen Werkzeuge vorhanden sind.²⁵ Die in Klosa (in diesem Band) geäußerte Replik auf meine diesbezügliche Anregung für die PC-BIBLIOTHEK (Storrer (1996)) zeigt, warum ein solches Desiderat so schwer erfüllbar ist: In der traditionellen Printlexikographie werden nicht alle Flexionsformen der Lemmata verzeichnet, sondern – um das Beispiel der Substantive zu nehmen – nur die Formen, die das Paradigma im Deutschen eindeutig identifizieren, d.h. Nominativ und Genitiv Singular sowie Nominativ Plural. Die Angaben sind, um Druckraum zu sparen, meist auch sehr kryptisch formuliert und können in recht unsystematischer Weise um Kommentare erweitert sein. Es ist deshalb nicht ohne Weiteres möglich, die prinzipiell vorhandenen Flexionsinformationen für die automatische Lemmatisierung zu nutzen, wie dies bei einer stärker formalisierten, linguistisch motivierten Datenmodellierung der Fall wäre.²⁶ Das einfache Beispiel dürfte bereits zeigen, dass eine linguistisch motivierte Datenmodellierung kein Desiderat aus dem lexikologischen oder metalexikographischen Elfenbeinturm ist. Sie schafft vielmehr die Voraussetzungen, um das Nachschlagen in digitalen Wörterbüchern, das ja immer häufiger im Zuge der Rezeption und Produktion von Texten am Bildschirm geschieht, flexibel und effizient an die Bedürfnisse in konkreten Nutzungssituationen anzupassen.

Die zweite These steht in engem Zusammenhang mit der ersten, bezieht sich aber nicht auf die Art und Weise der Datenmodellierung, sondern auf die sich daraus ergebenden Konsequenzen für den lexikographischen Arbeitsprozess.

These 2: Der lexikographische Arbeitsprozess sollte sich an den zu bearbeitenden lexikologischen Phänomenen und nicht am Alphabet orientieren.

²⁴ Dass die Beschreibungen in traditionellen Wörterbüchern unsystematisch und wenig konsistent sind, zeigten nicht zuletzt die Versuche, aus ihnen mit halbautomatischen Verfahren maschinenlesbare Lexika für die maschinelle Sprachverarbeitung zu gewinnen (vgl. dazu Boguraev/Briscoe (1989), Heyn (1992), Storrer/Feldweg et al. (1993)).

²⁵ Vgl. Feldweg (1997), Storrer (1998).

²⁶ Zu der Frage, ob die kryptischen grammatischen Angaben überhaupt von den menschlichen Benutzern problemlos dechiffriert werden können, gibt es m.W. noch keine empirische Studie; meine eigenen Erfahrungen mit Studierenden stimmen mich eher skeptisch.

Im Wörterverzeichnis gedruckter Wörterbücher muss die Abfolge der Wörterbuchartikel nach einem nachvollziehbaren Kriterium erfolgen, damit die Nutzer auf die Informationen rasch zugreifen können. Hierfür hat sich die alphabetische Anordnung bewährt, auch wenn sie seit langem als rein formal und nicht den lexikologischen Zusammenhängen entsprechend kritisiert wird. Die Kritik wird meist aus lexikologischer Warte formuliert; weitaus seltener werden die Nachteile thematisiert, die daraus entstehen, dass die alphabetische Anordnung im Buch in erheblichem Maße auch den lexikographischen Arbeitsprozess determiniert²⁷. Alphabetische Wörterbücher werden traditionell von A-Z nach sog. Buchstabenstrecken abgearbeitet; bei großen Wörterbuchprojekten sind daran mehrere Generationen von Lexikographinnen und Lexikographen beteiligt. Um das Wörterbuch einigermaßen konsistent zu halten, müssen relativ früh bereits der Wörterbuchplan, die Lemmaliste und das Artikelstrukturprogramm festgelegt werden. An diesen Vorgaben lässt sich dann nicht mehr viel verändern, auch wenn zwischen der Bearbeitung von Buchstabenstrecke A und der Bearbeitung von Buchstabenstrecke Z mehrere Jahrzehnte liegen. Dass bei diesem Vorgehen das Lemma „aufgehen“ und das Lemma „zugehen“ nicht mehr in der angemessenen Einheitlichkeit beschrieben sind, dass Verweise häufiger zu den bereits bearbeiteten Buchstabenstrecken führen als zu den nur geplanten, lässt sich mittlerweile mit elektronischen Auswertungsmethoden auch nachweisen.

Da es im digitalen Wörterbuch keine feste Anordnung der Lemmata, sondern viele Zugriffsmöglichkeiten auf die lexikalischen Informationen gibt, kann der lexikographische Arbeitsprozess nach lexikologisch motivierten Kriterien organisiert werden. Ein solches Vorgehen hat deutliche Vorteile gegenüber dem Abarbeiten von Buchstabenstrecken, weil das methodische Vorgehen bei der lexikographischen Arbeit in Abhängigkeit vom Typ des Lemmazeichens variiert – auf einen sehr einfachen Nenner gebracht: Bei der Beschreibung eines Verbs spielen andere Merkmale und Kategorien eine Rolle als bei der Beschreibung einer Gradpartikel. Wortschatzeinheiten könnten also wesentlich konsistenter bearbeitet werden, wenn der lexikographische Arbeitsprozess an Lemmazeichentypen ausgerichtet wird, wobei sich diese nach Kriterien der syntaktischen und/oder semantischen Zusammengehörigkeit beliebig fein untergliedern lassen. Das Abfassen der Wörterbuchartikel kann dann durch Computerwerkzeuge unterstützt werden, die Angaben zu jedem Lemma eines Lemmazeichentyps auf Konsistenz und auf Kompatibilität mit den Angaben der Lemmata desselben Typs überprüfen. Die Lexikographinnen und Lexikographen können sich jeweils auf die Besonderheiten des aktuell bearbeiteten Lemmazeichentyps konzentrieren und Forscher, die diesen Lemmazeichentyp untersuchen bzw. untersucht haben, an der Beschreibung der betreffenden Einträge mit beteiligen (vgl. These 7).

These 3. Die Verbindung zwischen lexikographischen Beschreibungen und lexikographischen Quellen muss für die Benutzer transparent und nachvollziehbar sein.

Das Quellen- und Belegprinzip gehört zu den wichtigsten Grundsätzen seriöser lexikographischer Arbeit. Die Lexikographinnen und Lexikographen dürfen sich nicht nur auf ihre Sprachkompetenz verlassen, sondern müssen ihre Beschreibungen durch authentisches Sprachmaterial belegen. Zu den wichtigen, aber auch zeit- und kostenintensiven Teilaufgaben lexikographischer Arbeitsprozesse gehört es deshalb, die Quelltexte auszuwählen, zu beschaffen und für die „eigentliche“ lexikographische Arbeit, das Abfassen der

²⁷ Die Darstellung ist stark vereinfachend; die Prozesse des lexikographischen Arbeitsprozesses sind detailliert und unter Berücksichtigung des Medienwandels in Wiegand (1999) beschrieben.

Wörterbuchartikel, aufzubereiten.²⁸ Vor dem Einsatz des Computers bei der Wörterbuchproduktion geschah dies im Allgemeinen durch das Exzerpieren von Belegen und deren anschließende Verzettelung und Archivierung in Zettelkästen (vgl. Lemberg (1996)). Inzwischen werden Belege und auch Corpora zunehmend computergestützt verwaltet²⁹. Im gedruckten Wörterbuch bekommt der Wörterbuchbenutzer selbst im günstigen Fall nur einen kleinen Bruchteil dieser Arbeit zu sehen, nämlich die Belegangaben, die von den Lexikographinnen und Lexikographen ausgewählt wurden. Bei ein- und mehrsprachigen Gebrauchswörterbüchern fällt die Angabe von authentischen Belegen oft ganz der Erfordernis zum Opfer, die entstehenden Produkte klein, handlich und im Preis erschwinglich zu halten; d.h., in den meisten Einbänden sind Belegangaben durch knappe, selbstkonstruierte Beispielangaben ersetzt.

Im digitalen Medium ist Speicherplatz kein relevanter Kostenfaktor. Deshalb können und sollten in digitalen Wörterbüchern die Wörterbuchartikel verknüpft sein mit den lexikographischen Quelltexten und/oder den daraus exzerpierten Belegen. Dies erlaubt es den Benutzern, den Weg vom Wörterbuch zu den Quellen zurückzugehen, die von den Lexikographen getroffenen Entscheidungen nachzuvollziehen und um eigenständige Recherchen zu ergänzen.³⁰ Der im lexikographischen Prozess ohnehin zu leistende Aufwand der Quellenbearbeitung kommt dann auch den Nutzern zugute – der Nutzwert der Wörterbücher wird bei gleichbleibenden Kosten erheblich gesteigert.

Der dafür zu betreibende Aufwand ist relativ gering, wenn die Quelltexte und Belegarchive bereits in digitalisierter Form vorliegen und vom Computerarbeitsplatz der beteiligten Lexikographen aus zugänglich sind. Ein computergestützter lexikographischer Arbeitsplatz muss ohnehin Werkzeuge bereitstellen, mit denen Corpus und Belegsammlung nach möglichst flexibel kombinierbaren Kriterien durchsucht und gefiltert werden können. Diese Werkzeugen lassen sich ohne erheblichen Mehraufwand zu Recherchertools für die künftigen Nutzer des Wörterbuchs weiterentwickeln. Bei einer solchen Weiterentwicklung sind zwei Dinge zu beachten: Im Gegensatz zu den Lexikographen, die ja täglich mit den Werkzeugen hantieren, werden die meisten Nutzer des digitalen Wörterbuchs nicht viel Zeit investieren wollen, um sich in die Suchfunktionen einzuarbeiten. Deshalb muss die Oberfläche einfacher gestaltet werden, d.h. zwischen leicht erlernbaren Grundfunktionen und spezielleren Zusatzfunktionen für die „power user“ unterscheiden. Wichtig ist es auch, dass der Benutzer bei der Recherche nicht von Belegen überschwemmt wird, sondern den Suchraum nach Bedarf einschränken kann.³¹ Voraussetzung hierfür ist die möglichst feinkörnige Annotation der Quelltexte und Belege mit Hilfe von Werkzeugen aus der Texttechnologie, d.h., dass die Quellen des lexikographischen Corpus mit bibliographischen Angaben versehen sind und die lexikalischen Einheiten in Beleg- und Quelltexten auf ihre Grundform zurückgeführt (lemmatisiert) und mit eindeutigen syntaktischen Kategorien versehen werden.³²

²⁸ Für eine detaillierte Beschreibung der hier sehr grob skizzierten Phasen lexikographischer Prozesse mit und ohne Computereinsatz verweise ich auf Wiegand (1999, Kap. 1.5.1. und 1.5.2.).

²⁹ Vgl. auch Schmidt (1997). Als Leitprojekt für einen erfolgreichen Umstieg vom Zettelkasten auf die Belegdatenbank kann das an der Heidelberger Akademie der Wissenschaften erarbeitete deutsche Rechtswörterbuch gelten (vgl. Speer (1994)).

³⁰ Die Vernetzung von Wörterbuch und Belegarchiv ist ein wichtiger Aspekt bei der Hypertextualisierung des DEUTSCHEN RECHTSWÖRTERBUCHS, vgl. dazu Lemberg/Petzold/Speer (1998) und Lemberg (in diesem Band).

³¹ Vgl. die „Maxime der optimalen quantitativen Korpusdokumentation“ in Richter (in diesem Band).

³² Das digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS), das momentan an der Berlin-Brandenburgischen Akademie der Wissenschaften vorbereitet wird (vgl.

These 4: Die Benutzerschnittstelle von digitalen Wörterbüchern sollte an Typen von Benutzungssituationen adaptierbar sein.

Am Beispiel von zweisprachigen Wörterbüchern wird die Motivation für diese These schnell deutlich: Ein Wörterbuch mit dem Sprachpaar deutsch-englisch wird benutzt von englischen und deutschen Muttersprachlern, zur Hin- und zur Herübersetzung, zum Verstehen und zum Produzieren fremdsprachlicher Texte, zur Suche nach Übersetzungsäquivalenten, nach grammatischen Eigenschaften fremdsprachlicher Einheiten oder nach typischen Kollokationen. Es ist metalexikographisch gut erforscht und beschrieben, dass in Abhängigkeit von solchen Parametern sehr unterschiedliche Informationen benötigt werden.³³ Der Wunsch nach vier bzw. sechs Wörterbüchern pro Sprachpaar ließ sich theoretisch gut begründen, galt jedoch im gedruckten Medium als praktisch nicht umsetzbar.

Anders im digitalen Wörterbuch: Es ist gerade die Stärke von Hypertextsoftware, aus ein und demselben Datenpool die Informationen herauszugreifen und in geeigneter Weise zu präsentieren, die für einen bestimmten Nutzungskontext typischerweise relevant sind. Dadurch wird es möglich, lexikographische Beschreibungen an den Informationsbedarf anzupassen, der für eine gegebene Nutzungssituation typisch ist. Statt der statischen Wörterbuchartikel im Printwörterbuch, in denen alle potenziell interessanten Informationen auf engstem Raum komprimiert sind, bekommt ein Nutzer eines digitalen Hypertext-Wörterbuchs Wörterbuchartikel „on demand“ mit den Angaben zusammengestellt, die für die aktuelle Wörterbuchbenutzungssituation relevant sind. Erste Prototypen derartiger kontextadaptiver Wörterbücher sind in Petelenz (in diesem Band) und in Thielen/Breidt/Feldweg (1998) beschrieben. Dass es sich bislang nur um Prototypen handelt, liegt weniger daran, dass nicht bekannt wäre, für welche usuellen Benutzungssituationen typischerweise welche Klassen von Angaben relevant werden.³⁴ Die Ursache liegt vielmehr darin, dass eine kontextadaptive Präsentation lexikalischer Informationen eine linguistisch motivierte und feinkörnige Modellierung der lexikographischen Daten voraussetzt (vgl. These 1). Eine derartige Modellierung erfordert, wenn sie auf der Grundlage eines gedruckten Wörterbuchs erfolgt, einen relativ hohen Auf- und Nachbereitungsaufwand (vgl. Breidt (1998), Feldweg (1997)) und lässt sich deshalb am schnellsten realisieren, wenn ein digitales Wörterbuch unabhängig von einer vorhandenen Printvorlage neu konzipiert werden kann.

These 5: Ein digitales Wörterbuch zur Gegenwartssprache sollte ein Ausbauwörterbuch sein.

In seinem sehr lesenswerten Diskussionspapier zum Computereinsatz in der Dialektlexikographie stellt Martin Schröder das konventionelle gedruckte Abschlusswörterbuch dem neuen elektronischen Ausbauwörterbuch gegenüber: „Während das zum Druck vorgesehene Wörterbuch die Informationsaufnahme notwendig mit der Bearbeitung einer bestimmten Wortstrecke abschließt bzw. hierfür bestenfalls einen Nachtrag bereithält, wird im datenbankorientierten Online-Wörterbuch die Bearbeitung der Wortstrecken grundsätzlich und an allen Stellen offen bleiben.“ (Schröder 1997, 16). Diese Offenheit bringt nicht nur für die Dialektlexikographie viele Vorteile mit sich, sondern ermöglicht es auch und vor allem der gegenwartsbezogenen Lexikographie, auf Veränderungen im Wortschatz zügig zu reagieren. Die Aktualisierung, die Revision und der Ausbau digital verwalteter lexikographischer Daten wird durch den modularen Aufbau und die computergestützte Verweisverwaltung enorm erleichtert. Dies bringt aber auch neue

http://www.bbaw.de/iag/dig_woerterbuch/index.html), wird auf einem derart aufbereiteten lexikographischen Corpus basieren und dieses auch für die künftigen Nutzer des Wörterbuchs recherchierbar machen.

³³ Vgl. Kromann/Riiber/Rosbach (1991) und Breidt (1998) für einen Überblick.

³⁴ Vgl. Kühn (1989) und in Wiegand (1999, Kap. 4.2.2.).

Probleme mit sich: Es muss auch in einem Ausbauwörterbuch gesichert sein, dass die abrufbaren Informationen verbindlich, verlässlich und zitierbar sind. Ein digitales Wörterbuch, das seine Wörterbuchartikel in vom Benutzer nicht nachvollziehbarer Weise verändert oder löscht, verliert für die Zwecke an Wert, bei denen es auf Verlässlichkeit und Zitierbarkeit ankommt. Werkzeuge und Verfahren der Versionsverwaltung können hier Abhilfe schaffen, wenn die Probleme frühzeitig erkannt und berücksichtigt werden.

These 6: Digitale Wörterbücher sollten die Option der Mehrfachkodierung sinnvoll einsetzen.

Die Argumente für diese These wurden eigentlich bereits in Abschnitt 2.2 genannt und sind auch weitgehend unumstritten: Die Möglichkeit, das bislang textdominierte Wörterbuch durch Bild-, Ton- und Videoobjekte zu ergänzen, kann die Verständlichkeit und Anschaulichkeit lexikographischer Angaben verbessern. Die Chancen für die Dialektlexikographie, die Erklärung fachsprachlicher Termini in ihrem fachsystematischen Kontext und vor allem auch für die muttersprachliche und fremdsprachliche Sprachdidaktik liegen auf der Hand. Die neuen Gestaltungsmittel bringen aber auch neue Anforderungen an die Wörterbuchmacher mit sich: Empirische Untersuchungen deuten darauf hin, dass eine schlecht koordinierte Verknüpfung von Text, Bild, Ton und Video die Informationsaufnahme verschlechtert statt sie zu verbessern (Weidenmann (1995)). Mehrfachkodierung zu nutzen, bedeutet deshalb nicht einfach, möglichst viele Text-, Ton-, Bild- und Videoobjekte zu verknüpfen. Wichtig für eine qualitativ hochwertige Anwendung ist vielmehr die Integration der unterschiedlichen Zeichentypen nach semantisch-funktionalen Prinzipien. Hier besteht im Bereich der Lexikographie noch ein erheblicher Forschungsbedarf.³⁵

These 7: Online-Wörterbücher sollten die Chancen der Verbindung von Information und Kommunikation gezielt für die Qualitätssicherung nutzen.

Als Online-Wörterbücher bezeichne ich digitale Wörterbücher, die über das World Wide Web zugänglich sind (vgl. auch Lemberg (in diesem Band)). Die Stärke des WWW liegt in der Verbindung von Information und Kommunikation: Mit WWW-Browsern kann man nicht nur Informationen abrufen, sondern auch die Kommunikationsdienste des Internet in Anspruch nehmen, von der elektronischen Post (E-mail) und den Postverteilern (Mailing-Listen) über die Diskussionsgruppen (Newsgroups), bis hin zu den Online-Konferenzen (Chat). Dies eröffnet Wörterbuchprojekten neuartige Möglichkeiten, in Kontakt mit ihren Nutzern zu treten und diese am Aufbau und an der Pflege des Wörterbuchs zu beteiligen. Die Formen der Partizipation in existierenden Online-Wörterbüchern reichen von der Bitte um Fehlerkorrektur und Rückmeldung, über den Aufruf zur Beteiligung an der Schließung von Lemmalücken, bis hin zu Wörterbüchern, die ganz oder überwiegend von den Beiträgen ihrer „Gäste“ leben (vgl. Storrer/Freese (1996) und Storrer (1998, Abschnitt 4.2)).

Ein weiterer Vorteil: Wenn das lexikographische Corpus bzw. die Belegsammlungen eines Projekts in digitalisierter Form über das WWW zugänglich sind, wird die räumliche Nähe der Lexikographen zu einem Belegarchiv unerheblich. Dies erleichtert den arbeitsteiligen Aufbau von Wörterbüchern in räumlich verteilten Arbeitsstellen erheblich. In These 2 wurde bereits erläutert, warum die Arbeitsteilung nicht nach Buchstabenstrecken sondern nach lexikologischen Kriterien erfolgen sollte. Im Idealfall kann bei einem verteilten Wörterbuchprojekt jeder Lemmazeichentyp genau von den Forschern bearbeitet werden, die sich mit dem betreffenden lexikologischen Phänomen bereits beschäftigt haben. Die eingangs

³⁵ Erste Überlegungen finden sich in Petelenz (in diesem Band) und Hupka (1989).

zitierte Anregung von H. Tiktin, das Prinzip der Arbeitsteilung auf den lexikographischen Arbeitsprozess zu übertragen, kann also im WWW erstmals effizient und in großem Stil umgesetzt werden. Tiktin hatte dabei vor allem das Problem vor Augen, dass fachsprachlicher Wortschatz im Grunde nur von Lexikographen beschrieben werden kann, die sich im jeweiligen Fachgebiet auch auskennen, dass aber die meist philologisch ausgebildeten Lexikographinnen und Lexikographen häufig in Naturwissenschaft, Handwerk und Technik wenig bewandert sind (Tiktin (1910, 248f)). Das WWW macht es nun möglich, beim Aufbau von allgemeinsprachlichen Wörterbüchern Spezialisten der jeweiligen Fachdisziplinen hinzuzuziehen und somit die Verlässlichkeit und Korrektheit der Erläuterungen zu verbessern. Dieser Aspekt ist wichtig, weil die verständliche Erläuterung von Fachvokabular maßgeblich dazu beiträgt, dass sich die Bürger in der sog. „Informationsgesellschaft“ über aktuelle Entwicklungen im wirtschaftlichen, technologischen und wissenschaftlichen Bereich ausreichend informieren können, um die politische Auseinandersetzung über die Folgen dieser Entwicklungen nachvollziehen und sich ein eigenes Urteil bilden zu können.

Dass sich der kollaborative Wörterbuchaufbau über das World Wide Web tatsächlich organisieren lässt, zeigen Wörterbuchprojekte, die – bislang weitgehend unbeachtet von der metalexikographischen Forschung und ganz im Geiste der Open-Source-Projekte – Spezialwörterbücher, aber auch allgemeinsprachliche bilinguale Wörterbücher aufbauen und kostenlos zur Verfügung stellen. Ein bemerkenswertes Beispiel für ein Wörterbuch dieser Art zum Sprachpaar Deutsch-Englisch ist LEO (<http://dict.leo.org/>). 1997 gestartet, erwies es sich bei der in Storrer/Freese (1996) durchgeführten Stichprobe als wenig zuverlässig. Inzwischen haben viele freiwillige Helfer das Wörterbuch deutlich erweitert und verbessert. Die Stärke von LEO liegt in einem sehr reichhaltigen Kollokations- und Phraseologieteil, der viele aktuelle Termini und Wendungen und fachsprachliches Vokabular aus Informationstechnik und Wirtschaft enthält; in diesen Bereichen schlägt es konventionelle bilinguale Printwörterbücher an Abdeckung und Aktualität. Es ist deshalb v.a. für Nutzer wertvoll, die zwar nicht professionell übersetzen, im beruflichen Alltag aber häufig englische Gebrauchstexte verfassen oder verstehen müssen – die Zahl von ca. 200.000 Zugriffen pro Tag (März 2000) zeigt die steigende Beliebtheit der Ressource.

Ein anderer Aspekt der Qualitätssicherung soll zum Abschluss noch angesprochen werden: Die Nutzung von Online-Wörterbüchern über das Internet wird protokolliert, d.h., es ist erstmals möglich, auf einfache Art und Weise herauszufinden, wieviele Nutzer welche Daten abgerufen haben. Auch wenn die in Lemnitzer (in diesem Band) beschriebene Studie eher ernüchternde Ergebnisse lieferte, sollte man das Potenzial nicht unterschätzen, das in der automatischen Protokollierung von Benutzeraktionen für die künftige Wörterbuchbenutzungsforschung liegt. Je erfolgreicher und qualitativ hochwertiger ein Wörterbuch, umso häufiger dürfte es usuell genutzt und nicht nur mit den Lemmata getestet werden, die Nutzern nach der Lemnitzer-Studie offensichtlich spontan zuerst einfallen. Eine entsprechende Studie mit dem Wörterbuch LEO dürfte bereits wesentlich interessantere Ergebnisse zeigen.

Insgesamt bietet die Verbindung von Information und Kommunikation im World Wide Web in bislang unbekannter Weise die Chance, Wörterbücher in Auseinandersetzung mit und unter Beteiligung von den Sprachbenutzern zu erarbeiten und die Qualität der lexikographischen Produkte bereits zu einem frühen Zeitpunkt mit den Nutzern zu testen. Wörterbücher dieser Art herzustellen und in Gang zu halten – schließlich geht es ja gemäß These 5 um Ausbauwörterbücher – wäre „eine Aufgabe, des Schweißes der Edlen wert“, wie es Tiktin (1910, S. 253)) für seine Vision von den Wörterbüchern der Zukunft formuliert hat.

4. Literatur

- BERGENHOLTZ, HENNING; TARP, SVEN; WIEGAND, HERBERT ERNST (1999): Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In: HOFFMAN, L.; KALVERKÄMPER, H.; WIEGAND, H. E. (Hgg.) (1999): Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung. Berlin/New York, 1762-1832.
- BLUM, JOACHIM; BUCHER, HANS-JÜRGEN (1998): Die Zeitung: Ein Multimedia. Textdesign - ein Gestaltungskonzept für Text, Bild und Grafik. Konstanz.
- BLUMENTHAL, ANDREAS; LEMNITZER, LOTHAR; STORRER, ANGELIKA (1988): Was ist eigentlich ein Verweis? - Konzeptuelle Datenmodellierung als Voraussetzung computergestützter Verweisbehandlung. In: HARRAS, G. (Hg.) (1988): Das Wörterbuch: Artikel und Verweisstrukturen. Düsseldorf, 351-373.
- BOGURAEV, BRAN; BRISCOE, TED (1989): Utilising the LDOCE grammar codes. In: BOGURAEV, B.; BRISCOE, T. (Hgg.) (1989): Computational Lexicography for Natural Language Processing. London/New York, 85-116.
- BOLTER, JAY DAVID (1991): Writing Space. The Computer, Hypertext and the History of Writing. Hillsdale NY.
- BREIDT, ELISABETH (1998): Neuartige Wörterbücher für Mensch und Maschine: Wörterbuchdatenbanken in COMPASS. In: WIEGAND, H. E. (Hg.) (1998): Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. Tübingen, 1-27.
- DROSDOWSKI, GÜNTHER (1977): Nachdenken über Wörterbücher: Theorie und Praxis. In: DROSDOWSKI, G.; HENNE, H.; WIEGAND, H. E. (Hgg.) (1977): Nachdenken über Wörterbücher. Mannheim, 103-143.
- FELDWEG, HELMUT (1997): Wörterbücher und neue Medien: Alter Wein in neuen Schläuchen? In: LiLi 27/107 (1997), 110-123.
- FELLBAUM, CHRISTIANE (Hg.) (1998): WORDNET: An electronic lexical database. Cambridge, MA.
- FREISLER, STEFAN (1994): Hypertext - eine Begriffsbestimmung. In: Deutsche Sprache 1 (1994), 19-50.
- GOEBEL, ULRICH; LEMBERG, INGRID; REICHMANN, OSKAR (1995): Versteckte lexikographische Information. Möglichkeiten ihrer Erschließung dargestellt am Beispiel des Frühneuhochdeutschen Wörterbuchs. Tübingen.
- GRIMM, JACOB (1854): Vorrede zum Deutschen Wörterbuch. In: GRIMM, J.; GRIMM, W. (Hgg.) (1854): Deutsches Wörterbuch. Leipzig, I-LXVIII.
- HAMP, BIRGIT; FELDWEG, HELMUT (1997): GermaNet -- A lexical semantic net for German. In: VOSSEN, P.; CALZOLARI, N. (Hgg.) (1997): Proceedings of the ACL/EACL-97 Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. Madrid, 9-15.
- HENNE, HELMUT (1977): Nachdenken über Wörterbücher: Historische Erfahrungen. In: DROSDOWSKI, G.; HENNE, H.; WIEGAND, H. E. (Hgg.) (1977): Nachdenken über Wörterbücher. Mannheim, 7-50.
- HEYN, MATTHIAS (1992): Zur Wiederverwendung maschinenlesbarer Wörterbücher - eine computergestützte metalexikographische Studie am Beispiel der elektronischen Edition des "Oxford Advanced Learner's Dictionary of Current English". Tübingen.
- HUPKA, WERNER (1989): Wort und Bild. Die Illustrationen in Wörterbüchern und Enzyklopädien. Tübingen.
- KROMANN, H.-P.; RIIBER, T.; ROSBACH, P. (1991): Principles of bilingual lexicography. In: HAUSMANN, F. J.; REICHMANN, O.; WIEGAND, H. E. ET AL. (Hgg.) (1991): Wörterbücher. Ein internationales Handbuch zur Lexikographie, 3. Teilband. Berlin/New York, 2711-2728.
- KUHLEN, RAINER (1991): Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank. Berlin et al.
- KÜHN, PETER (1989): Typologie der Wörterbücher nach Benutzungsmöglichkeiten. In: HAUSMANN, F. J.; REICHMANN, O.; WIEGAND, H. E. ET AL. (Hgg.) (1989): Wörterbücher. Ein internationales Handbuch zur Lexikographie, 1. Teilband. Berlin/New York, 111-127.
- LANG, EWALD (1983): Lexikon als Modellkomponente und Wörterbuch als lexikographisches Produkt. In: SCHILDT, J.; VIEHWEGER, D. (Hgg.) (1983): Die Lexikographie von heute und das Wörterbuch von morgen. Analysen - Probleme - Vorschläge. Berlin, 76-91.

- LEMBERG, INGRID (1996): Die Belegexzerption zu historischen Wörterbüchern am Beispiel des Frühneuhochdeutschen Wörterbuchs und des Deutschen Rechtswörterbuchs. In: WIEGAND, H. E. (Hg.) (1996): Wörterbücher in der Diskussion II. Tübingen, 83-192.
- LEMBERG, INGRID; PETZOLD, SYBILLE; SPEER, HEINO (1998): Der Weg des Deutschen Rechtswörterbuchs in das Internet. In: WIEGAND, H. E. (Hg.) (1998): Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. Tübingen, 265-289.
- NELSON, THEODOR H. (1972): As We Will Think. REPRINT IN: NYCE, J. M. K., PAUL (1991) (Hg.): From Memex to Hypertext: Vannevar Bush and the Mind's Machine: 245-259.
- RAYMOND, DARRELL R.; TOMPA, FRANK W.M. (1988): Hypertext and the Oxford English Dictionary. In: Communications of the ACM 31/7 (1988), 871-879.
- SCHMIDT, HARTMUT (1997): Plädoyer für eine moderne korpusbasierte deutsche Wortschatzforschung. In: Zeitschrift für Linguistik und Literaturwissenschaft 27/196 (1997), 19-29.
- SCHMITZ, ULRICH (1997): Schriftliche Texte in multimedialen Kontexten. In: WEINGARTEN, R. (Hg.) (1997): Sprachwandel durch den Computer? Opladen, 131-157.
- SCHRÖDER, MARTIN (1997): Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projekts "Deutsches Dialektwörterbuch". In: Zeitschrift für Dialektologie und Linguistik LXIV/1 (1997), 58-66.
- SLATIN, JOHN M. (1991): Composing Hypertext: A Discussion for Writing Teachers. In: BERK, E.; DEVLIN, J. (Hgg.) (1991): Hypertext / Hypermedia Handbook. New York et al., 55-64.
- SPEER, HEINO (1994): DRW to FAUST. Ein Wörterbuch zwischen Tradition und Fortschritt. In: Lexicographica 10 (1994). Tübingen, 171-213.
- STORRER, ANGELIKA (1996): Wörterbücher zum Anklicken. Ein kleiner Rundgang durch die PC-Bibliothek. In: Sprachreport 2/95 (1996), 9-10.
- (1998): Hypermedia-Wörterbücher: Perspektiven für eine neue Generation elektronischer Wörterbücher. In: WIEGAND, H. E. (Hg.) (1998): Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. Tübingen, 107-135.
- (2000): Was ist "hyper" am Hypertext? In: KALLMEYER, W. (Hg.) (2000): Sprache und neue Medien. Berlin u.a..
- STORRER, ANGELIKA; FELDWEG, HELMUT; HINRICHS, ERHARD (1993): Korpusunterstützte Entwicklung lexikalischer Wissensbasen. In: Sprache und Datenverarbeitung 17 (1993), 59-72.
- STORRER, ANGELIKA; FREESE, KATRIN (1996): Wörterbücher im Internet. In: Deutsche Sprache 24/2 (1996), 97-136.
- THIELEN, CHRISTINE; BREIDT, ELISABETH; FELDWEG, HELMUT (1998): COMPASS. Ein intelligentes Wörterbuchsystem für das Lesen fremdsprachiger Texte. In: STORRER, A.; HARRIEHAUSEN, B. (Hgg.) (1998): Hypermedia für Lexikon und Grammatik. Tübingen, 173-194.
- TIKTIN, H. (1910): Wörterbücher der Zukunft. In: Germanisch-Romanische Monatszeitschrift II. Jahrgang (1910), 243-253.
- WEIDENMANN, BERND (1995): Multicodierung und Multimodalität im Lernprozeß. In: ISSING, L. J.; KLIMSA, P. (Hgg.) (1995): Information und Lernen mit Multimedia. Weinheim, 65-84.
- WIEGAND, HERBERT ERNST (1977): Nachdenken über Wörterbücher: Aktuelle Probleme. In: DROSDOWSKI, G.; HENNE, H.; WIEGAND, H. E. (Hgg.) (1977): Nachdenken über Wörterbücher. Mannheim, 51-102.
- (1987): Wörterbuchartikel als Text. In: HARRAS, G. (Hg.) (1987): Das Wörterbuch - Artikel und Verweisungsstrukturen. Jahrbuch 1987 des Instituts für deutsche Sprache. 30-120.
- (1989): Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. In: HAUSMANN, F. J.; REICHMANN, O.; WIEGAND, H. E. ET AL. (Hgg.) (1989): Wörterbücher. Ein internationales Handbuch zur Lexikographie, 1. Teilband. Berlin/New York, 462-501.

- (1989): Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. In: HAUSMANN, F. J.; REICHMANN, O.; WIEGAND, H. E. ET AL. (Hgg.) (1989): Wörterbücher. Ein internationales Handbuch zur Lexikographie, 1. Teilband. Berlin/New York, 371-409.
- (1991): Printed Dictionaries and their Parts as Text. An Overview of more Research as an Introduction. In: Lexikographica 6 (1990) (1991), 1-124.
- (1996): Textual Condensation in Printed Dictionaries. A Theoretical Draft. In: Lexikos 6 (1996), 133-158.
- (1998): Altes und Neues zur Makrostruktur alphabetischer Printwörterbücher. In: WIEGAND, H. E. (Hg.) (1998): Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. Tübingen, 348-372.
- (1999): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband. Berlin/New York.