

Studien zur Deutschen Sprache

FORSCHUNGEN DES INSTITUTS FÜR DEUTSCHE SPRACHE



Carolin Müller-Spitzer

# Der lexikografische Prozess

Konzeption für die Modellierung der Datenbasis



Gunter Narr Verlag Tübingen



# Studien zur Deutschen Sprache

FORSCHUNGEN DES INSTITUTS FÜR DEUTSCHE SPRACHE

---



Herausgegeben von

Arnulf Deppermann, Ulrich Hermann Waßner und Stefan Engelberg

Band 42 · 2007

Carolin Müller-Spitzer

# Der lexikografische Prozess

Konzeption für die Modellierung der Datenbasis

**gn**V Gunter Narr Verlag Tübingen

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.d-nb.de>> abrufbar.

Dissertation, Neuphilologische Fakultät der Universität Heidelberg, 2005.

Ursprünglicher Titel: Der lexikografische Prozess. Konzept für die Modellierung einer lexikografischen Datenbasis.

© 2007 · Narr Francke Attempto Verlag GmbH + Co. KG

Dischingerweg 5 · D-72070 Tübingen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Gedruckt auf säurefreiem und alterungsbeständigem Werkdruckpapier.

Internet: <http://www.narr.de>

E-Mail: [info@narr.de](mailto:info@narr.de)

Satz: Hohwieler, Mannheim

Druck und Bindung: Hubert & Co., Göttingen

Printed in Germany

ISSN 0949-409X

ISBN 978-3-8233-6357-6

# Inhalt

|   |    |
|---|----|
| <b>Vorwort</b> .....  | 9  |
| <b>1. Das neue Medium und die Lexikografie: Eine Bandbreite an Möglichkeiten</b> .....                | 11 |
| <b>2. Zur Verortung und Eingrenzung des Themas dieser Arbeit</b> .....                                | 15 |
| <b>3. Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikografischen Prozessen</b> ..... | 21 |
| 3.1 Zum Thema des Kapitels .....  | 21 |
| 3.2 Arten elektronischer Wörterbücher.....  | 23 |
| 3.2.1 Müller, Edelmüller und Sondermüller .....   | 23 |
| 3.2.2 Unterscheidungsmöglichkeiten elektronischer Wörterbücher:<br>Bereits vorgelegte Typologien..... | 27 |
| 3.2.3 Was ist ein elektronisches Wörterbuch? .....  | 29 |
| 3.2.4 Elektronisches Wörterbuch – Wortschatzinformationssystem .....                                  | 32 |
| 3.2.5 Automatisch erstellte vs. lexikografisch bearbeitete<br>Wortschatzinformationssysteme .....     | 33 |
| 3.3 Arten von lexikografischen Prozessen.....   | 42 |
| 3.3.1 All in one: Die Duden-Ontologie.....  | 42 |
| 3.3.2 Arten von lexikografischen Prozessen nach H.E. Wiegand.....                                     | 43 |
| 3.3.3 Arten von lexikografischen Prozessen. Eine erweiterte Übersicht ....                            | 46 |
| 3.4 Ebenen im lexikografischen Prozess.....   | 53 |
| 3.5 Relevanz der getroffenen Unterscheidungen .....   | 56 |
| <b>4. Anforderungen an die Modellierung lexikografischer Daten</b> .....                              | 59 |
| 4.1 Funktionalitäten von Wortschatzinformationssystemen .....   | 60 |
| 4.2 Unterstützung im lexikografischen Prozess .....   | 68 |
| 4.3 Schnittstelle zur automatischen Sprachverarbeitung .....  | 72 |
| 4.4 Allgemeine Eigenschaften der Datenhaltung.....  | 73 |
| <b>5. Die XML-basierte Modellierung einer lexikografischen Datenbasis</b> .....                       | 77 |
| 5.1 Begründung der Wahl von XML und verwandten Standards.....   | 79 |
| 5.1.1 Was ist XML?.....   | 79 |
| 5.1.2 Vorteile des Einsatzes von XML .....  | 82 |
| 5.1.3 Einige Bemerkungen zum Gesamtaufbau einer XML-basierten<br>Redaktionsumgebung.....              | 87 |

|           |   |     |
|-----------|---|-----|
| 5.1.4     | XML als „Piece of a Larger Picture“ .....   | 89  |
| 5.1.5     | Notwendigkeit von Richtlinien für die Anwendung von XML .....                                   | 91  |
| 5.2       | Differenzierung verschiedener Ansätze einer DTD-Entwicklung .....                               | 95  |
| 5.2.1     | Maßgeschneiderte vs. Standard-DTD.....  | 95  |
| 5.2.2     | Strukturierung eines bestehenden Datenbestandes vs.<br>Neuerstellung eines Datenbestandes ..... | 99  |
| 5.2.3     | Layoutorientierte vs. inhaltsorientierte Datenauszeichnung .....                                | 100 |
| 5.2.4     | Layoutbasierte vs. konzeptuelle Inhaltsmodellierung.....  | 103 |
| 5.2.5     | Angemessenheit der Modellierung .....   | 106 |
| 5.2.6     | Strenge vs. weiche Struktur .....   | 107 |
| 5.3       | Rahmenbedingungen und Zielsetzungen des<br>Modellierungskonzeptes .....                         | 109 |
| 5.4       | Modellierung des Inhaltsstrukturenprogramms .....   | 117 |
| 5.4.1     | Inhaltliche Konzeption.....   | 117 |
| 5.4.2     | Standardbasierte Ausdrucksmöglichkeit: XML-DTDs .....   | 132 |
| 5.4.2.1   | Graphen und Bäume .....   | 132 |
| 5.4.2.2   | Kurzeinführung in die Syntax von XML.....   | 135 |
| 5.4.2.3   | „The Well-Dressed-DTD“: einige Voraussetzungen<br>für eine von Menschen lesbare DTD .....       | 146 |
| 5.4.2.3.1 | Optische Gestaltung .....   | 146 |
| 5.4.2.3.2 | Verteilung von Elementen, Attributen und Entities .....   | 148 |
| 5.4.2.3.3 | Modularisierung von DTDs.....   | 149 |
| 5.4.2.3.4 | Kommentierung der DTD.....  | 150 |
| 5.4.2.3.5 | Dokumentation der DTD.....  | 151 |
| 5.4.2.3.6 | Sprache.....  | 152 |
| 5.4.3     | Richtlinien für die XML-basierte Modellierung des<br>Inhaltsstrukturenprogramms .....           | 152 |
| 5.4.3.1   | Modellierung des hierarchischen Aufbaus des<br>Inhaltsstrukturenprogramms .....                 | 153 |
| 5.4.3.2   | Modellierung der terminalen Elemente.....   | 158 |
| 5.4.3.2.1 | Unterscheidung in Angaben <sub>v</sub> , Kommentare und Hinweise ...                            | 158 |
| 5.4.3.2.2 | Modellierung von Angaben <sub>v</sub> .....   | 160 |
| 5.4.3.2.3 | Modellierung von Kommentaren.....   | 167 |
| 5.4.3.2.4 | Modellierung von Hinweisen .....  | 168 |
| 5.4.3.2.5 | Nicht gekennzeichnete terminale Elemente .....  | 169 |
| 5.5       | Modellierung des Vernetzungsstrukturenprogramms .....   | 169 |
| 5.5.1     | Inhaltliche Konzeption.....   | 170 |
| 5.5.2     | Standardbasierte Ausdrucksmöglichkeit: XLink .....  | 180 |
| 5.5.3     | Richtlinien für die XML-basierte Modellierung des<br>Vernetzungsstrukturenprogramms .....       | 190 |

|           |   |            |
|-----------|---|------------|
| 5.6       | Mehr als ein Instruktionenbuch: Die Anwendung des Modellierungskonzeptes im lexikografischen Prozess .....    | 204        |
| 5.7       | Flexibilität der Darstellung der entsprechend modellierten Daten .....  | 210        |
| 5.7.1     | Grundsätzliches .....   | 210        |
| 5.7.2     | Standardbasiertes Verfahren: Darstellung mit XSL(T) .....   | 212        |
| 5.7.3     | Veranschaulichung anhand eines Beispiels .....  | 216        |
| 5.7.4     | Umgang mit Verdichtung in medienneutral konzipierten lexikografischen Prozessen .....                         | 229        |
| 5.8       | Perspektive Wörterbuchverbund .....   | 233        |
| <b>6.</b> | <b>Modellierung einer lexikografischen Datenbasis im Projekt <i>ellexiko</i>: ein Werkstattbericht .....</b>  | <b>243</b> |
| 6.1       | Kurzbeschreibung des Projekts .....   | 243        |
| 6.2       | Rolle des Modellierungskonzeptes bei der DTD-Entwicklung .....  | 246        |
| 6.3       | Vorgehensweise bei der DTD-Entwicklung .....  | 248        |
| 6.4       | Skizze der technischen Redaktionsumgebung .....   | 249        |
| 6.5       | Die Modellierung des Inhaltsstrukturenprogramms .....   | 250        |
| 6.5.1     | Grundsätzliche Aufteilung der DTDs .....  | 250        |
| 6.5.2     | Grundstruktur eines Artikels im <i>ellexiko</i> -Wörterbuch (bearbeitete Artikel) .....                       | 253        |
| 6.5.3     | Anwendung der Richtlinien zur Modellierung des Inhaltsstrukturenprogramms .....                               | 259        |
| 6.5.3.1   | Modellierung des hierarchischen Aufbaus des Inhaltsstrukturenprogramms .....                                  | 259        |
| 6.5.3.2   | Modellierung der terminalen Elemente .....  | 261        |
| 6.5.3.2.1 | Unterscheidung in Angaben <sub>v</sub> , Kommentare und Hinweise .....  | 261        |
| 6.5.3.2.2 | Modellierung von Angaben <sub>v</sub> .....   | 264        |
| 6.5.3.2.3 | Modellierung von Kommentaren .....  | 271        |
| 6.5.3.2.4 | Modellierung von Hinweisen .....  | 273        |
| 6.6       | Modellierung des Vernetzungsstrukturenprogramms .....   | 276        |
| 6.7       | Weitere Aspekte einer „Well-Dressed-DDT“ .....  | 287        |
| 6.7.1     | Benennung der Objekte .....   | 287        |
| 6.7.2     | Anordnung der Entities, Elemente und Attribute und Kommentierung der DTDs .....                               | 288        |
| 6.7.3     | Dokumentation der DTDs .....  | 289        |
| 6.8       | Darstellung der Artikel und Perspektiven für die Recherche .....  | 292        |
| 6.9       | Die DTD-Entwicklung im Spannungsfeld zwischen theoretischen Richtlinien und ihrer praktischen Umsetzung ..... | 298        |

|     |                               |     |
|-----|-------------------------------|-----|
| 7.  | <b>Schlussbemerkung</b> ..... | 301 |
| 8.  | <b>Literatur</b> .....        | 303 |
| 8.1 | Wörterbücher .....            | 303 |
| 8.2 | Forschungsliteratur .....     | 303 |

## Vorwort

Dieses Buch ist eine überarbeitete Fassung meiner im Juni 2005 an der Heidelberger Universität eingereichten Dissertation. Zum Gelingen dieser Arbeit haben viele Menschen beigetragen, bei denen ich mich hiermit – zumindest auszugsweise – bedanken möchte: Zunächst bedanke ich mich bei Herbert Ernst Wiegand für die interessanten, spannenden und auf gewisse Weise auch unterhaltsamen Seminare während meiner Studienzeit, die mich auf die Fährte der (Meta-)Lexikografie gebracht haben. Dabei habe ich besonders seine breit angelegte fachliche Neugierde schätzen gelernt, die auch einer solchen Arbeit auf der Grenze zwischen Metalexikografie und Texttechnologie offen gegenübersteht. Außerdem möchte ich meinen Kollegen und Kolleginnen am Institut für Deutsche Sprache für Anregungen, Unterstützung und kollegiale Kritik danken; genauso den anderen in- und ausländischen Kollegen im Arbeitsfeld der Wörterbuchforschung und Lexikografie.

Ein großes Projekt wie eine Doktorarbeit, die einige Zeit in Anspruch nimmt, wird jedoch auch immer vom persönlichen Umfeld getragen. Deshalb möchte ich mich als erstes bei meinen Eltern bedanken, die mich meinen eigenen Weg haben gehen lassen und dabei immer ein sicheres Netz für mich waren und sind. Ihnen habe ich viel zu verdanken. Ich bedanke mich nicht weniger bei meinem Mann Sebastian, der im Alltag durch Unterstützung in vielerlei Hinsicht sicher den größten Beitrag zum Gelingen der Arbeit geleistet hat. Auch möchte ich mich bei unseren Töchtern Marie und Anna bedanken, die auf ihre Weise die Entstehung der Doktorarbeit und die anschließenden Prüfungen befördert haben. Nicht zuletzt bedanke ich mich bei meinen Freundinnen und Freunden, insbesondere bei Eva und Ingrid, für ihre Zuneigung und Verbundenheit.

Und schon im Voraus bedanke ich mich bei allen Lesern des Buches für ihr Interesse und wünsche allen nach soviel Dankesworten eine hoffentlich anregende Lektüre.

Ich widme diese Arbeit meiner Familie.

Carolin Müller-Spitzer

Plankstadt, September 2007



## 1. Das neue Medium und die Lexikografie: Eine Bandbreite an Möglichkeiten

Der Computer kann [...] Wörterbücher rationalisieren, systematisieren, präzisieren, symmetrisieren, parametrisieren, modularisieren, homogenisieren, technisieren, operationalisieren, automatisieren und propagandieren – mit einem Wort, optimieren.

*Francis Knowles*

Die guten Leute, die sich über die technischen Fortschritte unserer Zeit so entsetzen, verwechseln, wie mir scheint, Zweck und Mittel. Wer nur um Gewinn kämpft, erntet nichts, was der Mühe wert ist. Aber die Maschine ist kein Ziel und darum ist auch das Flugzeug kein Zweck, sondern ein Werkzeug, ein Gerät, nichts anderes als der Pflug.

*Antoine de Saint-Exupéry*

Das elektronische Medium, welches so neu gar nicht mehr ist, hat die gesamte Publikationslandschaft tiefgreifend verändert. Es verheißt auch der Lexikografie eine Bandbreite an Möglichkeiten. Dabei kann der Computer jedoch nicht selbst aktiv werden, um diese Möglichkeiten zu nutzen; der Computer selbst kann nichts. Akteur bleibt der Mensch, der den Rechner als neues technisches Werkzeug zur Erreichung der von ihm definierten Handlungsziele einsetzen kann. Es gilt daher zu prüfen, wie der Einsatz des neuen Mediums so gestaltet werden kann, dass er nicht Selbstzweck, sondern sinnvoller Gebrauch dieses neuen Werkzeuges ist.

Das gilt auch für die Sprachlexikografie als eigenständige kulturelle Praxis, die hier Gegenstand der Betrachtung ist.<sup>1</sup> Hier wird das Potenzial des elektronischen Mediums für Wörterbücher besonders hoch eingeschätzt.

[...] It should not come as a surprise that nearly everyone involved in present-day dictionary making is enthusiastic about the potential of the electronic medium. ED [Electronic Dictionary, Anm. d. Verf.] dreams are indeed not without a solid basis. (Schryver 2003, S. 162)

---

<sup>1</sup> Vgl. Wiegand (1998a, S. 41). Näheres dazu siehe Kapitel 2.

Als Gründe werden – sehr allgemein formuliert – in der Forschung u.a. folgende Eigenschaften und Gebrauchsweisen von Wörterbüchern genannt: Wörterbücher werden nicht linear gelesen, sondern in ihnen wird punktuell nachgeschlagen, Wörterbuchartikel sind auf komplexe Weise miteinander vernetzt, was auf Papier oft nicht hinreichend umgesetzt werden kann, und Artikel von Sprachwörterbüchern sind in der Regel standardisiert in einzelne Positionen aufgeteilt, die meist nicht alle von einem Benutzer zur Beantwortung einer Suchanfrage zur Kenntnis genommen werden. In der Umsetzung dieser Eigenschaften wird dem elektronischen Medium ein potenzieller Mehrwert gegenüber dem Druckmedium zugesprochen.<sup>2</sup>

Neue Möglichkeiten hinsichtlich der Funktionalität werden in der Praxis jedoch nur langsam genutzt, wie es bei großen Medienveränderungen durch die Geschichte fast immer zu beobachten ist (vgl. Faulstich 1997). Sicherlich hat sich in den letzten Jahren in der lexikografischen Praxis viel verändert, sodass man nicht mehr unbedingt sagen kann, dass es sich bei den meisten elektronischen Wörterbüchern um eine Widerspiegelung der Printwörterbücher auf „reduziertem Niveau“ (Feldweg 1997, S. 110) handelt. Doch elektronische Wörterbücher ziehen gerade ihre Berechtigung daraus, sich sehr deutlich von gedruckten zu unterscheiden, indem sie die Möglichkeiten des neuen Mediums wirklich sinnvoll nutzen (vgl. auch Lehr 2000). Jedoch sind Beispiele für ganz grundsätzlich andere, neue Funktionalitäten elektronischer Wörterbücher gegenüber Printwörterbüchern, so wie sie z.B. Atkins fordert, nach wie vor kaum zu finden:

If new methods of access (breaking the iron grip of the alphabet) and a hypertext approach to the data stored in the dictionary do not result in a product light years away from the printed dictionary, then we are evading the responsibilities of our profession. (Atkins 1992, S. 521; zitiert nach Schryver 2003, S. 144)

Das neue Medium spielt in der Lexikografie jedoch nicht nur als Publikationsmedium eine Rolle. Ein wichtiger Bereich ist auch der Einsatz neuer Technologien in der Herstellung von Wörterbüchern, also im lexikogra-

---

<sup>2</sup> Einen sehr guten Überblick über das, was in der Literatur in Bezug auf die Möglichkeiten des elektronischen Mediums für die Lexikografie formuliert wird, bietet Schryver (2003). Aus der deutschsprachigen Wörterbuchforschung sind u.a. Storrer (2001) und die Beiträge aus Lemberg/Schröder/Storrer (Hg.) (2001) zu nennen. Zahlreiche Beispiele für Online-Wörterbücher sind über das Wörterbuchportal von Robert Beard und Paul J.J. Payack, [www.yourdictionary.com](http://www.yourdictionary.com) (Stand: Februar 2007), zu finden.

fischen Prozess. Dabei geht es nicht nur um die Herstellung von elektronischen Wörterbüchern, sondern genauso um die Herstellung von gedruckten Wörterbüchern oder von Wörterbüchern, die gleichzeitig in beiden Medien erscheinen. Ziel ist auch hier, im lexikografischen Prozess qualitative Verbesserungen durch den Einsatz des Computers zu erreichen, nicht nur quantitative Erleichterungen im Sinne einer schnelleren und reibungsloseren Herstellung. Doch können aus einfachen Quantitätssteigerungen nicht direkt Qualitätssteigerungen erwartet werden.

[...] Im computerunterstützten lexikografischen Prozess kann die Quantität niemals in Qualität (sozusagen 'von allein') umschlagen. Vielmehr kommt die Qualitätssteigerung im Beispiel dadurch zu Stande, dass ein aufgabengerechter und intelligenter Einsatz des Computers erfolgte, sodass die Datenverarbeitung zu einer Organisation des Materials führte, die eine bessere und zusammenhängende Wahrnehmung dieses Materials unter verschiedenen, für die Lösung der Aufgabe relevanten Perspektiven erlaubte, was eine wichtige Voraussetzung für eine adäquate Materialinterpretation durch die beiden Lexikografen darstellte. Es wäre wohl falsch zu folgern, die Qualitätssteigerung sei allein das Ergebnis der Materialinterpretation der beiden Lexikografen, auch wenn diese präzise Fragen stellten; vielmehr ist sie das Ergebnis einer gelungenen Mensch-Maschine-Kommunikation. (Wiegand 1998a, S. 160)

Das elektronische Medium bietet der Lexikografie also nicht schon an sich eine schöne neue Welt, sondern es muss etwas dafür getan werden, dass diese vielversprechenden Möglichkeiten die lexikografische Praxis verbessern und so zum Entstehen innovativer Produkte beitragen können. Dabei sind die beiden Bereiche – neue technische Möglichkeiten im lexikografischen Prozess und neues Potenzial bei der Publikation lexikografischer Daten – nicht losgelöst voneinander zu betrachten, sondern stehen in einem engen Verhältnis zueinander. Denn erst durch einen sinnvollen Einsatz des Computers im lexikografischen Prozess liegen die lexikografischen Daten in der Regel so vor, dass sie 'mediengerecht' im elektronischen Medium präsentiert werden können und – falls gewünscht – gleichzeitig auch ein gedrucktes Wörterbuch entstehen kann.

Es geht in dieser Arbeit daher allgemein um die Frage: Wie können neue technische Möglichkeiten so im lexikografischen Prozess eingesetzt werden, dass der Computer eine gute Unterstützung für die lexikografische Arbeit bietet und gleichzeitig eine Datenbasis entsteht, aus der neuartige elektronische Produkte entwickelt werden können? Eine Fragestellung ist in dieser Form jedoch noch ein zu 'weites Feld'. Ein wesentlicher Punkt im lexikogra-

fischen Prozess ist die Frage der Datenorganisation. Speziell die Aufgabe der Datenmodellierung ist, je nachdem, wie man Datenmodellierung definiert, als Aufgabe neu hinzugekommen bzw. hat sich grundlegend verändert. Meines Erachtens werden dabei in der Ausgestaltung der Datenmodellierung die entscheidenden Weichen dafür gestellt, ob die lexikografischen Daten am Ende so vorliegen, dass aus ihnen ein innovatives elektronisches Wörterbuch entwickelt werden kann und gleichzeitig eine gedruckte Präsentation dieser Daten grundsätzlich möglich ist. Darüber hinaus kann eine bestimmte Form von Datenmodellierung und entsprechender Datenerfassung wesentliche Qualitätsverbesserungen für die lexikografische Arbeitsumgebung bewirken. Allerdings ist bisher noch nicht umfassend aufgezeigt worden, wie eine solche Datenmodellierung und die darauf aufbauenden Verbesserungsmöglichkeiten in der lexikografischen Praxis aussehen können. Wie man eine solche Datenmodellierung entwickeln kann, welchen Richtlinien man dabei folgen sollte, welcher technischen Möglichkeiten man sich bei der Entwicklung bedienen kann, sind daher hier zu klärende Fragen. Das Thema dieser Arbeit ist somit darauf eingegrenzt, ein Konzept für die Modellierung einer lexikografischen Datenbasis zu entwickeln, die in benannten Bereichen Verbesserungsmöglichkeiten bewirken kann, und exemplarisch aufzuzeigen, wie solche Verbesserungen aussehen können.

Diese allgemeine Themenbeschreibung soll nun genauer eingegrenzt und innerhalb der Lexikografie und Wörterbuchforschung verortet werden.

## 2. Zur Verortung und Eingrenzung des Themas dieser Arbeit

Im Laufe meiner Beschäftigung mit Wörterbüchern [...] bin ich allmählich zu der Überzeugung gelangt, dass eine möglichst weitgehende theoretische und methodologische Durchleuchtung einer althergebrachten [...], pragmatisch eingespielten Praxis zu einem vertieften Verständnis dieser Praxis, zu einem höheren Grad ihrer Beherrschung und damit in manchen Bereichen auch zu ihrer Erneuerung führen kann.

*Herbert Ernst Wiegand*

Gegenstand dieser Untersuchung ist der lexikografische Prozess, genauer die Datenmodellierung im lexikografischen Prozess. Dabei sind nicht beliebige Arten von lexikografischen Prozessen im Blickpunkt, sondern solche, die zu Sprachnachschlagewerken führen. Unter diesen stehen wiederum die wissenschaftlichen Wörterbuchprojekte im Vordergrund, da hier eher der Raum für eine fundierte Konzeption bei der Modellierung der lexikografischen Daten gegeben ist. Für die genaue Erläuterung des Aufbaus dieser Arbeit ist es hilfreich, dieses Thema innerhalb der Lexikografie und Wörterbuchforschung zu verorten. Dazu sollen zunächst – anschließend an Wiegand – Sprachlexikografie und Wörterbuchforschung gegeneinander abgegrenzt werden.

Die wissenschaftliche Form der Sprachlexikografie kann mit Wiegand als eine „eigenständige kulturelle und wissenschaftliche Praxis“ (Wiegand 1998a, S. 62) charakterisiert werden. Sprachlexikografie als solche ist danach keine Wissenschaft, da „das Ziel der Sprachlexikographie [...] nicht darin“ besteht, „wissenschaftliche Theorien zu erarbeiten“ (ebd., S. 23). Natürlich können Lexikografen auch sprachwissenschaftliche Tätigkeiten ausführen. Dazu bringt Wiegand das Beispiel der Etymologie. Wenn ein etymologisches Wörterbuch erarbeitet würde, dann sei die „Erarbeitung einer Etymologie und damit die von etymologischem Wissen eine *sprachwissenschaftliche* Tätigkeit“, die damit „auch eine Tätigkeit eines bestimmten Lexikographen sein“ kann.

Die lexikographische Tätigkeit besteht u.a. in einer Verarbeitung des verfügbaren etymologischen Wissens zu einem lexikographischen Text. Die Erarbeitung einer Etymologie einerseits und deren Darstellung in einem Wörterbuchartikel andererseits sind (auch wenn sie von einer Person ausgeführt werden und ineinander greifen können) zwei verschiedene und deutlich unterscheidbare Dinge. (Wiegand 1998a, S. 23)

In der wissenschaftlichen Sprachlexikografie werden „in unterschiedlichem Ausmaß wissenschaftliche Methoden aus *verschiedenen* Wissenschaften angewendet und wissenschaftliche Ergebnisse aus *verschiedenen* Wissenschaften genutzt“ (Wiegand 1998a, S. 40). Nicht nur Ergebnisse z.B. aus der Lexikologie können demnach angewandt werden, sondern wie später zu sehen sein wird, auch Ergebnisse beispielsweise aus dem Bereich der Texttechnologie. Von lexikografischen Tätigkeiten deutlich zu unterscheiden sind aber solche

Tätigkeiten, die nicht als Teil eines lexikographischen Prozesses darauf ausgerichtet sind, daß Sprachnachschlagewerke entstehen, sodaß sie nicht mit dem Schreiben von Wörterbüchern in einem erkennbaren Handlungszusammenhang stehen (ebd., S. 45).

Diese Tätigkeiten solle „man nicht Lexikografie nennen“, da man „Wörterbücher nicht theoretisch schreiben“ kann (ebd., S. 45). Selbstverständlich enthält lexikografische Tätigkeit reflexive Komponenten, die aber immer noch zur praktischen Tätigkeit der Sprachlexikografie gezählt werden sollen (ebd., S. 46).

Die verschiedenen wissenschaftlichen Formen der Beschäftigung mit Sprachlexikografie sind daher nicht selbst Teil der Sprachlexikografie, sondern bilden den wissenschaftlichen Metabereich hierzu (Wiegand 1998a, S. 64). Die Sprachlexikografie ist demnach der „Gegenstandsbereich“, „aus dem für die *verschiedenen Forschungsgebiete* der Wörterbuchforschung die wissenschaftlichen Gegenstände gewonnen werden“ (ebd., S. 46). Ein Bereich der Wörterbuchforschung ist dabei auch der Einsatz des Computers im lexikografischen Prozess. „In diesem Sinne kann die Wörterbuchforschung als die Gesamtheit aller, im Schrifttum nachweisbaren, wissenschaftlichen Bemühungen aufgefaßt werden, die darauf abzielen, diejenigen theoretischen, methodischen, terminologischen, historischen, dokumentarischen, didaktischen und kulturpädagogischen Fragen zu beantworten, die sich stellen“ u.a. „beim Einsatz des Computers in der Lexikographie und Wörterbuchforschung“ (ebd., S. 79). Sinn und Ziel dieser Unterteilung von Sprachlexikografie und Wörterbuchforschung ist bei Wiegand nicht die „unfruchtbare Aufteilung in Praktiker und Theoretiker“, sondern „die Unterscheidung zweier Arten von Tätigkeiten“ (ebd., S. 88). Die Abgrenzung dieser beiden Arten von Tätigkeiten nach Wiegand ist deshalb referiert, um das hier behandelte Thema den verschiedenen Bereichen zuordnen zu können.

Einen großen Teil dieser Arbeit macht die Entwicklung eines Konzepts für die Modellierung einer lexikografischen Datenbasis aus, die nicht auf einen bestimmten lexikografischen Prozess bezogen ist, sondern auf möglichst viele konkrete lexikografische Prozesse anwendbar sein sollte. Dabei soll das Konzept so theoretisch fundiert sein, dass es nachvollziehbar wird. Die Entwicklung dieses Konzepts für eine Datenmodellierung ist daher der Wörterbuchforschung zuzurechnen, da sie nicht mit dem Ziel verbunden ist, ein konkretes Wörterbuch zu erarbeiten. Insofern ist die Arbeit auch als „Beitrag zur Neuordnung lexikographischer Theoriebildung unter Bezugnahme auf die Möglichkeiten des Rechnereinsatzes“ (Geeb 2001, S. 56) zu sehen. Da die Intention, die hinter der Entwicklung dieses Konzepts steht, die Verbesserung der lexikografischen Praxis ist, wird die Umsetzung dieses Konzepts an einem konkreten sprachlexikografischen Projekt illustriert. Diese Entwicklung einer Datenmodellierung in einem konkreten lexikografischen Prozess kann der Sprachlexikografie zugeordnet werden, denn sie ergänzt und ersetzt teilweise die früheren Instruktionsbücher und Arbeitsanweisungen. Zu Letzteren bemerkt Wiegand:

Die Ergebnisse der lexikographischen Tätigkeiten werden zum lexikografischen Prozeß gerechnet, und zwar nicht nur die Endprodukte, die Wörterbücher, sondern auch – und dies gilt speziell für die wissenschaftliche Lexikographie – alle Teil- oder Zwischenergebnisse wie Arbeitsanweisungen, Instruktionen, Protokolle, Konzeptionspapiere, Probeartikel, Musterartikel, Korrekturunterlagen, Werkstattberichte usw., also alle jene Schriftstücke, in denen auch die selbstreflexive Komponente der Lexikographie als einer eigenständigen, kulturellen Praxis konkret faßbar wird. (Wiegand 1998a, S. 77)

Nun ist z.B. die Modellierung einer Datenbank nicht ein Schriftstück wie eine Arbeitsanweisung, aber sie ist Teil eines konkreten lexikografischen Prozesses; ihre Entwicklung ist daher als lexikografische Tätigkeit aufzufassen. Der Werkstattbericht im sechsten Kapitel dieser Arbeit ist damit der Sprachlexikografie zuzuordnen, da es sich hier um den Bericht über Ausschnitte aus einem konkreten lexikografischen Prozess handelt.

Das Thema dieser Arbeit ist damit sowohl ein metalexikografisches als auch ein lexikografisches: Einerseits die Entwicklung eines Konzepts für die Modellierung einer lexikografischen Datenbasis; andererseits die Anwendung dieses Konzeptes an einem konkreten Fallbeispiel. Das Modellierungskonzept soll dabei dazu dienen,

- die Lexikografen im lexikografischen Prozess, d.h. bei der Datenerarbeitung und Dateneingabe zu unterstützen,
- die so erarbeiteten lexikografischen Daten in innovativer Weise für Wörterbuchbenutzer im elektronischen Medium zugreifbar zu machen, und
- den Weg offen zu halten, die so erarbeitete lexikografische Datenbasis in gedruckter Form präsentieren zu können.

Um das Konzept zu entwickeln, werden zunächst im folgenden Kapitel einige terminologische und ordnende Überlegungen angestellt, die den Themenbereich dieser Arbeit betreffen, jedoch auch über ihn hinausweisen (vgl. auch Müller-Spitzer 2003). Sie sind deshalb notwendig, um eingrenzen und benennen zu können, welche Arten von elektronischen Nachschlagewerken hier Gegenstand der Betrachtung sind, um welche Arten von lexikografischen Prozessen es geht und welche Ebenen im lexikografischen Prozess im Rahmen dieser Arbeit unterschieden werden. Im vierten Kapitel werden in einem Forschungsüberblick Anforderungen an die Modellierung lexikografischer Daten aufgezeigt. Dieser Überblick soll zusätzlich zu den oben formulierten grundsätzlichen Zielen der Modellierung einer lexikografischen Datenbasis sehr viel detaillierter zusammentragen, was in diesem Zusammenhang von (Meta-)Lexikografen gefordert wird und welchen Ansprüchen damit das in dieser Arbeit entwickelte Konzept genügen muss. Im darauf folgenden Hauptkapitel 5 werden die Richtlinien für die XML-basierte Modellierung einer lexikografischen Datenbasis entwickelt. Wie dabei genau vorgegangen wird und wie die Gliederung des Hauptkapitels motiviert ist, wird am Anfang des Kapitels erläutert. In Kapitel 6 wird in einem Werkstattbericht exemplarisch gezeigt, wie dieses Konzept im Projekt *ellexiko* des Instituts für Deutsche Sprache<sup>3</sup> umgesetzt worden ist.

Ziel dieser Arbeit ist es damit, bekannte Forschungen aus dem Bereich der Printlexikografie mit neuen Verfahren und Techniken aus dem Bereich der XML-Modellierung zu einem eigenständigen, praxisorientierten und theoretisch fundierten Ansatz für ein Konzept der maßgeschneiderten Modellierung einer lexikografischen Datenbasis zusammenzubringen und dabei die Relevanz und Auswirkungen eines solchen Ansatzes für die und anhand der lexikografischen Praxis aufzuzeigen. Damit eng verbunden ist das Ziel,

---

<sup>3</sup> Für Projektinformationen siehe [www.ellexiko.de](http://www.ellexiko.de) (Stand: Februar 2007).

Wege aufzuzeigen, wie bessere elektronische Wörterbücher hergestellt werden können, mit der zusätzlichen Möglichkeit, aus der dafür erarbeiteten Datenbasis auch ein gedrucktes Wörterbuch herzustellen. Die Arbeit bietet damit keine Analyse des Bestehenden, sondern entwickelt einen Leitfaden für eine Modellierung, die zu einer verbesserten lexikografischen Praxis unter Ausnutzung der heutigen technischen und medialen Bedingungen führen soll.



### 3. **Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikografischen Prozessen**

Es gibt nichts rundum Zutreffendes, Eindeutiges und Stichhaltiges, das ich über mich sagen, gar ohne Wenn und Aber in einem einzigen Wort ausdrücken könnte. Ich *unterscheide*, dies ist das A und O meiner Logik.

*Michel de Montaigne, Essais*

#### 3.1 **Zum Thema des Kapitels**

Wie eingangs schon gesagt, hat sich die Wörterbuchlandschaft durch das Hinzukommen des elektronischen Mediums in den letzten ein bis zwei Jahrzehnten massiv verändert. Waren es erst zaghafte Versuche, gedruckte Wörterbücher in sehr ähnlicher Form einfach elektronisch verfügbar zu machen, so steht dem heute eine sehr große Vielfalt an elektronischen Wörterbüchern, Wortschatzinformationssystemen – oder wie immer man sie auch bezeichnen mag – gegenüber. Das elektronische Medium spielt darüber hinaus an den verschiedensten Stellen im lexikografischen Prozess eine Rolle, ob bei der computerunterstützten Erarbeitung gedruckter Wörterbücher, bei der Datengewinnung aus elektronischen Textkorpora oder als Publikationsmedium bei elektronischen Wörterbüchern. Diese neuen Produkte stellen eine Herausforderung an die Wörterbuchforschung dar. Vor allem die Wörterbuchkritik hat hier ein stark erweitertes Feld vor sich und Fragen zu beantworten wie: Können Bewertungsmaßstäbe, die sich für gedruckte Wörterbücher herausgebildet haben, auf elektronische Wörterbücher übertragen werden? Und wenn dies gelten soll, gilt das für alle Arten elektronischer Wörterbücher? Oder wie soll man elektronische Nachschlagewerke klassifizieren, damit hier besser differenziert werden kann? Ist die automatische Gewinnung von Daten aus elektronischen Textkorpora als lexikografische Tätigkeit zu bezeichnen? Was ist mit der Verwendung lexikografischer Daten in der maschinellen Sprachverarbeitung? Gehört Letzteres auch zu Themen der Wörterbuchforschung?

Gerade die letzte Frage ist auch für diese Arbeit wichtig. Denn es ist zu klären, inwiefern in dem hier zu entwickelnden Konzept für die Modellierung einer lexikografischen Datenbasis Belange der automatischen Sprachverarbeitung zu berücksichtigen und zu integrieren sind. Jedoch können und sollen auch die anderen Fragen in diesem Kapitel nicht außer acht gelassen werden. Dabei wird vor allem die Perspektive der möglichen Einordnung elektronischer Wörterbücher eingenommen.

Die hier dargestellten Überlegungen gliedern sich in drei Themenbereiche mit unterschiedlicher Gewichtung: Abschnitt 2 des Kapitels beschäftigt sich mit der Unterscheidung von automatisch erstellten vs. menschlich bearbeiteten elektronischen Wörterbüchern und den Herausforderungen, die so entstehende Produkte u.a. an die Wörterbuchkritik stellen. Warum gerade diese Art der Wörterbucherarbeitung eine besondere Schwierigkeit für die metalexikografische Kritik darstellt, wird an einem vorangestellten Ausgangsbeispiel veranschaulicht. In diesem Zusammenhang wird auch der Versuch von terminologischen Klärungen in Bezug auf elektronische Wörterbücher unternommen, die für den weiteren Verlauf der Arbeit wichtig sind. Der zweite Themenbereich zeigt in einer Gesamtschau Arten lexikografischer Prozesse, wie sie heute in der Wörterbuchlandschaft vorzufinden sind. Auch hier wird ein Beispiel vorangestellt, welches veranschaulichen soll, wie heute lexikografische Prozesse aussehen können und in welche Richtung demnach ein Gesamtüberblick über Arten lexikografischer Prozesse erweitert werden muss. Daneben soll dieser Teil deutlich machen, für welche Arten lexikografischer Prozesse das Modellierungskonzept entwickelt wird. In Abschnitt 4 wird ein Thema herausgegriffen, welches besonders für den medienneutral konzipierten lexikografischen Prozess (zu diesem Terminus siehe 3.3) und den computerlexikografischen Prozess relevant ist: eine mögliche Sicht der Ebenen im lexikografischen Prozess. Dieser Abschnitt ist wiederum die Voraussetzung dafür, anschließend in Kapitel 4 zwei wichtige Vorgaben des hier entwickelten Modellierungskonzepts verdeutlichen zu können. Alle drei Themenbereiche bieten darüber hinaus Einordnungsmöglichkeiten für die Wörterbuchkritik. Dabei sind sie nicht für alle Arten elektronischer Wörterbücher gleichermaßen relevant.

Um es deutlich herauszustellen: Es geht in diesem Kapitel insgesamt nicht um einzelne Kriterien, nach denen elektronische Wörterbücher bewertet werden können. Vorschläge dazu sind u.a. in Lehr (1996), Klosa (2001) und Engelberg/Lemmitzer (2001, bes. S. 194ff.) genannt. Die hier entwickelten terminologischen Unterscheidungen verstehen sich als Ergänzung zu diesen Arbeiten in dem Sinne, dass ein elektronisches Wörterbuch – bevor es im Einzelnen bewertet wird – nach den hier zu entwickelnden Unterscheidungen eingeordnet werden kann. Insgesamt handelt es sich demnach um Betrachtungen, die helfen sollen, den neuen und sehr vielfältigen Gegenstandsbereich der elektronischen Wörterbücher und lexikografischer Prozesse aus verschiedenen Perspektiven zu ordnen.

## 3.2 Arten elektronischer Wörterbücher

### 3.2.1 Müller, Edelmüller und Sondermüller

Die automatische Analyse elektronischer Textkorpora macht die Entwicklung ganz neuer Produkte möglich. Dies ist ein neuer Gegenstandsbereich für die Wörterbuchforschung, da die Möglichkeit, überhaupt solche wortschatzbezogenen Daten in der Weise *automatisch* erarbeiten zu können, durch die umfassende Anwendung des neuen Mediums bedingt ist. Hinzu kommt, dass durch das Internet eine sehr preiswerte Publikationsmöglichkeit zur Verfügung steht, die zudem eine sehr hohe Verbreitung garantiert. Dadurch kommen Produkte auf den Markt, die in der Weise oder in der Form nicht gedruckt worden wären. Welche neuen Fragen dies für die Wörterbuchkritik aufwerfen kann, soll am folgenden Beispiel, dem Projekt *Deutscher Wortschatz* der Universität Leipzig,<sup>4</sup> demonstriert werden.

Es handelt sich nach Projektinformationen hierbei um ein „umfangreiches Vollformenwörterbuch des Deutschen“ (Quasthoff/Wolff 1999, S. 1), in dem

die typischen Inhalte und Funktionen unterschiedlicher Wörterbuch- und Lexikontypen [...] zur Verfügung stehen (Nachschlagen von Begriffen; Querverweise; morphologische, syntaktische, semantische und pragmatische Information, statistische Daten; Einarbeitung von Ontologien) und durch die zusätzlichen Möglichkeiten des elektronischen Mediums (automatische linguistische Analyseverfahren, Recherche, Hypertextualisierung, automatische Generierung unterschiedlich strukturierter Einträge, Visualisierung von Relationen zwischen Einträgen) ergänzt werden. (ebd.)

Auf der Webseite wird das Produkt auch allgemein als ein „Nachschlagewerk für Wörter und ihren Gebrauch“ bezeichnet.<sup>5</sup>

Interessiert man sich nun beispielsweise für seinen Nachnamen, so wie ich hier für „Müller“, so erhält man den in Abbildung 1 gezeigten Eintrag.

Die Lemmatisierung von Eigennamen kann für allgemeine, einsprachige Wörterbücher als unüblich angesehen werden, so heißt es z.B. in der Einleitung zum Duden-Universalwörterbuch: „Personennamen“ wurden nur „aufgenommen, wenn sie als Appellativa (Gattungsbezeichnungen) [...] gebraucht werden, z.B. Casanova“ (DUW, S. 7). Diese Regel kann als gängige Praxis angesehen werden. Im Duden-Universalwörterbuch findet sich dementspre-

---

<sup>4</sup> Siehe <http://wortschatz.uni-leipzig.de/> (Stand: Februar 2007).

<sup>5</sup> Siehe <http://wortschatz.uni-leipzig.de/html/inhalt> (Stand: Februar 2007).

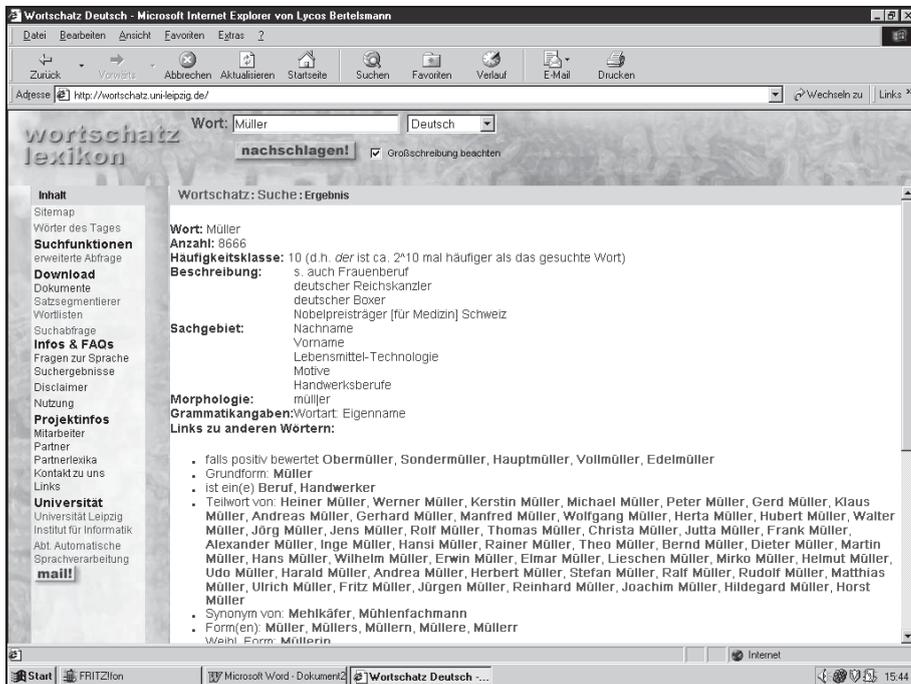


Abb. 1: Eintrag „Müller“ aus dem Wortschatz-Lexikon (Stand: 27. März 2003)

chend auch unter „Müller“ nur die Erläuterung der Berufsbezeichnung (DUW, S. 1041). Das Wortschatz-Lexikon wird von den eigenen Mitarbeitern, wie oben zitiert, jedoch als „Vollformenwörterbuch“ bezeichnet; wobei anscheinend Eigennamen zu den Vollformen gerechnet werden.

Wirklich überraschend wird es aber unter der Rubrik „Links zu anderen Wörtern“. Beispielsweise steht da zu „Müller“: „falls positiv bewertet: Obermüller, Sondermüller, [...] Edelmüller“. Verfolgt man diese Angabe und klickt z.B. auf „Edelmüller“, so erhält man den in Abbildung 2 gezeigten Eintrag.

Als Sachgebiet für „Edelmüller“ wird „Nachname“ angegeben; das einzige Beispiel, was aufgeführt wird, stammt aus dem Telefonbuch, welches anscheinend in das Korpus aufgenommen wurde. Daneben sind „MÜLLER“ und „müller“ als separate Einträge im Wortschatz-Lexikon zu finden, wobei „müller“ auf die Grundform „müll“ zurückgeführt wird und als „signifikante Kookkurrenzen“, „heiner“ und „traumtext“ aufgeführt sind.

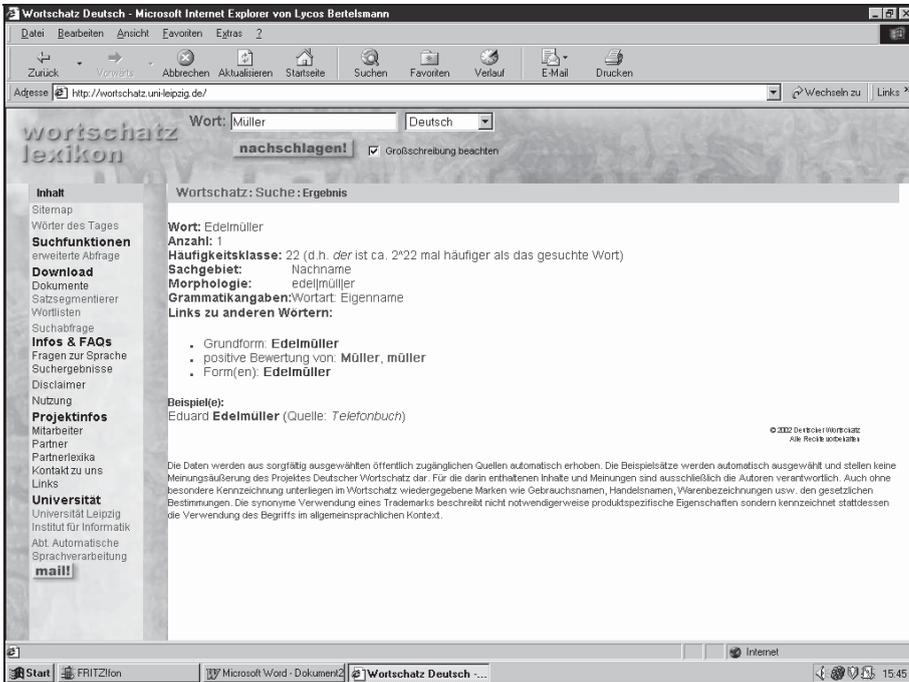


Abb. 2: Eintrag „Edelmüller“ aus dem Wortschatz-Lexikon (Stand: 27. März 2003)

Dieses einzeln herausgegriffene, bewusst strittige Beispiel soll in diesem Zusammenhang zeigen: Wollte man das Wortschatz-Lexikon rezensieren und würde ganz allgemein die Bewertungsmaßstäbe, die sich für gedruckte Wörterbücher etabliert haben, übertragen,<sup>6</sup> dann müsste diese Kritik verheerend ausfallen. Denn die Vernetzung, die oben gezeigt wurde, ist unter lexikografischen Gesichtspunkten nicht nachvollziehbar. Aber ist es überhaupt legitim, diese Anforderungen an das Wortschatz-Lexikon zu stellen? Dafür ist der Passus unter den jeweiligen Einträgen zu beachten. Dort heißt es: „Die Daten werden aus sorgfältig ausgewählten öffentlich zugänglichen Quellen automatisch erhoben [...]“. Es gilt also bei den Einträgen nicht – wie sonst üblich – zu fragen: Wie kommen die Lexikografen dazu, diese Angabe in der Form anzusetzen? Sondern vielmehr müssen die Fragen in folgende Richtung gehen: Mit welchen Analysemethoden werden die Textkorpora analysiert? Wie werden die Angaben im Detail automatisch erstellt? Wie ist

<sup>6</sup> In diesem Sinne z.B. Engelberg/Lemnitzer (2001, S. 194): „Auch für elektronische Wörterbücher gelten deshalb zunächst die Bewertungskriterien und -maßstäbe, die wir weiter oben für Printwörterbücher aufgezählt und kommentiert haben.“

das zu Grunde gelegte Textkorpus zusammengesetzt? Ist diese Zusammensetzung nachvollziehbar? Welche Ziele verfolgen die Projektverantwortlichen mit ihrem Produkt? etc. Die Fragen stimmen also zum großen Teil vom Ansatzpunkt nicht mit denen überein, die man üblicherweise bei einer Wörterbuchrezension stellt (vgl. z.B. Engelberg/Lemnitzer 2001, S. 161ff.).

Leider finden sich zu der Gewinnung der einzelnen Angaben auf der Webseite relativ wenig Hinweise, jedoch schreiben Wolff und Quasthoff:

Die Angaben zu Flexion, Morphologie, Sachgebiet, Synonymie etc., die vor allem aus strukturierten Quellen wie Lexika und durch computerlinguistische Analyseprogramme ermittelt werden, sind der jeweiligen Grundform eines Wortes zugeordnet [...]. Momentan haben diese Angaben noch nicht überall die gewünschte Qualität, da sie teilweise mit automatischen Mitteln erzeugt wurden und bei der Anzahl der Einträge eine vollständige intellektuelle Überarbeitung nicht möglich ist. (Quasthoff/Wolff 1999, S. 2)

Aus diesem Grund sollte man natürlich auch vorsichtig mit der Behauptung sein, das Wortschatz-Lexikon erfülle die Funktionen eines üblichen Wörterbuchs, denn zu denen gehört – zumindest im Rahmen der wissenschaftlichen Lexikografie (vgl. Wiegand 1998a, S. 40ff.) – auch die Verantwortung für die Zuverlässigkeit der lexikografischen Daten (vgl. Wiegand 1997, S. 195). Doch allein vom Produkt her kann man den Anspruch an eine solche Zuverlässigkeit der Daten nicht stellen, da deutlich gekennzeichnet ist, dass die Daten rein automatisch erhoben sind und nicht intellektuell nachbearbeitet wurden. Hier scheinen also übliche Maßstäbe – wie sie sich für die (Print-)Lexikografie etabliert haben – nicht angemessen. Deshalb ist es notwendig, terminologische Unterscheidungsmöglichkeiten zu entwickeln, mit denen man deutlich machen kann, dass es sich beim Wortschatz-Lexikon um eine andere Art von elektronischem Nachschlagewerk handelt als beispielsweise bei einer elektronischen Version des „Großen Wörterbuchs zur deutschen Sprache“ aus dem Dudenverlag und dementsprechend auch andere Bewertungsmaßstäbe angelegt werden müssen. Solche begrifflichen Unterscheidungen bieten dabei selbst noch keine Bewertungsmöglichkeiten, sondern sollen mögliche Ordnungsvorstellungen liefern, nach denen dann Bewertungskriterien entwickelt oder schon entwickelte gruppiert werden können. Im Folgenden wird es daher um erweitertes ‘terminologisches Handwerkszeug’ für die Wörterbuchkritik gehen.

Bevor die Form der Erarbeitung der lexikografischen Daten in eine mögliche Unterscheidung elektronischer Wörterbücher eingearbeitet wird, sollen nun zunächst kurz einige grundlegende Typologisierungsmöglichkeiten elektronischer Wörterbücher referiert werden.

### 3.2.2 Unterscheidungsmöglichkeiten elektronischer Wörterbücher: Bereits vorgelegte Typologien

Die Bewertung und Einordnung elektronischer Wörterbücher ist für die Wörterbuchforschung ein neuer wichtiger Bereich. Den aktuellsten Vorschlag einer Typologie mit umfangreicher Zurkenntnisnahme der internationalen (meta-)lexikografischen Literatur bietet Schryver in seinem bereits einmal zitierten Aufsatz „Lexicographer's Dreams in the Electronic-Dictionary Age“ (Schryver 2003). Schryver konstatiert darin, dass die bisherigen Typologien für die heutige Variabilität elektronischer Wörterbücher nicht hinreichend seien. Er schlägt daher eine eigene dreistufige Typologie vor, die nach einer dreigliedrigen Grundfrage entwickelt ist:

We would therefore like to suggest a typology based on one main, rigid criterion: the way in which dictionaries are *accessed*. More particularly, in designing this typology, we had one question in mind: ‘Who accesses what where?’ The resulting three-step typology is thought to be flexible enough to cater for future innovations. (Schryver 2003, S. 147)

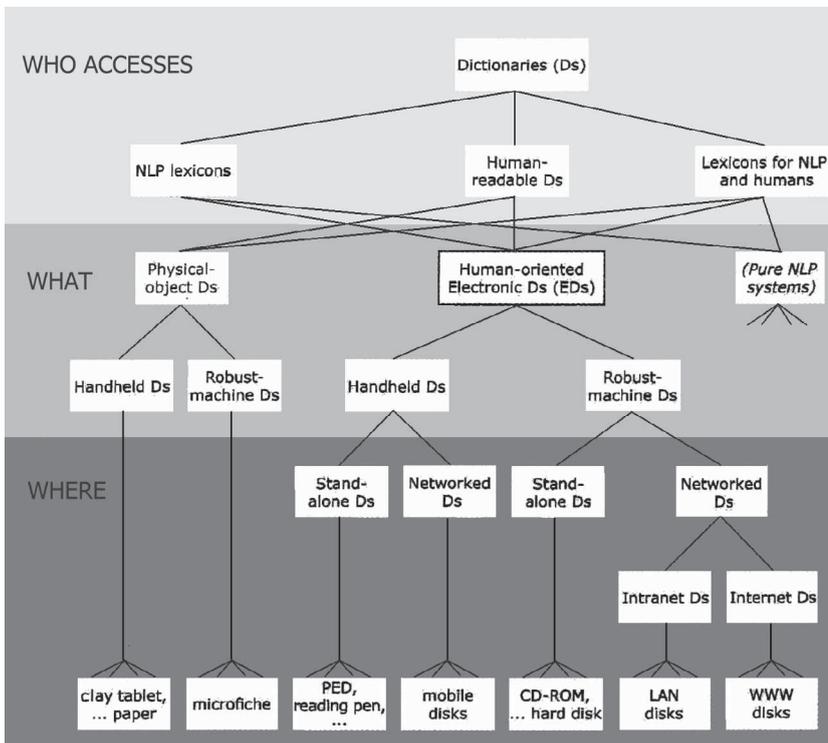


Abb. 3: Dreistufige Typologie zur Einordnung elektronischer Wörterbücher nach Schryver (Schryver 2003, S. 150)

In einem ersten Schritt wird demnach gefragt, wer auf das Wörterbuch zugeht. Hier bestehen nach dieser Typologie zwei grundsätzliche Möglichkeiten: Ein Mensch oder eine Maschine greift auf die Daten zu.<sup>7</sup> Der zweite Teil der Frage ist, worauf zugegriffen wird – „What is accessed?“ (Schryver 2003, S. 149) – d.h. wie das physikalische Objekt beschaffen ist. Im dritten Schritt wird dann geklärt, wie die Daten gespeichert sind, d.h.: „Where does one access the dictionary data?“ (ebd., S. 149). Diese Frage zielt damit auf den „type of storage“ (ebd.). Ein elektronisches Wörterbuch kann mit dieser Typologie nach vielen Aspekten klassifiziert und eingeordnet werden.

Zusätzlich zu dieser Typologie kann nach Schryver eine von Lehr vorgeschlagene metalexikografische Bewertung vorgenommen werden: „In addition, for each of the EDs one could add what Lehr (1996) termed a ‘metalexikographic evaluation’.“ (Schryver 2003, S. 150). Dies soll kurz erläutert werden: Zur Eröffnung der neuen Lexicographica-Rubrik „Electronic Dictionaries“ wurde von Andrea Lehr die sinnvolle und für Rezensionen schon mehrfach angewandte Unterscheidung zwischen papierorientierten vs. innovativ gestalteten elektronischen Wörterbüchern getroffen:

In (meta-)lexikographischer Hinsicht müssen wir zwischen elektronischen Wörterbüchern, die auf ein Papierwörterbuch zurückgehen und solchen, die Neuentwicklungen sind, unterscheiden. Erstere lassen sich außerdem danach subklassifizieren, ob sie eine wesentliche Veränderung bezüglich der Erscheinungsform ihrer Wörterbuchartikel erfahren haben oder nicht, und letztere danach, ob bei der Gestaltung der Wörterbuchartikel an traditionelle lexikographische Formen angeknüpft oder ob ein neuer Weg beschritten wurde – wir sprechen beide Male von *papierorientierten* vs. *innovativen* elektronischen Wörterbüchern. (Lehr 1996, S. 314)

Diese Art der Bewertung kann daher als eine weitere Folie über die Typologie Schryvers gelegt werden.

Genauso sollen die im Folgenden vorgestellten terminologischen Klärungen als eine Ergänzung zu der eben ausgeführten Typologie gesehen werden, da hier ein Aspekt aufgegriffen wird – nämlich die Art der Erarbeitung der lexikografischen Daten, der bisher noch nicht in Unterscheidungsmöglichkeiten

---

<sup>7</sup> Schryver versteht damit unter einem „electronic dictionary“ sowohl ein Produkt für Menschen als auch für Maschinen, auch wenn er an einer Stelle sagt, dass „the term ‘ED’ will therefore stand for ‘human-oriented electronic dictionary’“ (Schryver 2003, S. 146). Dies unterscheidet sich von der hier vorgeschlagenen Auffassung, dass unter einem elektronischen Wörterbuch nur ein Produkt für Menschen verstanden werden soll (vgl. Abschnitt 3.2.3).

elektronischer Wörterbücher eingearbeitet ist. Um eine solche Unterscheidung in verschiedene Arten der Erarbeitung elektronischer (Sprach-) Nachschlagewerke vornehmen zu können, soll zunächst geklärt werden, was in diesem Untersuchungsrahmen unter einem elektronischen Wörterbuch verstanden werden soll und was nicht.

### 3.2.3 Was ist ein elektronisches Wörterbuch?

Die Bezeichnung *elektronisches Wörterbuch* ist – allgemeinsprachlich formuliert – genau genommen nicht passend für den Bezugsgegenstand, da ein elektronischer Datenträger kein *Buch* ist. Die Übertragung der Bezeichnung *Wörterbuch* auf die Publikation lexikografischer Daten auf einem elektronischen Datenträger kann aber dann hilfreich sein, wenn damit auch bestimmte Eigenschaften des gedruckten Wörterbuchs auf das elektronische Wörterbuch übertragen werden können, d.h., wenn damit zur kommunikativ adäquaten Verwendung beigetragen wird. Dies ist dann gegeben, wenn die übertragenen Eigenschaften dem Bezugsgegenstand entsprechen.<sup>8</sup>

Um genauer zu prüfen, inwiefern und auch für welche Art von elektronisch publizierten lexikografischen Daten die Bezeichnung *elektronisches Wörterbuch* sinnvoll ist, ist zunächst zu fragen, wie ein gedrucktes Sprachwörterbuch charakterisiert werden kann. Wiegand<sup>9</sup> definiert zunächst ein Nachschlagewerk:

Def. 1: „Ein *Nachschlagewerk* ist ein *Buch* (hier verstanden als etwas Gedrucktes) mit wenigstens einer definierten äußeren Zugriffsstruktur, dessen genuiner Zweck darin besteht, daß ein potentieller Benutzer aus den lexikographischen Textdaten Informationen zum Gegenstandsbereich des Nachschlagewerkes gewinnen kann.“ (Wiegand 1998a, S. 58)

---

<sup>8</sup> Vgl. zu Metaphern im Bereich einer neuen Technologie Freisler (1994, S. 33, Anm. 37); hier in Bezug auf Sprachmetaphern im Bereich Computer: „Weingarten weist mit Recht darauf hin, daß die enorme Vielfalt der Dialogmetaphern, die in bezug auf den Computer existieren, wohl am besten als Deutungsversuche betrachtet werden können, um eine neue Technologie dem alltäglichen und wissenschaftlichen Verstehenshorizont zugänglich zu machen. Es ist bekannt, daß solche Eingliederungsversuche bzw. die Subsumtion von Unbekanntem unter Vertrautes in einer ersten Phase meist mit Hilfe von metaphorischen Übertragungen vorgenommen werden.“ Insofern könnte auch die Benennung elektronischer Sprachnachschlagewerke als elektronische Wörterbücher eine Übergangserscheinung sein.

<sup>9</sup> Die folgenden Überlegungen gehen im Wesentlichen von Wiegands *Wörterbuchforschung* (Wiegand 1998a) aus, da die Ausführungen dort meines Wissens die ausführlichsten und genauesten Überlegungen zu der Frage darstellen, was ein Sprachwörterbuch ist.

Darauf aufbauend definiert Wiegand ein Sprachwörterbuch:

Def. 2: „Ein *Sprachwörterbuch* ist ein *Nachschlagewerk*, dessen genuiner Zweck darin besteht, daß ein potentieller Benutzer aus den lexikographischen Textdaten Informationen zu sprachlichen Gegenständen gewinnen kann.“ (Wiegand 1998a, S. 58)

Folgt man dieser Definition, ist der Terminus *elektronisches Wörterbuch* demnach nur dann für eine kommunikativ adäquate Verwendung im fachsprachlichen Kontext hilfreich, wenn die Definitionen Wiegands für gedruckte Nachschlagewerke und Sprachwörterbücher folgendermaßen auf die Definitionen für elektronische Nachschlagewerke und Wörterbücher übertragen werden können:

Def. 3: Ein *elektronisches Nachschlagewerk* ist ein *elektronisch verfügbares Werk* mit wenigstens einer definierten äußeren Zugriffsstruktur, dessen genuiner Zweck darin besteht, dass ein potenzieller Benutzer aus den lexikografischen Textdaten Informationen zum Gegenstandsbereich des elektronischen Nachschlagewerkes gewinnen kann.

Def. 4: Ein *elektronisches Sprachwörterbuch* ist ein *elektronisches Nachschlagewerk*, dessen genuiner Zweck darin besteht, dass ein potenzieller Benutzer aus den lexikografischen Textdaten Informationen zu sprachlichen Gegenständen gewinnen kann.

Nun bezeichnet aber Wiegand z.B. auch Komponenten eines Sprachübersetzungssystems, welche als eine Art Wörterbuch fungieren, als Wörterbuch, genauer als Maschinenwörterbuch (Wiegand 1998a, S. 241), bei dem es entweder keine Benutzer gibt oder der Computer sozusagen der Benutzer ist. Ist es sinnvoll, auch in diesem Fall von einem Wörterbuch zu sprechen? Der Benutzer in den Definitionen 1-4 ist ein Handelnder, der „Informationen gewinnen“ kann; er muss daher ein Akteur mit kognitiven Fähigkeiten sein. Der Computer ist jedoch kein Akteur mit kognitiven Fähigkeiten, sondern eine Maschine, die vom Menschen dazu programmiert werden kann, Daten zu verarbeiten; ein Computer kann aus Daten keine Informationen gewinnen.<sup>10</sup> Davon ausgehend ist es meines Erachtens sinn-

<sup>10</sup> Vgl. Wiegand (1998a, S. 171): „Menschen verarbeiten Daten zu Informationen und Informationen, über welche sie bereits verfügen, zu anderen Informationen; sie *erarbeiten* sich Informationen (und damit Kenntnisse von etwas) in Wahrnehmungs- und Denkprozessen, um sich in der Welt orientieren und insbesondere, um zielorientiert handeln zu können.“

voll, elektronische Wörterbücher nur dann als Wörterbücher zu bezeichnen, wenn sie für einen menschlichen Benutzer gemacht sind. Nur dann können grundsätzliche Eigenschaften eines gedruckten Wörterbuchs sinnvoll auf ein elektronisches Wörterbuch übertragen werden. Ist der Computer als Benutzer gedacht, sollte besser von „lexikalischen Ressourcen für sprachtechnologische Produkte“ (Engelberg/Lemnitzer 2001, S. 230)<sup>11</sup> gesprochen werden, z.B.: ein automatisches Übersetzungsprogramm basierend auf einer lexikalischen Ressource. Die Definitionen 1-4 sollten demnach dahingehend ergänzt werden, dass mit „potenziellem Benutzer“ immer ein Mensch gemeint ist. Dass damit die Verwendung des Terminus *elektronisches Wörterbuch* an die Verwendung des Terminus *Wörterbuch* in der printlexikografisch orientierten Wörterbuchforschung angelehnt werden soll, ist dabei nicht primär durch den angestrebten Anschluss an diese Forschungen motiviert. Vielmehr scheint mir diese Einengung deshalb angemessen, weil nur so grundsätzliche Eigenschaften und Kriterien, die Wörterbücher ausmachen und die über die Präsentation in verschiedenen Medien Bestand haben, sinnvoll übertragen werden können.

Nun wurde der Terminus (*elektronisches*) *Wörterbuch* in seiner Verwendung dahingehend eingegrenzt, dass als Benutzer immer ein Mensch vorausgesetzt wird; d.h., dass es ein Produkt *für Menschen* ist. Das Beispiel aus dem Wortschatz-Lexikon hatte jedoch mehr zum Gegenstand, inwieweit die Daten *von Menschen* erarbeitet wurden. Letzteres soll in Abschnitt 3.2.5 in die Unterscheidung von automatisch erstellten vs. lexikografisch, d.h. menschlich bearbeiteten elektronischen Wörterbüchern aufgenommen werden. Zunächst wird jedoch ein alternativer Benennungsvorschlag für ein elektronisches Wörterbuch eingeführt.

---

Wahrnehmungs- und Denkprozesse gelten daher als *informationsverarbeitende*, intraindividuelle Prozesse. Computer werden in Handlungszusammenhängen als handlungsunterstützendes Mittel zur Erreichung von Handlungszielen von Menschen derart eingesetzt, daß sie in *datenverarbeitenden* extraindividuellen Prozessen Daten zu anderen Daten *verarbeiten*. Da alle computerinternen Daten als Mengen von numerischen Zeichenexemplaren aufgefaßt werden können, sind datenverarbeitende niemals informationsverarbeitende Prozesse.“

<sup>11</sup> Engelberg/Lemnitzer sprechen in diesem Zusammenhang neben der genannten Redeweise auch von „lexikalischen Komponenten sprachverarbeitender Systeme“ (Engelberg/Lemnitzer 2001, S. 204).