

## TEI-basierte Modellierung von Retrodigitalisaten (am Beispiel des Trierer Wörterbuchnetzes)

Vera Hildenbrandt [hildenbr@uni-trier.de](mailto:hildenbr@uni-trier.de), Tel.: +49 651 201-3790

### 1. Einführung

Mit den Möglichkeiten des World Wide Web und den technologischen Entwicklungen seit Mitte der neunziger Jahre des 20. Jahrhunderts haben Stichwörter wie Performanz, Zugriffsmöglichkeiten, Interoperabilität, Hypertextualisierung, Langzeitarchivierung oder Nachhaltigkeit längst Einzug in die lexikographisch-lexikologische Diskussion gehalten und den lexikographischen Prozess, aber auch die Ansprüche von Nutzern an Wörterbücher verändert. Es stellen sich u.a. folgende Fragen: Wie können Wörterbücher im Internet präsentiert werden? Welche Zugriffsmöglichkeiten auf ihre Inhalte können geboten werden? Wie können Wörterbuchdaten sinnvoll miteinander vernetzt werden? Wie können sie nachhaltig vorgehalten und langfristig archiviert werden? Wie müssen Wörterbücher aufbereitet werden, um dauerhaft flexible vernetzte Präsentations- und Recherchemöglichkeiten zu schaffen?

Diese Fragen stellen den Bearbeiter von Wörterbüchern, die nicht von Anfang an digital publiziert, sondern erst durch Retrodigitalisierung ins elektronische Medium überführt werden, vor besondere Herausforderungen. Wie kann der Spagat zwischen Web-Technologien und Wörterbüchern, die ursprünglich nicht für die Internetpublikation konzipiert wurden, gelingen? Wie können retrodigitalisierte Wörterbücher so modelliert werden, dass die darin gebotenen lexikographischen Informationen auf verschiedene Weise präsentierbar und auf innovative Weise recherchierbar werden?

Fragen wie diese führten Ende 1987 unter der Ägide der Association for Computers and the Humanities, der Association for Computational Linguistics und der Association for Literary and Linguistic Computing zur Gründung der Text Encoding Initiative (TEI). Als Initiative der Wissenschaft für die Wissenschaft verfolgt die TEI seitdem das Ziel, hard- und software-unabhängige Methoden für die Codierung, den Austausch und die langfristige Archivierung geisteswissenschaftlicher Daten zu entwickeln. Im Jahr 1994 veröffentlichte die TEI die erste offizielle Version ihrer *Guidelines for the Encoding and Interchange of Machine-Readable Texts* (TEI P3),<sup>9</sup> die, basierend auf der Metasprache SGML, Codierungsregeln und -empfehlungen für eine Reihe verschiedener Textsorten und eine Vielzahl von Anwendungen vorschlug. Diese Regeln und Empfehlungen wurden seither ständig überarbeitet und erweitert; im November 2007 wurde die zurzeit aktuelle, XML-basierte und die Vorteile von XML-Schema nutzende Version P5 der Guidelines herausgegeben.<sup>10</sup> Die TEI-Guidelines beschreiben somit nicht nur ein in den Geisteswissenschaften mittlerweile sehr gängiges Dokumentenformat für Datencodierung und -austausch, sondern sie sind auch konform mit aktuellen Standards. Auf diese Weise können beispielsweise andere XML-Sprachen (z.B. MathML, SVG) ebenso in TEI-Dokumenten verwendet werden, wie TEI in anderen XML-Sprachen (z.B. MEI) genutzt werden kann.

<sup>9</sup> Die 1990 erschienene erste Version der Richtlinien (TEI P1) und die 1992 veröffentlichte Version P2 hatten nach dem Selbstverständnis der Verfasser keinen offiziellen, sondern „Entwurfscharakter“ (vgl. Schmidt 1997, S. 351).

<sup>10</sup> Vgl. zur Geschichte der Text Encoding Initiative die von der TEI selbst gebotene *History* unter <http://www.tei-c.org/About/history.xml>, zu den Anfängen außerdem Schmidt (1997), S. 351 f.

Die TEI-Guidelines wurden zur Modellierung der im Trierer Wörterbuchnetz versammelten retrospektiv digitalisierten Wörterbücher herangezogen. Im Folgenden soll eine allgemeine kurze Einführung in das TEI-Modul für Wörterbücher verdeutlichen, aus welchen Gründen das Codierungsschema der TEI in Trier eingesetzt wurde, bevor anschließend anhand einiger konkreter Beispiele die Vorzüge und Schwierigkeiten, die bei der Anwendung der TEI-Richtlinien auf retrodigitalisierte lexikographische Ressourcen auftreten, dargestellt werden.

## 2. Das TEI-Modul für Wörterbücher

An den Guidelines der TEI arbeiten seit der Gründung der Initiative Geistes- und Kulturwissenschaftler verschiedener Fachdisziplinen. Rückmeldungen der Anwender und die Arbeit verschiedener Special-Interest-Groups tragen zur ständigen praxisnahen Weiterentwicklung eines vielseitigen Regelwerks bei, das Standards setzt, den Nutzern aber dennoch Freiräume lässt.<sup>11</sup> Das Codierungsschema der TEI besteht aus verschiedenen Modulen, die jeweils eine bestimmte Anzahl von XML-Elementen (Tags) deklarieren, die gegebenenfalls durch Attribute spezifiziert werden.<sup>12</sup> Neben einem sogenannten „Core-Tagset“, das in jedem TEI-Dokument zur Verfügung steht, gibt es einige „Base-Tagsets“, so unter anderem ein Tagset zur Auszeichnung von Prosatexten, eines zur Auszeichnung von Verstexten, eines zur Auszeichnung von Manuskripten und eben auch eines zur Auszeichnung von Wörterbüchern.

Die Entwicklung der Richtlinien für die Codierung von Wörterbüchern stellte die zuständige Arbeitsgruppe vor besondere Herausforderungen, denn es galt, ein Schema zu entwickeln, das so allgemein ist, dass es für eine Vielzahl von Wörterbüchern anwendbar ist und zugleich die Besonderheiten des einzelnen Wörterbuchs abdeckt.<sup>13</sup> Das Schema muss berücksichtigen, dass Wörterbücher – auch typographisch – zu den komplexesten Textsorten gehören, die von der TEI behandelt werden, dass sie sehr stark und tief strukturiert und die enthaltenen Informationen sehr verdichtet sein können und dass ebenso die Struktur von Artikeln verschiedener Wörterbücher wie auch die Struktur von Artikeln innerhalb eines Wörterbuchs variieren kann. Etymologische Informationen etwa können an unterschiedlichen Stellen und auf unterschiedlichen Hierarchieebenen innerhalb des Wörterbuchartikels erscheinen, oder Informationen auf einer niedrigeren Hierarchieebene können Informationen auf einer höheren Hierarchieebene ersetzen. Das heißt, dass sich dieselben Informationsklassen auf allen Hierarchieebenen eines Wörterbuchartikels finden können. Die TEI-Richtlinien für Wörterbücher müssen also zulassen, dass alle XML-Elemente auf allen Ebenen eines Wörterbuchartikels vorkommen können und dass im Umkehrschluss alle Elemente, die eine Artikelebene kennzeichnen, denselben Inhalt haben dürfen.<sup>14</sup>

Aus der Bewältigung dieser Herausforderungen durch die Entwickler des Wörterbuchmoduls der TEI, das heißt aus der Umsetzung der Erkenntnisse über die mögliche(n) Struktur(en) von Wörterbüchern und ihren Artikeln in ein Regelwerk ergibt sich ein Teil der Gründe, die für den Einsatz der TEI-Guidelines bei der Modellierung von Retrodigitalisaten sprechen: Erstens

---

<sup>11</sup> Vgl. dazu Jannidis (1997), S. 153/154.

<sup>12</sup> Auf eine Beschreibung des Aufbaus einer TEI-Datei sei an dieser Stelle mit Verweis auf Jannidis (1997), S. 155-158, verzichtet. Obwohl bereits 1997 erschienen, darf die allgemeine, knappe Einführung in die Teile einer TEI-Datei auch heute noch Gültigkeit beanspruchen und ist gerade für Neuanwender sehr gut verständlich.

<sup>13</sup> Entwickelt wurde das Wörterbuchtagset zunächst für Wörterbücher mittlerer Größe und westlicher Sprachen wie den *Petit Larousse*, den *Petit Robert* und das *Collins English Dictionary* (vgl. Ide/Véronis 1996b, S. 170).

<sup>14</sup> Zu den Herausforderungen, die bei der Entwicklung von Richtlinien für die Auszeichnung von Wörterbuchartikeln zu bewältigen waren, vgl. Ide/Véronis (1996a) und Ide/Véronis (1996b), S. 173-175.

sind die Richtlinien ganz bewusst so allgemein gehalten, dass sie auf unterschiedlich konzipierte Wörterbücher angewendet werden können. Zweitens ermöglichen sie benutzerdefinierte Erweiterungen. Drittens erlauben sie eine nur geringe Explizitheit der Auszeichnung, das heißt, nicht jeder Wörterbuchartikel muss bis in seine feinsten Verästelungen hinein codiert sein, bevor eine elektronische Publikation vorgenommen werden kann. Viertens sind durch das Befolgen der mittlerweile international anerkannten und bei der Codierung einer Vielzahl von historischen Wörterbüchern eingesetzten Guidelines gute Voraussetzungen für die Interoperabilität der Wörterbuchdaten und damit für die Vernetzung mit anderen Nachschlagewerken geschaffen. Fünftens kann mittels der Richtlinien nicht nur die Oberflächenstruktur der Wörterbücher (Layout, Typographie, lineare Abfolge von Zeichen und Informationen) codiert werden, wodurch es beispielsweise möglich wird, in der Präsentation einer vor der digitalen Version existierenden gedruckten Fassung, wie sie bei Retrodigitalisaten immer gegeben ist, treu zu bleiben. Vielmehr kann das Wörterbuch durch die XML-basierte Codierung auch als Datenbank betrachtet werden, die der Nachschlagende nicht von A bis Z liest, sondern in der er gezielt sucht und selektiert.<sup>15</sup> Die TEI-Codierung von Wörterbüchern schafft folglich nicht nur die nötigen Voraussetzungen für eine flexible Präsentation der Wörterbuchinhalte, sondern bereitet auch verschiedene Zugriffsmöglichkeiten auf diese Inhalte vor. Und sechstens schließlich wurden die Guidelines bereits in vielen Projekten erfolgreich erprobt und nicht zuletzt bei der Codierung (retrodigitalisierter) älterer Wörterbücher eingesetzt.<sup>16</sup>

Das TEI-Modul „Dictionaries“ erlaubt die Abbildung der Makro- und Mikrostruktur von lexikographischen Ressourcen verschiedener Art. Für die Makrostruktur von gedruckten Wörterbüchern sind analog zu anderen gedruckten Texten die Tags <text>, <front>, <body>, <back> und <div> vorgesehen. Der Wörterbuchartikel selbst wird in der Regel durch das Element <entry> bezeichnet; Artikel, die in höchstem Maße unstrukturiert sind und die mit dem <entry>-Tag verbundenen Bedingungen nicht erfüllen, dürfen in Ausnahmefällen durch den <entryFree>-Tag gekennzeichnet werden. Neben diesen Elementen sind weitere die Hierarchie eines Wörterbuchartikels abbildende Tags der Homographentag <hom>, der zur Codierung aller Informationen, die sich auf einen Homographen innerhalb eines Wörterbuchartikels beziehen, herangezogen wird, und das Element <sense>, das zur Markierung aller auf eine bestimmte Bedeutung des Stichwortes bezogenen Angaben eingesetzt wird. In diese größeren hierarchisierenden Struktureinheiten können dann weitere lexikographische Informationseinheiten eingeschachtelt sein. Hier sehen die TEI-Guidelines u.a. Tags vor zur Auszeichnung von Informationen über die geschriebene und gesprochene Form eines Stichwortes (<form>), von morphosyntaktischen Informationen (Wortart, Genus, Numerus, Kasus etc.) (<gramGrp>), von Bedeutungsangaben (<def>), von Zitaten inklusive bibliographischer Angaben (<cit>), von Angaben zum Gebrauch des Stichwortes (<usg>), von Querverweisen innerhalb des Wörterbuchs oder auf andere Werke (<xr>), von etymologischen Informationen (<etym>), von Verweisen auf Komposita oder Derivate des Stichwortes (<re>) und schließlich zur Auszeichnung von Anmerkungen (<note>). Die verschiedenen durch diese Elemente jeweils gekennzeichneten Informationen, die die TEI als Hauptkonstituenten eines Wörterbuchartikels betrachtet, können durch weitere Tags feiner ausgezeichnet werden. So kann etwa bei den Informationen zu einem Stichwort weiter differenziert werden zwischen der orthographischen Form des Lemmas (<orth>), Angaben zur Aussprache (<pron>), zur Silbentrennung (<syll>) usw. Zusätzlich können weitere, in allen TEI-Dokumenten verfügbare Elemente zur Kennzeichnung von typographischen und Layout-Merkmalen – zum Beispiel des Zeilen-

<sup>15</sup> Vgl. Ide/Véronis (1996b), S. 174.

<sup>16</sup> Vgl. etwa den Bericht von Susan Rennie über die Codierung des *Scottish National Dictionary* (Rennie 2000) sowie die Liste der Projekte, die das Codierungsschema der TEI nutzen, unter <http://www.tei-c.org/Activities/Projects/>.

umbruchs (<lb>), des Seitenumbruchs (<pb>) oder des Spaltenwechsels (<cb>) – genutzt werden, um den Bezug zur Printfassung zu erhalten. Doch sollen die allgemeinen Ausführungen über die Richtlinien der TEI für die Codierung von Wörterbüchern an dieser Stelle nicht weiter vertieft werden,<sup>17</sup> vielmehr soll im Folgenden die Anwendung der Guidelines bei der Auszeichnung retrodigitalisierter Wörterbücher am Beispiel des Trierer Wörterbuchnetzes im Zentrum stehen.

### 3. Codierung retrodigitalisierter Wörterbücher nach den Richtlinien der TEI

Im Trierer Wörterbuchnetz ist zwischenzeitlich eine Reihe verschiedener Nachschlagewerke versammelt, von denen die Mehrzahl erst lange nach Erscheinen der gedruckten Version ins digitale Medium überführt wurde. Mit Förderung durch die Deutsche Forschungsgemeinschaft (DFG) wurden am Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier (Trier Center for Digital Humanities) seit 1997 das *Mittelhochdeutsche Wörterbuch* von Georg Friedrich Benecke, Wilhelm Müller und Friedrich Zarncke (BMZ, 1854-1866), das *Mittelhochdeutsche Handwörterbuch* von Matthias Lexer (Lexer, 1872-1878), Lexers Nachträge zum *Handwörterbuch*, das *Findebuch zum mittelhochdeutschen Wortschatz* (Findebuch, erarbeitet zwischen 1986 und 1992), die Erstbearbeitung des *Deutschen Wörterbuchs* von Jacob und Wilhelm Grimm (<sup>1</sup>DWb, 1852-1960), das *Pfälzische Wörterbuch* (PfälzWb, 1965-1997), das *Rheinische Wörterbuch* (RheinWb, 1928 und 1971), das *Wörterbuch der deutsch-lothringischen Mundarten* (LothrWb, 1909) und das *Wörterbuch der elsässischen Mundarten* (ElsWb, 1899-1907) digitalisiert. Ebenfalls mit Förderung durch die DFG wurden parallel zur Ausarbeitung der weiteren Bände des Wörterbuchs die zwischen 1966 und 1998 veröffentlichten ersten drei Bände des *Goethe-Wörterbuchs* (GWb) digitalisiert und im Internet publiziert. Zu diesen Ressourcen kommen mit *Meyers Großem Konversationslexikon* und Adelungs *Grammatisch-Kritischem Wörterbuch der Hochdeutschen Mundart* weitere Nachschlagewerke, deren Daten im Rahmen des vom Bundesministerium für Bildung und Forschung geförderten Verbundprojektes Text-Grid erworben werden konnten, sowie eine Reihe externer Ressourcen (*Mittelhochdeutsches Wörterbuch*, *Mittelhochdeutsche Begriffsdatenbank*, *Deutsches Rechtswörterbuch*, *Oekonomische Encyclopädie* von Krünitz, *Lexikon der Luxemburger Umgangssprache*, *Wörterbuch der luxemburgischen Mundart*, *Luxemburger Wörterbuch*), auf die aus dem Trierer Wörterbuchnetz und vice versa verlinkt wird. Im Rahmen dieses Beitrags stehen vor allem die Wörterbücher im Fokus, deren digitale Version im Rahmen von DFG-geförderten Projekten entstanden ist – also BMZ, Lexer, Findebuch, <sup>1</sup>DWb, PfälzWb, RheinWb, LothrWb, ElsWb und GWb.

Gemeinsam ist diesen Wörterbüchern, dass sie – es wurde oben bereits erwähnt und wird zudem durch die Erscheinungsdaten deutlich – ausschließlich im Druck veröffentlicht und mit Ausnahme des GWb bereits abgeschlossen wurden, als noch nicht an eine elektronische Version und deren Möglichkeiten gedacht wurde. Teilweise sind diese Wörterbücher über einen sehr langen Zeitraum hinweg entstanden; als prominentestes Beispiel muss hier sicherlich das <sup>1</sup>DWb angeführt werden, dessen erste Lieferung 1852 und dessen letzte Lieferung im Jahr 1960 erschien. Der in diesen Zeiträumen erfolgte Wandel sprachwissenschaftlicher und lexikographischer Grundsätze, aber auch den Lexikographen auferlegte Straffungskonzepte, die einen rascheren Abschluss der Unternehmen bewirken sollten, führten im Verlauf der Erarbeitung der Nachschlagewerke zu Änderungen in der lexikographischen Praxis. Auch hier ragt

---

<sup>17</sup> Für weiterführende Informationen sei auf Kapitel 9 „Dictionaries“ der TEI-Guidelines verwiesen. Eine PDF-Datei ist zugänglich unter: <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.

das <sup>1</sup>DWb aus der Riege der genannten Wörterbücher hervor, denn bereits „[v]on Beginn an, also seit Jacob und Wilhelm Grimm, gibt es ungewollte und gewollte Differenzen im Aufbau und in der Aussageleistung der Artikel.“ (Schmidt 2004, S. 25) Auch eine erste Reorganisation der Wörterbucharbeit im Jahr 1908 sowie eine zweite im Jahr 1930 änderten nur wenig an der Heterogenität in der Artikelstruktur.<sup>18</sup> Aber auch ein jüngeres Wörterbuch wie das GWb wurde Straffungskonzepten unterworfen, die zu Modifikationen in der lexikographischen Praxis führten.<sup>19</sup> Gemeinsam ist den hier genauer betrachteten Nachschlagewerken also auch, dass es sich um in mehr oder minder geringem Maße standardisierte Wörterbücher handelt, deren Artikel sich – etwa im Fall von BMZ und <sup>1</sup>DWb – durch einen eher diskursiven Stil und zum Teil durch eine wenig stringente Artikelstruktur auszeichnen. Durch die Diskursivität in der Darbietung der Informationen wird es schwierig bis unmöglich, die Zugehörigkeit der Informationen zu einer bestimmten Informationsklasse programmgestützt zu bestimmen. Zudem werden Informationen nicht immer in derselben Reihenfolge angeboten. Im BMZ beispielsweise können Angaben zur Morphologie, zur Bedeutung und zur Etymologie in beliebiger Abfolge und an beliebigen Stellen eines Wörterbuchartikels vorkommen, unter Umständen können sie auch mehrmals innerhalb desselben Artikels auftreten.<sup>20</sup> Und gemeinsam sind den Wörterbüchern schließlich auch gewisse Inkonsistenzen – so zum Beispiel bei den zitierten Siglen oder den Einleitungen eines Querverweises.

Diese Wörterbücher mit ihrer jeweils spezifischen Geschichte, ihren Gemeinsamkeiten und ihren Problemen wurden retrodigitalisiert mit dem Ziel, sie als Open-Access-Publikationen im Internet anzubieten. Dabei sollten sie zum einen als elektronische Abbilder der gedruckten Versionen ins Internet gestellt werden, um auch aus der digitalen Publikation heraus die Printausgaben referenzier- und zitierbar zu halten. Das heißt, dass bei der Aufbereitung und Auszeichnung der Daten weder der Wörterbuchtext an sich noch die Struktur der Artikel verändert werden durfte. Zum anderen sollten die Wörterbücher jedoch als Datenbank begriffen werden, in der bestimmte lexikographische Informationseinheiten gezielt such- und durchsuchbar sein sollten. Darüber hinaus sollten zumindest die bereits im Druck eng aufeinander bezogenen Wörterbücher, also zum einen BMZ, Lexer und Findebuch und zum anderen PfälzWb, RheinWb, LothrWb und ElsWb, in der digitalen Version miteinander vernetzt werden. Zur Umsetzung dieser Ziele standen jeweils nur ein begrenzter Zeitraum und ein begrenztes Budget zur Verfügung. Dies bedeutete, dass nach der Retrodigitalisierung der Wörterbücher, die im Double-Keying-Verfahren erfolgte, der Prozess der Auszeichnung auf möglichst ökonomische Weise, das heißt weitgehend automatisiert durchgeführt werden musste. Es wurden daher Programmroutinen entwickelt, die – auf den bei der Eingabe miterfassten typographischen Merkmalen aufsetzend – das TEI-Markup nach der Top-down-Methode in die Wörterbuchdaten einbrachten. Das heißt, zuerst wurden die oberen Ebenen der Artikelstruktur ausgezeichnet und von hier aus dann stufenweise die weiteren lexikographischen Informationseinheiten in Angriff genommen. Dabei wurde die Tatsache, dass die meisten der genannten Wörterbücher wenig standardisiert sind, zu einer Herausforderung für die computergestützte Auszeichnung.

---

<sup>18</sup> Eine kurzgefasste Geschichte des <sup>1</sup>DWb bieten Christmann/Hildenbrandt/Schares (2001), S. 14-20.

<sup>19</sup> So wurde etwa ab Band 3 des GWb die in den Verweisreihen der ersten beiden Bände gemachte Unterscheidung zwischen Synonymen und im weiteren Sinne bedeutungsverwandten Wörtern und Ausdrücken aufgegeben, und ab Lieferung IV,11 wurden die Artikelstruktur gestrafft sowie die Behandlung hochfrequenter Wörter (z.B. Präpositionen, Konjunktionen, Modalverben) radikal auf die Darstellung Goethescher Besonderheiten reduziert (vgl. Schmidt/Reinitzer/Kühlmann 2004).

<sup>20</sup> Vgl. zum Problem der geringen Standardisierung der BMZ-Artikel Burch/Fournier (2001), S. 140 ff.

Wie bei den Ausführungen über das Modul „Dictionaries“ der TEI-Guidelines dargestellt, wurde bei der Entwicklung eines Codierungsstandards für elektronische Wörterbücher zwar grundsätzlich berücksichtigt, dass Wörterbuchartikel in ihrer Struktur variieren können – so dürfen etwa die Elemente <form>, <gramGrp>, <sense> und <etym> ohne zwingend vorgeschriebene Reihenfolge wiederholt innerhalb eines <entry>-Tags vorkommen. Die Tatsache aber, dass bestimmte typographische Eigenschaften unterschiedliche Arten von Textelementen abbilden, das Fehlen eindeutiger struktureller Marker, Variantenreichtum in Bezug auf eine bestimmte Information und vor allem im Wörterbuch nur implizit gegebene Informationen, die sich von einem menschlichen Benutzer ohne Probleme erschließen lassen, stellen – einige Beispiele werden dies in der Folge verdeutlichen – eine große Hürde für ein maschinelles Markup dar. Für die in Trier digitalisierten Wörterbücher bedeutete dies, dass mittels computergestützter Verfahren innerhalb des jeweils zur Verfügung stehenden Bearbeitungszeitraums je nach Umfang und Struktur des Wörterbuchs unterschiedliche Auszeichnungstiefen erreicht werden konnten. Das <sup>1</sup>DWb mit seinen 16 Bänden in 32 Teilbänden, das auf nahezu 70.000 Spalten rund 320.000 Stichwörter in heterogen strukturierten und vielfach sehr diskursiven Wörterbuchartikeln behandelt, steht hier am einen Ende der Skala, das GWb, das in voraussichtlich neun Bänden rund 90.000 Stichwörter in relativ konsistent strukturierten Artikeln beschreiben wird, am anderen Ende. Anhand dieser beiden Wörterbücher seien daher im Folgenden die Möglichkeiten und Grenzen der computergestützten TEI-basierten Modellierung von Retrodigitalisaten demonstriert.

**Buch** **I** *das zu einer Werkeinheit gebundene od geheftete, geschriebene od gedruckte Literaturwerk* **1** *konkret, bes im Hinblick auf Format u Bindeart, Typographie u Ausstattung* **a** *als handschriftl Buch* Das schön geschriebne B. 6,38 Vs 4 DivHafis B25,184,8 Lorsbach 31.1.15 B21,14,21 Christiane 28.7.09 uö **b** *als gedrucktes Buch; mehrf '(ein)gebundenes, ungebundenes, geheftetes, gefalztes, rohes B.', vereinzelt 'durchschossenes B.'* Wenn typographisch allgemach die Bücher sich steigern, darf wohl auch der Buchbinder ehrenvoll als Künstler hervortreten 49<sup>2</sup>,135,2 KLehmanns Buchbinderarb Zu der andern Ausgabe [von HermDor] bin ich ganz wohl mit der hierbey zurückkommenden lateinischen Schrift zufrieden, nur wünsche ich einen breiten Steg und überhaupt viel Rand, als die wahre Zierde jedes B-es B12,135,11 Böttiger 3.6.97 Der Druck [von KUA] nimmt sich wie bisher mit den wenigen Abänderungen sehr gut aus; alles kömmt auf's Papier an, woran denn freylich unsre meisten Hefte und Bücher kranken B41,248,12 FJFrommann 9.12.26 K die Kupferstecher tractiren alles was zu einem B-e gehört so leicht und lose B12,368,13 Schiller 2.12.97 A(Lerche 70) ChAVulpius 11.6.00 [G/Voigt BiblKomm] B9,38,25 CarlAug 8.10.88 30,87,25 ItR B10,335,16 Schiller 21.11.95 B8,277,16 Gösehen [27.10.87] uö **c** *in sekundärer Verwendung als Aufbewahrungsort für empfindl Papiere, insbes zum Pressen u Bewahren von Pflanzen u Insekten* B25,338,13 Meyer 17.5.15 N6,336,9 Bryophyllum calycinum N6,419,3 MetamInns uö **2** *inhaltl*

Abb. 1: Der Anfang des Artikels *Buch* im GWb (Band 2, Spalte 923 ff.)

Der Beginn des Artikels *Buch* in der Printversion des GWb (vgl. Abb. 1) verdeutlicht, wie sehr sich die inhaltliche Strukturierung in der Typographie dieses Wörterbuchs widerspiegelt: Das Lemma ist halbfett gesetzt, die lexikographischen Interpretamente sind kursiv und halbfett gedruckt, die Belegzitate werden durch eine andere Schrifttype gekennzeichnet, die Belegstellenangaben durch einen kleineren Schriftgrad und eine geringere Zeichenbreite. Dank dieser sehr stark differenzierten typographischen Gestaltung und dem trotz der bereits mehr als 60 Jahre währenden Bearbeitungszeit sehr regelhaften Aufbau der Wörterbuchartikel konnte die TEI-konforme Codierung des GWb weitgehend automatisiert und vergleichsweise zügig durchgeführt werden, und es konnte eine beträchtliche Auszeichnungstiefe erreicht werden.

Dies wird illustriert durch die Präsentation des Wörterbuchs im Internet (vgl. Abb. 2). Die typographischen Merkmale der Druckfassung sind weitgehend erhalten, allerdings mit dem Unterschied, dass verschiedene Informationseinheiten in verschiedenen Farben markiert sind. Lexikographische Interpretamente sind rot, Belegzitate grün und Belegstellenangaben blau gehalten. Eine Präsentationsform, die Gloning und Schlaps bei ihren Vorüberlegungen zu möglichen *Prototypen für ein elektronisches Goethe-Wörterbuch* bedachten, ist damit umgesetzt: Es liegt „derselbe Datenbestand vor wie im gedruckten Wörterbuch, die Ansicht ist jedoch deutlich stärker strukturiert.“ (Gloning/Schlaps 1999, S. 26).

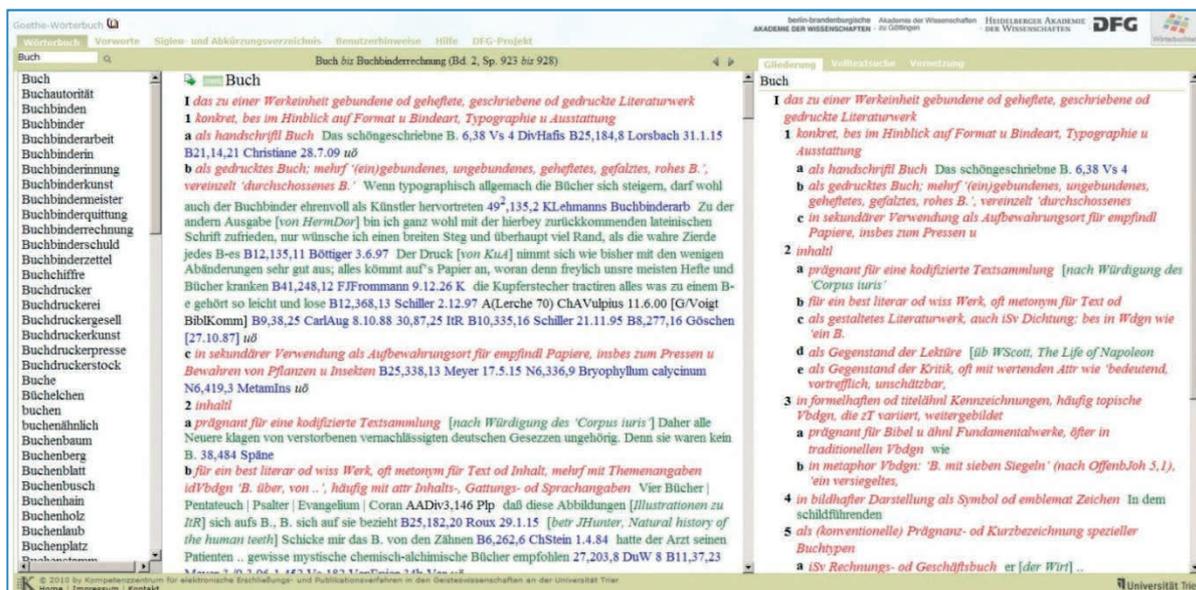


Abb. 2: Graphische Oberfläche des GWb im Internet

Zusätzlich werden im rechten Anzeigebereich die Wörterbuchartikel auf ihr hierarchisches Bedeutungsgerüst, also Lemma, Nummerierung und lexikographische Interpretamente, reduziert, im linken Anzeigebereich werden lediglich die Lemmata angeführt. Auf diese Weise vermag sich der Nutzer einen rascheren Überblick über die Artikelinhalte zu verschaffen. Neben diesen verschiedenen Präsentationsmöglichkeiten wurden aber auch verschiedene Recherchemöglichkeiten geschaffen (vgl. Abb. 3). Der Benutzer kann im Volltext suchen, er kann die Suche aber auch auf die Stichwörter, die Bedeutungserklärungen, die Belegzitate und die Belegstellenangaben beschränken, wobei die Kombination der verschiedenen Suchfelder ebenso erlaubt ist wie der Einsatz von Wildcards.



Abb. 3: Graphische Oberfläche des GWb im Internet mit Suchmaske

Diese verschiedenen Präsentations- und Recherchemöglichkeiten setzen voraus, „dass die entsprechenden Informationen [...] in expliziter Weise im Datenbestand enthalten sind.“ (Gloning/Welter 2001, S. 118). Gloning und Welter bringen diese Notwendigkeit auf den Punkt in dem knappen Satz: „What you mark is what you get“ (Gloning/Welter 2001, S. 128). Im GWb sind folgende Informationseinheiten markiert und damit gezielt „greifbar“: der Artikelkopf mit Lemma und Vorbemerkung, der Artikelkörper mit den verschiedenen Bedeutungsblöcken, Gliederungsmarken und lexikographischen Interpretamenten, Belegzitate und Belegstellenangaben und der Artikelanhang mit dem Verweisblock, der auf Synonyme, Derivate und Komposita des Stichworts verweist, aber auch Anmerkungen enthalten kann (vgl. Abb. 4).

```

<entry xml:id="JB04816" n="20923.48">
  <form type="artkopf">
    <ref type="DWB" target="B2S0466Z79"></ref>
    <form type="lemma">Buch</form>
  </form>
  <sense rend="artkoerper">
    <sense n="1" level="2" value="20923.48">
      <def rend="leitbem">das zu einer Werkeinheit gebundene od
      geheftete, geschriebene od gedruckte Literaturwerk</def>
      <sense n="1" level="3" value="20923.49">
        <def rend="leitbem">konkret, bes im Hinblick auf Format u
        Bindeart, Typographie u Ausstattung</def>
        <sense n="a" level="4" value="20923.50">
          <def rend="leitbem">als handschriftl Buch</def>
          <cit>
            <quote>Das sch&ouml;ngeschriebne
            B.</quote>
            <bibl>6,38 Vs 4 DivHafis</bibl>
          </cit>
        </sense>
      </sense>
    </sense>
    <sense rend="artanhang">
      <sense rend="verweisblock">
        <xr type="derivblock">&derivsign;
        <ref type="deriv" target="B1S0017Z54">ABC-Buch</ref>
      </sense>
      <xr type="anmblock">
        <note type="footnote" n="1">1) <note><hi rend="italic">Briefw
        von Riemer,
      </note>
      <ref type="artikelautor" rend="H. U." n="Horst Umbach"></ref>
    </sense>
  </entry>

```

Abb. 4: Auszug aus dem XML-codierten Artikel *Buch* des GWb

Eine so feinkörnige Modellierung der Daten konnte für das <sup>1</sup>DWb nicht erreicht werden, was unter anderem damit zusammenhängt, dass das Ausgangsmaterial, die gedruckte Fassung, weniger konsistent strukturiert und typographisch weit weniger differenziert ist als die des GWb. Im <sup>1</sup>DWb wechseln sich im Wesentlichen recte und kursiv gesetzte Passagen ab (vgl. Abb. 5). Lediglich die Lemmata, die Autornamen und die Verszitate heben sich durch Versalien, Kapitalälchen bzw. Einrückung vom umgebenden Wörterbuchtext ab.

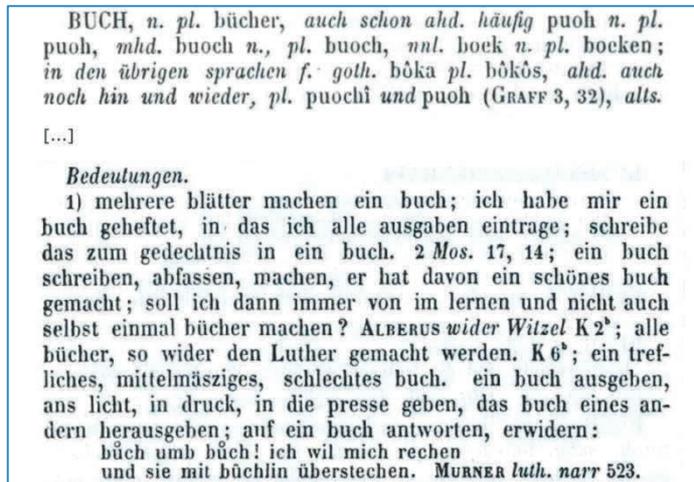


Abb. 5: Auszug aus dem Artikel *Buch* im <sup>1</sup>DWb (Band 2, Spalte 466 ff.)

Maschinell lassen sich hier daher auch nur die Lemmata, die den Lemmata folgenden Wortartangaben, die Autornamen, die Verszitate und die Gliederungsmarken markieren (vgl. Abb. 6).

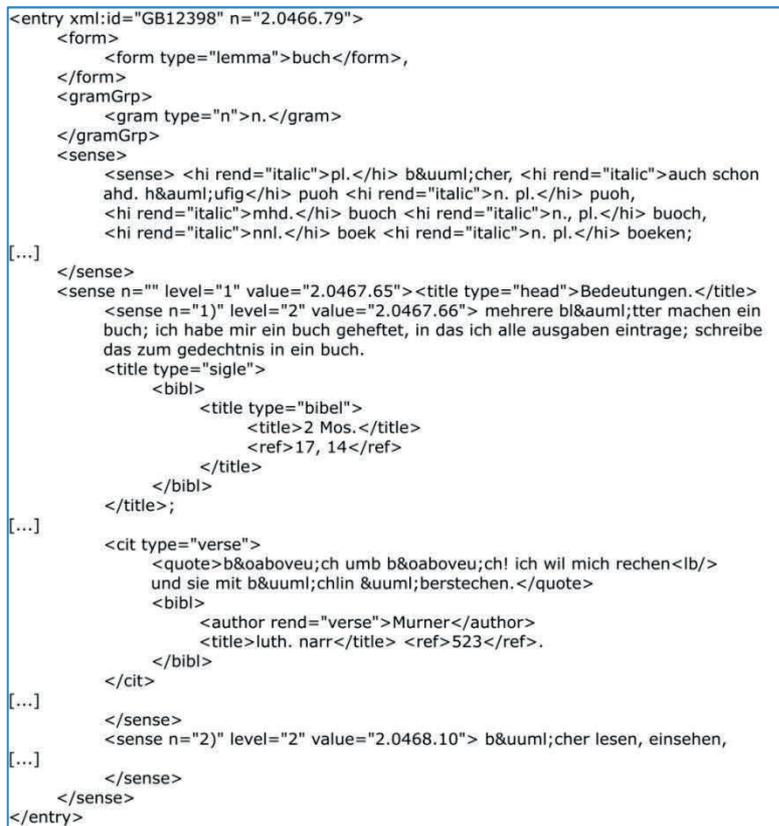


Abb. 6: Auszug aus dem XML-codierten Artikel *Buch* des <sup>1</sup>DWb



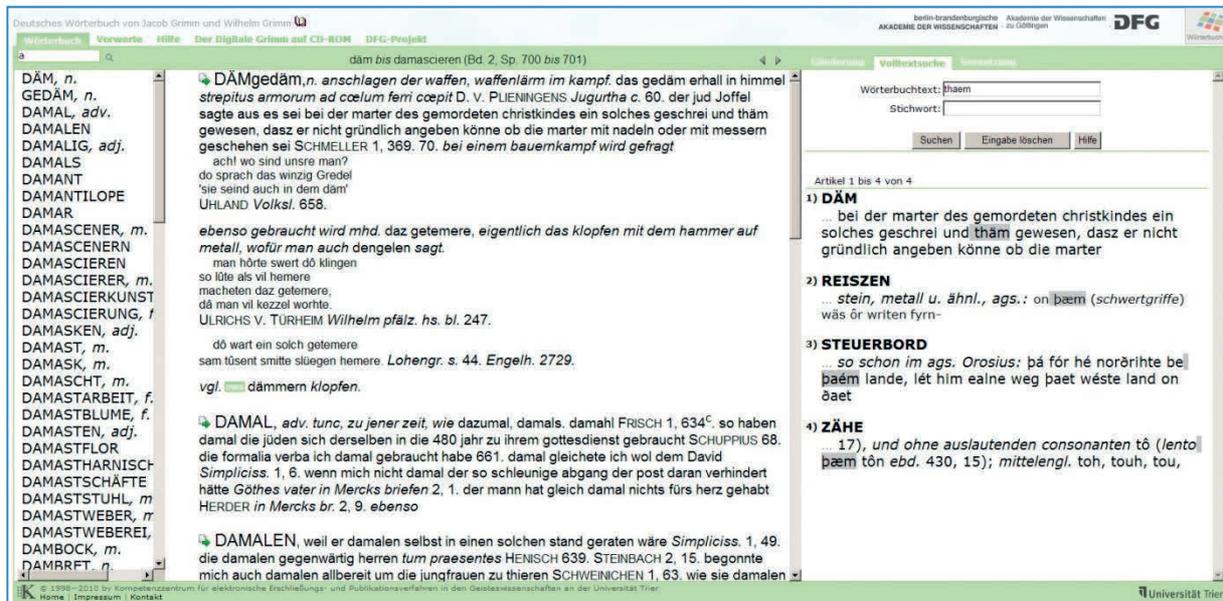


Abb. 8: Graphische Oberfläche des <sup>1</sup>DWb im Internet mit Suchmaske

Die Möglichkeit, die Artikelansicht auf das Gliederungsgerüst zu reduzieren, konnte im digitalen <sup>1</sup>DWb nur mit sehr hohem Aufwand realisiert werden, da das Gliederungsschema im <sup>1</sup>DWb im Vergleich beispielweise zu dem des GWb sehr viel weniger konsistent ist. Zwar können die Gliederungsmarken im <sup>1</sup>DWb automatisch auffindig gemacht werden, da sie in der Regel etwas eingerückt am Beginn einer neuen Zeile stehen, doch ist es sehr schwierig, ihre Position in der Artikelhierarchie zu bestimmen. Die „1)“ kann die erste Gliederungsebene einleiten, ebenso gut aber auch die zweite oder dritte Ebene bezeichnen, je nachdem, ob der Artikel mit „I. A. 1)“ oder „1) a)“ beginnt. Auch die Abfolge der Gliederungsmarken variiert. Im Artikel *herz* beispielsweise wird die oberste Gliederungsebene durch Großbuchstaben eingeleitet, auf der nächstniedrigeren Ebene gefolgt von römischen Ziffern („A. I. 1)“); im Artikel *gott* dagegen kennzeichnen römische Ziffern die oberste, Großbuchstaben die zweite und arabische Ziffern die dritte Gliederungsebene („I. A. 1)“). Je differenzierter und tiefer die Artikel untergliedert sind, desto mehr häufen sich die Besonderheiten.<sup>22</sup> Das genaue Erfassen der Position einer Gliederungsmarke in der Gliederungshierarchie stellte daher eine große Herausforderung bei der Programmierung der Markup-Routinen dar.

Ähnlich problematisch gestaltete sich die Auszeichnung der Quellensiglen in einigen der retrodigitalisierten Wörterbücher. Im BMZ beispielsweise wird Lorenz Diefenbachs *Vergleichendes Wörterbuch der gotischen Sprache* in mindestens zwölf Varianten angeführt. Es begegnet als „Diefenb. g. wb.“, „Diefenb. g. w.“, „Diefenb. g. wtrb.“, „Diefenb. g. wrtbch.“, „Diefenb. g. wtrbch.“, „Diefenb. g. wörterb.“, „Diefenb. goth. w.“, „Diefenb. goth. wb.“, „Diefenb. goth. wtrbch.“, „Diefenb. goth. wörterb.“, „Diefenbach g. wb.“, „Diefenbach goth. wtrbch.“ und „Diefenbach goth. wörterb.“ Versehentlich fehlende Abkürzungspunkte erhöhen die Varianz, die sich auch durch Zuhilfenahme des Quellenverzeichnisses nicht auflösen lässt, da dessen Zusammensteller Nellmann nur die erste der angeführten Siglen aufgenommen hat.<sup>23</sup> Ähnlich ist der die Siglen betreffende Befund im <sup>1</sup>DWb. Abraham a Santa Clara etwa wird als „ABRAHAM A S. CL.“, „ABRAHAM VON S. CLARA“, „ABRAH. A SANTA CLARA“, „ABR. A ST. CLARA“, „ABR. A. S. CL.“, „ABR. A. S. C.“ usw. angeführt. Das Quellenverzeichnis führt von den mehr als zwanzig Varianten lediglich eine auf. Um dem Nutzer des digitalen

<sup>22</sup> Zur Problematik der Gliederungsmarken im <sup>1</sup>DWb vgl. Bartz/Burch/Christmann/Gärtner/Hildenbrandt/Schares/Wegge (2004), S. 81 f.

<sup>23</sup> Vgl. dazu Burch/Fournier (2001), S. 143.

<sup>1</sup>DWb die erfolgreiche Suche nach „Abraham a Santa Clara“ zu ermöglichen, mussten alle Varianten ausfindig gemacht und auf die normalisierte Form abgebildet werden.

Als sehr aufwendig erwies sich bei einigen Wörterbüchern auch die Auszeichnung der Binnenverweise. Sie als Hyperlinks zu realisieren, gehört eigentlich zu den Mindestanforderungen einer Internetversion, im PfälzWb und im <sup>1</sup>DWb beispielsweise erwies sich die dafür notwendige Datenaufbereitung allerdings keineswegs als trivial. Denn wo sich das GWb mit einem Verweispfil, einem „vgl.“ oder einem „Syn.“ zur Einleitung eines Verweises auf ein anderes Stichwort begnügt, wurden für das PfälzWb an die hundert verschiedene Verweiseinleitungen von „vgl.“ und „s.“ über „dafür“ und „hierfür“ bis hin zu „im Unterschied zu“ und „im Ggs. zu“ ermittelt. Das <sup>1</sup>DWb überbietet diesen Befund noch, indem es neben dem gängigen „s.“ oder „vgl.“ in allen Variationen („vgl. oben“, „vergleiche“, „vergl. auch“, „s.“, „siehe“, „s. d. und vgl. unten“) zahllose diskursive Wendungen wie beispielsweise „früher hiesz es auch“, „man sehe die ausführung unter“ oder „wofür man doch lieber sagt“ als Verweiseinleitungen nutzt. Ein Ermitteln aller Verweiseinleitungen erscheint hier ebenso unmöglich wie die klare Abgrenzung der Verweislemmata, die häufig nur unzureichend durch typographische oder anderweitige Strukturmarker von den umgebenden Informationen geschieden sind. Vollständigkeit ist in der Auszeichnung der Binnenverweise daher bei einem Wörterbuch mit dem Umfang und der Anlage des <sup>1</sup>DWb selbst durch manuelle Nacharbeit nicht zu erreichen.

Auch eine Vernetzung verschiedener lexikographischer Ressourcen auf Lemmaebene stellt eine Herausforderung dar in den Fällen, in denen die Lemmata der einzelnen Wörterbücher nach verschiedenen Grundsätzen angesetzt sind oder sehr viele homographe Lemmata vorkommen. Solche Probleme wurden virulent bei der Verlinkung des GWb mit dem <sup>1</sup>DWb. Das Phänomen einer abweichenden Typographie der Lemmata in beiden Wörterbüchern konnte teilweise durch eine „Normalisierung“ der Stichwörter abgefangen werden, das heißt, Umlaute wurden auf die Grundbuchstaben reduziert, Akzente eliminiert usw. Dadurch konnten bei einem automatischen Vergleich der Lemmalisten von GWb und <sup>1</sup>DWb eine Reihe von Übereinstimmungen und somit Verweisziele ermittelt und es konnte zum Beispiel das Stichwort *abbüßen* im GWb mit dem Stichwort *abbüsen* im <sup>1</sup>DWb verlinkt werden. Bei homographen Stichwörtern dagegen konnte die Zuordnung nicht automatisch erfolgen. Hier musste ein verstehender Leser die mittels automatisierter Verfahren eingetragenen Verweise in einem manuellen Nachbearbeitungsschritt kontrollieren und gegebenenfalls korrigieren, um beispielsweise zu verhindern, dass aus dem GWb-Artikel *Buch* in den <sup>1</sup>DWb-Artikel *buch* im Sinne von „buk, *backte*“ (<sup>1</sup>DWb, Band 2, Spalte 466) verlinkt wurde.

Nahezu komplett versagen muss der Versuch eines maschinellen Auszeichnungsverfahrens bei nur implizit gegebenen Informationen. Sich häufig auf Belegstellenangaben beziehende Rückverweise wie „ebend.“, „ebd.“, „das.“, „ebendas.“ lassen sich von „einem ‚kontextsensitiven‘ menschlichen Bearbeiter“ (Burch/Fournier 2001, S. 144) erschließen. Dieses kontextsensitive Erschließen des Verweisziels in eine Auszeichnungsroutine zu fassen, ist dagegen ein schwieriges, zum Teil sicherlich unlösbares Unterfangen. Unter Punkt 10) des Artikels *gericht* beispielsweise führt das <sup>1</sup>DWb einen mehrfachen Rückverweis auf: „gericht *am wagen*, *lateralia* SCHMELLER a. a. o.; *ingricht*, *vorrichtung am wagen*. *ebd.* aus dem *bairischen wald*.“ Herauszufinden, auf welches Werk Schmellers sich „*ebd.*“ bezieht, ist selbst für den menschlichen Benutzer nicht ganz einfach, denn der nächste Hinweis auf Schmeller findet sich in der gedruckten Fassung gut vierzig Zeilen über dem angeführten Zitat und dort leider ebenfalls mit der Angabe „a. a. o.“. Dies wiederholt sich zehn Zeilen zuvor, so dass im Artikel *gericht* insgesamt zwar viermal auf Schmeller verwiesen wird, aber jedes Mal ohne Nennung des Werktitels. Ein kundiger Wörterbuchnutzer vermag zu erschließen, dass es sich

hierbei um Schmellers *Bayerisches Wörterbuch* (Band 2/1 [1877], Spalte 35 ff., Artikel *richten*) handeln muss, einem maschinellen Markup sind hier jedoch definitiv Grenzen gesetzt.

#### 4. Resümee

Insgesamt hat sich die Entscheidung, bei der Auszeichnung der retrodigitalisierten Wörterbücher auf die TEI-Guidelines zu setzen, bewährt. Sicherlich eignen sie sich nicht zuletzt deshalb für die Codierung ursprünglich nur gedruckt erschienener lexikographischer Ressourcen, weil diese die Ausgangsbasis für die Entwicklung des Wörterbuchmoduls waren. Auch dass die Guidelines bewusst allgemein gehalten sind – ein Umstand, der andere, digital entstehende Vorhaben von einer TEI-basierten Modellierung Abstand nehmen lässt<sup>24</sup> –, erweist sich im Falle der retrodigitalisierten Wörterbücher als Vorzug. Probleme bei der TEI-konformen Auszeichnung dieser Wörterbücher liegen weniger im Codierungsschema begründet als vielmehr in der Anlage der Wörterbücher. Diskursivität und geringe Standardisierung von Wörterbuchartikeln, das Fehlen eindeutiger struktureller Marker für bestimmte lexikographische Informationseinheiten und Varianz sind Phänomene, die sich mittels computergestützter Verfahren nicht abfangen lassen. Nicht die Anforderungen der potentiellen Nutzer an ein elektronisches Wörterbuch und die Funktionalitäten des Computers entscheiden in diesen Fällen über die Art und Tiefe der Textauszeichnung, sondern vor allem die Beschaffenheit des gedruckten Ausgangsmaterials und die für dessen Codierung zur Verfügung stehenden personellen und finanziellen Ressourcen. Das „Wunschziel“ einer „linguistisch motivierte[n] und feinkörnige[n] Modellierung“ würde – darin ist Storrer voll und ganz zuzustimmen –, „wenn sie auf der Grundlage eines gedruckten Wörterbuchs erfolgt, einen relativ hohen Auf- und Nachbearbeitungsaufwand“ (Storrer 2001, S. 64) erfordern. Für ein Wörterbuch vom Umfang des <sup>1</sup>DWb beispielsweise ist jedoch eine solche Auf- und Nachbearbeitung gänzlich undenkbar und unfinanzierbar. Dennoch sollte auf eine Retrodigitalisierung und TEI-konforme Auszeichnung nicht aus dem Grund verzichtet werden, dass der Datenmodellierung Grenzen gesetzt sind. Denn auch wenn unter Umständen keine hohe Auszeichnungstiefe erreicht werden kann, bieten sie doch andere Zugriffs- und Auswertungsmöglichkeiten als ihre gedruckten Pendant,<sup>25</sup> Zugriffs- und Auswertungsmöglichkeiten, die – das belegen die Nutzerzahlen<sup>26</sup> – gerne genutzt werden.

#### 5. Literatur

##### 5.1 Wörterbücher im Trierer Wörterbuchnetz

Adelung, Johann Christoph (1793-1801): Grammatisch-Kritisches Wörterbuch der Hochdeutschen Mundart mit beständiger Vergleichung der übrigen Mundarten, besonders aber der oberdeutschen. Zweyte, vermehrte und verbesserte Ausgabe. 4 Bände. Leipzig.

BMZ = Benecke, Georg Friedrich/Müller, Wilhelm/Zarncke, Friedrich (1854-1866): Mittelhochdeutsches Wörterbuch. Leipzig. [Nachdruck mit einem Vorwort und einem zusammengefaßten Quellenverzeichnis von Eberhard Nellmann sowie einem alphabetischen Index von Erwin Koller, Werner Wegstein und Norbert Richard Wolf. Stuttgart 1990.]

<sup>1</sup>DWb = Grimm, Jacob und Wilhelm (1854-1960. 1971): Deutsches Wörterbuch. 16 Bde. in 32 Teilbänden. Nebst Quellenverzeichnis. Leipzig.

<sup>24</sup> Im Projekt *ellexiko* beispielsweise entschied man sich aus diesem Grund gegen die von der TEI vorgeschlagene Standard-Modellierung von Wörterbüchern. Vgl. Müller-Spitzer (2005), S. 28.

<sup>25</sup> S. dazu oben Anm. 13.

<sup>26</sup> Monatlich sind rund eine Million Zugriffe auf das Trierer Wörterbuchnetz zu verzeichnen.

- ElsWb = Martin, Ernst/Lienhart, Hans (1899-1907): Wörterbuch der elsässischen Mundarten. 2 Bände. Straßburg. [Ndr.: Berlin/New York 1974.]
- Findebuch = Gärtner, Kurt/Gerhardt, Christoph/Jaehrling, Jürgen/Röll, Walter/Timm, Erika/Hanrieder, Christoph (Datenverarbeitung) (1992): Findebuch zum mittelhochdeutschen Wortschatz. Mit einem rückläufigen Index. Stuttgart.
- GWb = Goethe-Wörterbuch (1966 ff.). Hg. v. der Berlin-Brandenburgischen Akademie der Wissenschaften, der Akademie der Wissenschaften zu Göttingen und der Heidelberger Akademie der Wissenschaften. Band 1 ff. Stuttgart.
- Lexer = Lexer, Matthias (1872-1878): Mittelhochdeutsches Handwörterbuch. Leipzig. [Nachdruck mit einer Einleitung von Kurt Gärtner. Leipzig 1992.]
- LothrWb = Follmann, Michael Ferdinand (1909): Wörterbuch der deutsch-lothringischen Mundarten. Leipzig 1909. [Ndr.: Hildesheim/New York 1971.]
- Meyers Großes Konversationslexikon. Ein Nachschlagewerk des allgemeinen Wissens. Sechste, gänzlich neu bearbeitete und vermehrte Auflage. Leipzig und Wien 1905-1909.
- PfälzWb = Pfälzisches Wörterbuch. Begründet von Ernst Christmann, fortgeführt von Julius Krämer, bearbeitet von Rudolf Post unter Mitarbeit von Josef Schwing. 6 Bände. Wiesbaden/Stuttgart 1965-1997.
- RheinWb = Rheinisches Wörterbuch. Im Auftrag der Preußischen Akademie der Wissenschaften, der Gesellschaft für Rheinische Geschichtskunde und des Provinzialverbandes der Rheinprovinz auf Grund der von Johannes Franck begonnenen, von allen Kreisen des Rheinischen Volkes unterstützten Sammlung bearbeitet und herausgegeben von †Josef Müller, †Heinrich Dittmaier, Rudolf Schützeichel und Matthias Zender. 9 Bände. Bonn/Berlin 1928-1971.

## 5.2 Forschungsliteratur

- Bartz, Hans-Werner/Burch, Thomas/Christmann, Ruth/Gärtner, Kurt/Hildenbrandt, Vera/Schares, Thomas/Wegge, Klaudia (2004): Wie das Deutsche Wörterbuch in den Computer kam. In: Der Digitale Grimm. Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Elektronische Ausgabe der Erstbearbeitung. Bearbeitet von Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner, Vera Hildenbrandt, Thomas Schares, Klaudia Wegge. Hg. vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. 2 CD-ROMs, Benutzerhandbuch, Begleitbuch. Frankfurt/Main, S. 73-90.
- Burch, Thomas/Fournier, Johannes (2001): Zur Anwendung der TEI-Richtlinien bei der Retrodigitalisierung mittelhochdeutscher Wörterbücher. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher (= Lexicographica: Series maior; 107). Tübingen, S. 133-153.
- Christmann, Ruth/Hildenbrandt, Vera/Schares, Thomas (2001): Ein „heiligthum der sprache“ digitalisiert: Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. In: Benito, Nicolás Castrillo/Stahl, Peter (Hg.): TUSTEP educa. Actas de Congreso del International Tustep User Group: Peñaranda de Duero (Burgos) Octubre 1999. Burgos, S. 13-35.
- Burnard, Lou/Bauman, Syd (Hg.) (2010): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Oxford, Providence, Charlottesville, Nancy. Internet: <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>. (Stand: November 2011).
- Gloning, Thomas/Schlaps, Christiane (1999): Prototypen für ein elektronisches Goethe-Wörterbuch. In: Sprache und Datenverarbeitung. International Journal for Language Data Processing 32, 2, S. 21-34.
- Gloning, Thomas/Welter, Rüdiger (2001): Wortschatzarchitektur und elektronische Wörterbücher: Goethes Wortschatz und das Goethe-Wörterbuch. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. (= Lexicographica: Series maior; 107). Tübingen, S. 117-132.
- Ide, Nancy/Véronis, Jean (1996a): Encodage des dictionnaires électroniques: problèmes et propositions de la TEI. In: Piotrowski, David (Hg.): Lexicographie et informatique. Autour de l'informatisation du Trésor de la Langue Française. Actes du Colloque International de Nancy. Paris, S. 239-261. Internet: <http://sites.univ-provence.fr/~veronis/pdf/1996Tlf.pdf>. (Stand: November 2011).
- Ide, Nancy/Véronis, Jean (1996b): Codage TEI des dictionnaires électroniques. In: Cahiers GUTenberg n° 24 (spécial TEI), S. 170-176.

- Jannidis, Fotis (1997): Wider das Altern elektronischer Texte: philologische Textauszeichnung mit TEI. In: editio. Internationales Jahrbuch für Editionswissenschaft 11, S. 152-177.
- Müller-Spitzer, Carolin (2005): Die Modellierung lexikographischer Daten und ihre Rolle im lexikographischen Prozess. In: Haß, Ulrike (Hg.): Grundfragen der elektronischen Lexikographie. *lexiko – das Online-Informationssystem zum deutschen Wortschatz*. (Schriften des Instituts für Deutsche Sprache; 12) Berlin/New York, S. 21-54.
- Rennie, Susan (2000): Encoding a Historical Dictionary with the TEI. (With reference to the Electronic Scottish National Dictionary Project). In: Heid, Ulrich/Evert, Stefan/Lehmann, Egbert/Rohrer, Christian (Hg.): Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart, S. 261-271. Internet: [http://www.euralex.org/elx\\_proceedings/Euralex2000/031\\_Susan%20RENNIE\\_Encoding%20a%20Historical%20Dictionary%20with%20the%20TEI%20\(With%20reference%20to%20the%20Electronic%20Scottish%20National%20Dictionary%20Project\).pdf](http://www.euralex.org/elx_proceedings/Euralex2000/031_Susan%20RENNIE_Encoding%20a%20Historical%20Dictionary%20with%20the%20TEI%20(With%20reference%20to%20the%20Electronic%20Scottish%20National%20Dictionary%20Project).pdf). (Stand: November 2011).
- Schmidt, Frieder (1997): Neuland für die Buchgeschichte – Quellenaufbereitung im Zeitalter des WWW. Hypertext Markup Language (HTML), Standard Generalized Markup Language (SGML) und die Guidelines for Electronic Text Encoding and Interchange der Text Encoding Initiative (TEI). In: Leipziger Jahrbuch zur Buchgeschichte 7, S. 343-365.
- Schmidt, Hartmut (2004): Das Deutsche Wörterbuch. Gebrauchsanleitung. In: Der Digitale Grimm. Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Elektronische Ausgabe der Erstbearbeitung. Bearbeitet von Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner, Vera Hildenbrandt, Thomas Schares, Klaudia Wegge. Hg. vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. 2 CD-ROMs, Benutzerhandbuch, Begleitbuch. Frankfurt/Main, S. 25-64.
- Schmidt, Hartmut/Reinitzer, Heimo/Kühlmann, Wilhelm (2004): Vorwort. In: Goethe-Wörterbuch. Vierter Band, Lieferung 12. Stuttgart.
- Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. (= *Lexicographica: Series maior*; 107). Tübingen, S. 53-69.

