

Einführung

Die moderne Wörterbuchforschung beschreibt Lexikographie, speziell die wissenschaftliche Sprachlexikographie, als eigenständige, kulturelle und wissenschaftliche Praxis. Diese Praxis orientierte sich über Jahrhunderte hinweg nahezu ausschließlich an dem Medium Buch. Bücher waren meist die Träger der lexikographischen Quellentexte, aus denen Belege exzerpiert und in Belegarchiven gesammelt wurden. In Büchern wurden die Endprodukte der lexikographischen Arbeit, die Wörterbuchartikel, publiziert und vermarktet. Lange Zeit war das Medium Buch auch das geeignetste Trägermedium für Nachschlagewerke, da es den gezielten, punktuellen Zugriff auf bestimmte Teiltex-te besser unterstützt als beispielsweise ein Film, ein Zettelkasten oder eine Papyrusrolle.

Dies hat sich durch die Computertechnik, vor allem durch Datenbank- und Hypertextsysteme sowie durch Computernetzwerke wie das Internet, geändert. In digitalen Wörterbüchern nimmt der Computer den Nutzerinnen und Nutzern das zeitraubende Blättern und Suchen ab, ein Verweis kann durch einen einfachen Mausklick verfolgt werden, das Durchsuchen von mehreren Nachschlagewerken und die Volltextsuche in den gesamten Artikeltexten ist kein Problem. Von diesen Vorteilen können die Nutzerinnen und Nutzer teilweise bereits dann profitieren, wenn elektronische Wörterbücher nur als digitale Kopien der gedruckten Vorläufer publiziert werden. Doch dies ist erst der Beginn eines Prozesses, bei dem gedruckte Wörterbücher sukzessive durch leistungsfähige digitale Nachschlagewerke ergänzt und für viele Benutzungssituationen oder Benutzungsanliegen ersetzt werden. Wer diesen Prozess mitgestalten und dabei die Chancen und Perspektiven des digitalen Mediums in der Lexikographie ausreizen möchte, muss kritisch hinterfragen, inwieweit die am Buch orientierten Traditionen der Wörterbuchgestaltung, z.B. die Makro-, Mikro-, Zugriffs- und Verweisstrukturen, im digitalen Medium noch zweckmäßig sind und inwieweit das neue Medium nicht nach anderen Gestaltungs- und Organisationsprinzipien verlangt. Speicherplatz ist im digitalen Wörterbuch nicht annähernd so teuer wie Druckraum im gedruckten Wörterbuch. Viele Techniken zur Textverdichtung, insbesondere die Textkomprimierung, die in der Printlexikographie notwendig und zum Teil auch sinnvoll waren, um auf möglichst wenig Platz möglichst viel lexikographische Information unterzubringen, werden im digitalen Medium überflüssig. Statt dessen werden Techniken der strukturierten und redundanzfreien Informationsmodellierung benötigt, wie sie z.B. in der Datenbanktheorie verwendet werden, sowie Strategien, um aus dem digitalen lexikalischen Datenpool diejenigen Angaben herauszufiltern und am Bildschirm zu präsentieren, die in einer bestimmten Benutzungssituation benötigt werden. Dadurch wird es möglich, lexikographische Angaben verständlicher zu formulieren und den Aufbau der Wörterbuchartikel flexibel an den aktuellen Informationsbedarf der Nutzerinnen und Nutzer anzupassen.

Das digitale Medium beendet auch die Diskussion um die „richtige“ Anordnung des Wortschatzes, die sich wie ein roter Faden durch die Geschichte der Lexikographie zieht. Die alphabetische Anordnung, zurecht als rein formal und nicht den lexikologischen Zusammenhängen entsprechend kritisiert, war bislang effizient in ihrer Funktion, den gezielten Zugriff auf die lexikographischen Informationen zu erleichtern. Diese Funktion entfällt in digitalen Wörterbüchern, in denen die Suche nach Informationen auf Techniken des *Information Retrieval* basiert. Statt dessen wird nun zu diskutieren sein, wie eine adäquate

linguistisch motivierte Modellierung lexikographischer Angaben im digitalen Medium aussehen soll.

Die Computertechnik verändert auch den lexikographischen Arbeitsprozess. Der Computereinsatz in der Wörterbuchwerkstatt erlaubt es bei entsprechender technischer Infrastruktur, Abläufe effizienter und flexibler zu gestalten und damit gerade umfangreiche Wörterbuchprojekte schneller, qualitätsvoller und kostengünstiger abzuschließen. Wenn das lexikographische Korpus und die Belegsammlungen in Datenbanken verwaltet werden, können Lexikographinnen und Lexikographen von ihren Arbeitsplatzrechnern aus darauf zugreifen. Dies macht den langwierigen und kostspieligen Aufbau von Belegarchiven und die Anmietung von Räumen für deren Aufbewahrung künftig überflüssig. Liegt einmal das lexikographische Korpus zusätzlich mit korpuslinguistischen Methoden aufbereitet vor und ist der lexikographische Arbeitsplatz mit leistungsfähigen Suchwerkzeugen ausgestattet, dann verfügen die Lexikographinnen und Lexikographen über Recherche- und Auswertungsmöglichkeiten, von denen ihre Vorgängergenerationen nur hätten träumen können. Die Verwaltung lexikographischer Beschreibungsergebnisse durch ein Datenbanksystem hilft zudem bei der Konsistenzprüfung und der Verweiskontrolle. Die digitale Verfügbarkeit von Wörterbuchbasis und Beschreibungsergebnissen macht es möglich, Arbeitsabläufe zu dezentralisieren, d.h., auf räumlich verteilte Arbeitsstellen zu verteilen. Die Kommunikationsdienste des Internet können dabei nicht nur projektintern, sondern auch zur Außendarstellung und Qualitätskontrolle genutzt werden, indem z.B. Probeartikel getestet und Rückmeldungen künftiger Nutzerinnen und Nutzer bei der Wörterbuchplanung berücksichtigt werden.

Die hier nur grob skizzierten Vorteile des digitalen Mediums sollten laufenden und geplanten Wörterbuchprojekten genug Anlass geben, die Technikhürde zu überwinden, und die Chancen und Perspektiven der Computertechnik für die Lexikographie zu nutzen. Dieser Band enthält Beiträge von Autorinnen und Autoren aus der lexikographischen Praxis und der historischen und computerlinguistischen Sprachwissenschaft, die sich unter verschiedenen Aspekten mit diesen Herausforderungen auseinander setzen. Drei grundlegende Konzepte der computergestützten Lexikographie stehen dabei im Mittelpunkt: die Einsatzmöglichkeiten der Auszeichnungssprachen SGML/XML für die lexikographische Informationsmodellierung, die Nutzung des Internet als Publikationsmedium für lexikographische Produkte und Hypertext/Hypermedia als Organisationsform für lexikalisches Wissen. Die Beiträge im ersten Teil des Bandes führen in diese Konzepte ein und erörtern deren Chancen und Perspektiven für die computerunterstützte Produktion und Publikation von Wörterbüchern. Im zweiten Teil des Bandes werden theoretische und methodische Aspekte diskutiert, die sich beim Einsatz dieser Konzepte in konkreten Wörterbuchprojekten ergeben.

Im ersten Beitrag erläutern *Gregor Büchel* und *Bernhard Schröder* grundlegende informationstechnische Konzepte und Techniken zur Strukturierung und Verwaltung lexikographischer Daten: Datenbanksysteme, Datenmodelle und Verfahren der konzeptionellen Datenmodellierung sowie die Grundlagen der textuellen Informationsmodellierung mit SGML und XML. Die Aspekte der konzeptionellen Datenmodellierung und des Einsatzes von SGML/XML in lexikographischen Produktions- und Publikationsprozessen werden im nachfolgenden Beitrag von *Ingrid Schmidt* und *Carolin Müller* vertieft. Auf der Basis einer kritischen Auseinandersetzung mit den von der Text Encoding Initiative (TEI) vorgeschlagenen Dokumentstrukturgrammatik (Document Type Definition, DTD) für Print-Wörterbücher entwerfen sie ein lexikographisches Modell, das durch eine modulare Inhaltsstrukturmodellierung flexible Sichten auf die Daten ermöglicht und – im Sinne des „multiple

media publishing“ – unabhängig von den Randbedingungen und Konventionen eines bestimmten Trägermediums bleibt.

Der flexible und interaktive Zugriff auf lexikographische Daten ist auch eine der zentralen Zielsetzungen des Hypertext-Konzepts von *Angelika Storrer*. Sie erläutert die wesentlichen Merkmale von Hypertext-Wörterbüchern und diskutiert dann in Thesenform, wie das Mehrwertpotential des neuen Mediums für die Erarbeitung innovativer digitaler Wörterbücher optimal ausgeschöpft werden kann. Um die Mehrwerteigenschaften, die speziell das Internet und sein hypermedialer Dienst World Wide Web (WWW) gegenüber dem traditionellen Medium Buch aufweist, geht es in dem Beitrag von *Ingrid Lemberg*. Durch einschlägige Beispiele veranschaulicht sie Chancen und Perspektiven, die das neue Publikationsmedium vor allem für die wissenschaftliche Lexikographie eröffnet. Der Beitrag gibt einen Überblick über den aktuellen Stand der Entwicklungen im deutschsprachigen Raum und greift künftige Forschungsfragen in diesem Bereich auf. Komplementär dazu diskutiert *Annette Klosa* in ihrem Beitrag grundlegende Aspekte der Publikation von Wörterbüchern auf CD-ROM aus der Perspektive der Verlagslexikographie. Unter Bezugnahme auf einschlägige Rezensionen entwickelt sie Kriterien zur Qualitätsbewertung von CD-ROM-Wörterbüchern und erläutert, wie diese in der Praxis der kommerziellen Wörterbuchherstellung umgesetzt werden können und wie sich die medialen Veränderungen auf die Arbeit der Lexikographinnen und Lexikographen in der Verlagslexikographie auswirken.

Im zweiten Teil des Bandes werden theoretische und methodische Aspekte beim praktischen Einsatz von SGML/XML, Internet und Hypertext/Hypermedia am Beispiel konkreter Projekte diskutiert; der Schwerpunkt liegt dabei auf der wissenschaftlichen Lexikographie des Deutschen. Ein umfangreiches lexikalisches Informationssystem zur deutschen Gegenwartssprache, das Projekt LEKSIS bildet den Hintergrund für die im Beitrag von *Ulrike Haß-Zumkehr* angestellten Überlegungen zur Mikrostruktur im Hypertext-Wörterbuch. Sie macht deutlich, welche methodischen Herausforderungen mit der Nutzung des Hypertext-Konzepts verbunden sind und welche neuen und interessanten sprachtheoretischen und metalexikographischen Fragestellungen sich ergeben. Während LEKSIS bei der Datenmodellierung nicht an ein vorgängig vorhandenes Print-Wörterbuch gebunden ist, geht es in den folgenden Beiträgen um Aspekte der digitalen Aufbereitung von bereits gedruckten Wörterbüchern. Der Beitrag von *Thomas Gloning* und *Rüdiger Welter* beschreibt am Beispiel des Goethe-Wörterbuchs, wie die vielfältigen Aspekte der Architektur eines Wortschatzes, die durch die alphabetischen Anordnung der Lemmata oft verdeckt werden, durch eine mehrdimensionale SGML-Annotierung der Print-Vorlage expliziert werden können und inwiefern eine solche Aufbereitung die Nutzerinteressen besser und flexibler bedienen kann als die gedruckte Publikation. Der Beitrag von *Thomas Burch* und *Johannes Fournier* basiert auf den Erfahrungen in einem großen Projekt zur retrospektiven Digitalisierung mittelhochdeutscher Wörterbücher. Der Schwerpunkt liegt dabei auf den Vorzügen, Schwierigkeiten und Nachteilen, die sich im Zuge der Anwendung der bereits im Beitrag von *Schmidt* und *Müller* (s.o.) diskutierten Vorgaben der Text Encoding Initiative (TEI) ergeben haben. Das Mittelhochdeutsche ist auch Gegenstand des Beitrags von *Ralf Plate* und *Ute Recker*, allerdings geht es dabei nicht um die retrospektive Digitalisierung gedruckter Wörterbücher, sondern um die Produktions- und Publikationsprozesse eines neuen mittelhochdeutschen Belegwörterbuchs, das von Beginn an computergestützt erarbeitet wird. Die Autoren diskutieren Komponenten und Verfahren des verwendeten Publikationssystems und beschäftigen sich mit konzeptionellen Fragen der geplanten elektronischen Publikation.

Die Vorteile und Mehrwerte von Hypertext und Hypermedia für die wissenschaftliche Lexikographie werden von *Gerd Richter* am Beispiel der Hypertextualisierung des Hessischen Flurnamenwörterbuchs sehr eindrücklich aufgezeigt. Der Beitrag diskutiert Fragen der Integration digitalisierter lexikographischer Quellen in das Hypertextwörterbuch sowie Strategien der Konversion von Verweisangaben in Hypertext-Verknüpfungen (Links). Der Beitrag von *Krzysztof Petelenz* nähert sich dem Thema „Hypertext“ aus der Perspektive der zweisprachigen Lexikographie. Zunächst werden das konzeptionelle Datenmodell und die Architektur eines polnisch-deutschen Online-Wörterbuchs skizziert, das mit dem Hypertextsystem SchemaText entworfen und verwaltet wird. Im Detail wird dann die Modellierung von Mehrwortlexemen und Kollokationen sowie der Rolle von Abbildungen im Hypermedia-Wörterbuch erörtert. *Claudia Kunze* und *Andreas Wagner* beschreiben in ihrem Beitrag den Aufbau, die Informationstypen und die Anwendungsmöglichkeiten des lexikalisch-semantischen Netzes GermaNet. Die in Anlehnung an das englische WordNet organisierte und in einem übergreifenden europäischen Projekt (EuroWordNet) multilingual vernetzte Ressource kann nicht nur von Menschen, sondern auch für die maschinelle Sprachverarbeitung genutzt werden – zur semantischen Disambiguierung, zur Verschlagwortung von Texten sowie zur Akquisition von linguistischen Informationen aus Textkorpora. Abschließend verdeutlicht die Fallstudie von *Lothar Lemnitzer*, in der die Zugriffsprotokolle von zwei im WWW publizierten bilingualen Wörterbüchern ausgewertet wurden, dass das Internet nicht nur ein schnelles und kostengünstiges Publikationsmedium für Wörterbücher ist, sondern auch interessante Perspektiven für die Wörterbuchbenutzungsforschung eröffnet.

Die Idee zu diesem Sammelband entstand durch Diskussionen im Anschluss an das Symposium *Computergestützte Produktion und Publikation von Wörterbüchern*, das wir mit der Unterstützung der Heidelberger Akademie der Wissenschaften, der Gesellschaft für linguistische Datenverarbeitung (GLDV) und dem Institut für deutsche Sprache in Mannheim im September 1998 in Heidelberg veranstaltet haben. Unser herzlicher Dank gilt allen Autorinnen und Autoren für die Beiträge zu diesem Band sowie Kurt Thomas für seine verdienstvolle Arbeit am Layout, an der Indexierung und an der Herstellung der reprodiziblen Druckvorlage. Bei der Herstellung des Bandes hat uns dankenswerterweise die GLDV finanziell unterstützt. Für die Aufnahme in die Reihe *Lexicographica*, Series Maior, danken wir den Herausgebern der Reihe.

Heidelberg/Bonn, Oktober 2000

*Ingrid Lemberg
Bernhard Schröder
Angelika Storrer*