

**Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner,
Vera Hildenbrandt, Thomas Schares, Klaudia Wegge**

Abschlußbericht

Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet

Gefördert durch die DFG im Rahmen des Programms
„Retrospektive Digitalisierung von Bibliotheksbeständen“
November 1998 bis Oktober 2003

Inhalt

Vorbemerkungen	3
1 Ein „Wörterbuch zum Hausbedarf“: Zur Digitalisierung des Deutschen Wörterbuchs	5
2 Datenerfassung und Fehlerkorrektur	8
3 Kodierung mit SGML nach den Richtlinien der TEI	16
3.1 Auszeichnung nach internationalen Standards: SGML/TEI	16
3.2 Sukzessive Datenkodierung	16
Auszeichnung mit Hilfe von TUSTEP	17
Datenkodierung im <i>topdown</i> -Verfahren	17
Kodierung der Artikelstruktur	18
Unterscheidung der Trenn- und Bindestriche	21
Auszeichnung der Stichwörter in Reihen- oder Nestartikeln	21
3.3 Auszeichnung des Quellenverzeichnisses	24
4 Die elektronische Publikation	25
4.1 Datenbanken	26
4.2 Publikation im Internet	29
4.3 Publikation auf CD-ROM	31
5 Publikationen und Öffentlichkeitsarbeit	34
5.1 Projektpräsentationen	34
5.2 Publikationen	36
5.3 Kolloquium in der Akademie der Wissenschaften und der Literatur Mainz am 15./16. Mai 2003	39
6 Der digitale Grimm auf der <i>Documenta11</i> in Kassel, 8. Juni bis 15. September 2002	40
7 Berichte in Presse, Internet, Rundfunk und Fernsehen	41
8 Perspektiven	45
Anhang: SGML-Dokumentation (Tabelle)	49

Vorbemerkungen

An der Universität Trier wurde von November 1998 bis Juli 2004 die Erstbearbeitung des *Deutschen Wörterbuchs* (DWB) der Brüder Grimm digitalisiert und für eine elektronische Publikation aufbereitet. Seit Juni 2002 ist das digitale DWB im Internet allgemein zugänglich, im Juli 2004 wurde es durch den Verlag Zweitausendeins (<http://www.zweitausendeins.de>) auf CD-ROM mit erweiterten Recherchefunktionen veröffentlicht. Die Projektziele wurden damit erreicht.

Für die Durchführung des Projekts war ein Zeitraum von insgesamt sechs Jahren vorgesehen, für eine Förderung durch die *Deutsche Forschungsgemeinschaft* (DFG) im Rahmen ihres Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ wurden fünf Jahre mit zwei zweijährigen und einer abschließenden einjährigen Förderphase veranschlagt. Der Erstantrag an die DFG wurde nach einer im Januar 1998 positiv beschiedenen Voranfrage am 7.4.1998 eingereicht und am 14.9.1998 bewilligt. Auch die beiden Fortsetzungsanträge wurden bewilligt, so daß die durch die DFG geförderten Arbeiten zügig und ohne Unterbrechung durchgeführt und die in den Förderanträgen aufgestellten Arbeits- und Zeitpläne weitestgehend eingehalten werden konnten.

Das Projekt wurde mit der Unterstützung des vom Land Rheinland-Pfalz geförderten „Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier“ durchgeführt. Die Zusammenarbeit mit diesem Zentrum, das auch das von der DFG geförderte lexikographische Vorgängerprojekt, den digitalen Verbund mittelhochdeutscher Wörterbücher (www.MWV.uni-trier.de) betreute, wurde in hohem Maße begünstigt durch die Kumulierung von vergleichbaren Projekten und die dadurch ermöglichten Synergieeffekte. Besonders bedeutsam war die enge Zusammenarbeit mit den Mitarbeitern des Vorgängerprojekts, die ihre Erfahrungen und Kompetenzen wirkungsvoll vermitteln konnten. Das Kompetenzzentrum wird die geplanten Updates des digitalen DWB mitentwickeln. Die langfristige Pflege der im Projekt erarbeiteten Daten wird durch die Abteilung „Elektronische Medien“ der Universitätsbibliothek Trier übernommen. Die Kommission für deutsche Philologie der Akademie der Wissenschaften und der Literatur Mainz wird auch bei der Weiterentwicklung des digitalen DWB beratend tätig sein.

Im folgenden Abschlußbericht für die DFG soll ausführlich über die Konzeption des Projekts, die Hauptarbeitsschritte, die Projektpublikationen und die Öffentlichkeitsarbeit, die Präsentation des digitalen DWB als Exponat auf der *Documenta11* und die weiteren Perspektiven, die sich durch die Schaffung und Weiterentwicklung des elektronischen DWB eröffnen, berichtet werden. Die im Anhang zu-

sammengestellte SGML-Dokumentation gibt einen Einblick in die Komplexität der Arbeiten und Aufgaben, die im Bereich des elektronischen Publizierens auf internationalem Niveau zu bewältigen sind. Der Bericht ist von den im Projektteam tätigen Wissenschaftlichen Mitarbeiterinnen und Mitarbeitern und dem Antragsteller gemeinsam verfaßt.

Zum erfolgreichen Abschluß des Projekts haben viele Personen und Institutionen beigetragen, denen für ihre Unterstützung zu danken ist. Ausführliche Danksagen an die das Projekt unterstützenden Personen und die Kooperationspartner finden sich in meiner Einleitung zum Begleitband der CD-ROM, S. 19-23 (vgl. 5.2 Publikationen). Ausdrücklich genannt werden sollen hier die Universität Trier, die Mainzer Akademie und besonders das Ministerium für Wissenschaft, Weiterbildung, Forschung und Kultur des Landes Rheinland-Pfalz, das durch erhebliche finanzielle Unterstützung im Rahmen seiner Fördermaßnahmen zur Einrichtung von Kompetenzzentren die Gründung des Trierer „Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften“ und damit die zur Durchführung des Projekts entscheidende interdisziplinäre Zusammenarbeit zwischen Informatikern und Geisteswissenschaftlern ermöglichte. Vor allem ist schließlich der Deutschen Forschungsgemeinschaft für eine effiziente Antragsberatung und die entscheidende und großzügige Förderung zu danken. Sie erlaubte es der Trierer Arbeitsgruppe, die Möglichkeiten der neuen Informations- und Kommunikationstechnologie konsequent und kreativ zu nutzen und so die konkurrenzlose Informationsfülle, die das *Deutsche Wörterbuch* zur deutschen Sprache bietet, in der digitalen Version in ganz neuer Weise zu erschließen und in eine Verteilte Digitale Forschungsbibliothek einzubringen.

Trier, im September 2004

Kurt Gärtner

1 Ein „Wörterbuch zum Hausbedarf“: Zur Digitalisierung des *Deutschen Wörterbuchs*

Das von Jacob und Wilhelm Grimm begründete *Deutsche Wörterbuch* (DWB) mit seinen von 1854 bis 1960 erschienenen 32 Wörterbuchbänden und dem 1971 in einem weiteren Band veröffentlichten Quellenverzeichnis bietet die bei weitem umfangreichste Erfassung der deutschen Sprache. Während der langen Bearbeitungszeit haben Generationen von Lexikographen etwa 325.000 Stichwörter auf rund 70.000 Spalten behandelt; das Quellenverzeichnis enthält mehr als 25.000 Titel und Verweise.

Im Jahr 1838 konnten die Brüder Grimm vom Leipziger Verleger Salomon Hirzel für den Plan eines großen neuhochdeutschen Wörterbuchs gewonnen werden und fanden sich nach den *Kinder- und Hausmärchen* (1812/15) noch einmal zu einer gemeinsamen Arbeit zusammen. Daraus sollte wieder ein Werk für alle werden; sein Gegenstand sollte die Darstellung des hochdeutschen schriftsprachlichen Wortschatzes sein in seiner Entwicklung und seinem Gebrauch von der Mitte des 15. Jahrhunderts, d. h. von der Erfindung des Buchdrucks um 1450 und die dadurch ausgelöste massenhafte Verbreitung von Literatur, bis zur Bearbeitungsgegenwart, die für die Brüder noch durch die sprachmächtige Gestalt Goethes bestimmt war. Sie wollten mit ihrem Werk ein Hausbuch für „leser jedes standes und alters“, ein „wörterbuch zum hausbedarf“ schaffen (Band I, Vorrede, Sp. XII f.), das sich im Besitz eines jeden an der deutschen Sprache Interessierten befinden und innerhalb der Familien zur Erbauung gelesen werden sollte.

Als die Brüder, die 1838 bereits Anfang fünfzig waren, mit der Ausarbeitung des Wörterbuchs begannen, veranschlagten sie dessen Umfang auf sechs bis sieben Bände, die Arbeitszeit auf zehn Jahre. Erst 1854 jedoch lag der 1. Band abgeschlossen vor, und allein die drei ersten Bände und der Anfang des vierten beschäftigten die Brüder bis an ihr Lebensende: Wilhelm vollendete die Buchstabenstrecke D kurz vor seinem Tode 1859, Jacob gelangte bis zum Artikel FRUCHT im Anfangsteil des vierten Bandes. Die Fortsetzung des Werks erfolgte unter wechselnden wissenschaftlichen und politischen Umständen, und erst 1960 wurde das Unternehmen in einer gemeinsamen Anstrengung von Ost und West durch die Berlin-Brandenburgische und die Göttinger Akademie der Wissenschaften abgeschlossen.

Durch die lange Bearbeitungszeit wurde das DWB zu einem unvergleichlichen Zeugnis für die Entwicklung wissenschaftlicher Theorien und Methoden des 19. und 20. Jahrhunderts, die die Arbeit verschiedener Lexikographengenerationen prägten, und durch Anlage und Umfang wurde es zu einem worthistorischen Grundlagenwerk der deutschen Sprache und zu einem gewaltigen Informations-

speicher für sprachwissenschaftliche Forschungen. Bedingt durch die lange Ausarbeitungszeit wurde es aber auch ein sehr heterogenes Werk. Diese Heterogenität hat jedoch nicht verhindert, daß dem DWB ein ungebrochenes Interesse entgegengebracht wird. Dies wird durch den zwischen 1984 und 1999 bereits in vier Auflagen erschienenen fotomechanischen Nachdruck, der bei dtv erschien, bestätigt. Dieser eroberte die Gelehrtenstuben und verhalf dem Werk trotz des kleinen, dem normalen Auge gerade noch zumutbaren Schriftgrades und trotz der unterschiedlichen Artikelstrukturen und Artikelumfänge zu einem beispiellosen Erfolg in der akademischen Welt. Für die meisten Benutzer des vielbändigen Großwörterbuchs war dennoch bis vor kurzem der Gang in die nächste öffentliche Bibliothek unvermeidlich.

Um die allgemeine Verfügbarkeit des bedeutendsten nationalsprachlichen Wörterbuchs des Deutschen noch weiter zu erhöhen und das DWB zu einem wirklich populären Werk, zu einem „wörterbuch zum hausbedarf“ im Sinne Jacob Grimms zu machen und seine Benutzung in vieler Hinsicht – von der Lesbarkeit und Übersichtlichkeit bis zur Einrichtung neuartiger Zugriffsstrukturen – erheblich zu erleichtern, lag es nahe, eine Digitalisierung dieses wichtigen Grundlagenwerks in Angriff zu nehmen. Gerade die elektronischen Medien bieten die Möglichkeit, dem großen Wörterbuch eine vor wenigen Jahren noch nicht vorstellbare Verbreitung zu sichern und damit das Vermächtnis der Brüder Grimm, denen die Popularität des von ihnen begonnenen Werks ein erklärtes Ziel war, zu erfüllen.

Der von Jacob Grimm gehegte Traum, das Wörterbuch könnte „zum hausbedarf, und mit verlangen, oft mit andacht gelesen werden“ (Band I, Vorrede, Sp. XIII), ist nun greifbarer denn je, nachdem der PC und ein Anschluß an das Internet immer mehr zur Ausstattung auch eines gewöhnlichen Haushalts gehören. Durch die elektronische Publikation des DWB im Internet und – mit erweiterten Recherchefunktionen – als Offlinemedium wird dieses für den Wissenschaftler und alle an der deutschen Sprache und ihrer Geschichte Interessierten wichtige Werk nicht nur unabhängig von den Öffnungszeiten der Bibliotheken einem großen Benutzerkreis zugänglich gemacht, sondern auch in seinem ganzen Reichtum an Informationen erschlossen. Das über die alphabetische Anordnung der Stichwörter führende und dem Druckwerk nachempfundene Nachschlagen wird ergänzt durch eine Vielzahl von Recherchemöglichkeiten, die von der einfachen Volltextsuche bis zu gezielt auf einzelne oder mehrere Informationspositionen eingrenzenden Suchvorgängen reichen. Das elektronische DWB kann ebenso als umfangreichstes rückläufiges Wörterbuch der deutschen Sprache benutzt werden wie als Morphemwörterbuch.

Der Erfolg der ersten Internetversion, die im Januar 2002 für die Buchstaben A bis G zur Verfügung gestellt werden konnte, und der vollständigen Internetversion, die seit Juni 2002 zugänglich ist, spiegelt sich in den ständig wachsenden Zugriffszahlen. Durch die in Kooperation mit Ecker Bonk entwickelte Präsentation

des digitalen DWB als Exponat auf der *Documenta11*, der weltweit größten Ausstellung moderner Kunst, stiegen die Zugriffsraten noch einmal erheblich an (vgl. dazu unten 4.2). Mit der Online-Version, die im Gegensatz etwa zum *Oxford English Dictionary* (OED) frei im Internet zugänglich ist, wurden die Ziele der *Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen* vom Oktober 2003 beispielhaft verwirklicht (www.zim.mpg.de/openaccess-berlin/berlindeclaration.html).

Von vornherein war eine Offline-Version geplant, um das digitale DWB für alle Interessierten, die nicht über einen Zugang zum Internet verfügten, attraktiv zu machen und eine noch weitergehende Verbreitung der mit erheblichen Fördermitteln geschaffenen elektronischen Ressource zu erreichen. Der Bedarf hierfür wurde besonders deutlich durch den Erfolg der für ein halbes Jahr lauffähigen Betaversion der CD-ROM, die im Dezember 2002 an mehr als hundert interessierte Benutzer verschickt wurde. Die Rückmeldungen zeigten zum einen die Notwendigkeit einer preisgünstig gestalteten offline-Publikation durch einen Verlag mit einem großen Vertriebsnetz, zum anderen trugen sie zur Optimierung der CD-ROM-Version bei. Diese wird zusammen mit einem Begleitbuch und einem Benutzerhandbuch vom Kompetenzzentrum in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften herausgegeben und seit dem 8. Juli 2004 durch den Verlag Zweitausendeins vertrieben. Die Offline-Version ist sowohl unter Windows als auch unter Linux lauffähig; eine Version für Macintosh-Systeme wird derzeit entwickelt und ist voraussichtlich ab Mitte Oktober 2004 erhältlich. Der Erfolg der CD-ROM ist ein weiterer Beweis für die große und ungebrochene Popularität des DWB, das durch die neuen Medien ein wirkliches „wörterbuch zum hausbedarf“ zu werden verspricht.

Wie schon in der Vorbemerkung angekündigt, sollen im folgenden der Projektablauf dargestellt und die Hauptarbeitsschritte ausführlich dokumentiert werden, die zur Erstellung des digitalen DWB durchzuführen waren. Dabei soll auch darauf eingegangen werden, welche Probleme nach und nach gelöst werden mußten und wie die Öffentlichkeit auf die Publikation der elektronischen Versionen reagierte. Es wird zunächst die Vorgehensweise bei der Erfassung und Fehlerkorrektur erläutert, die den ersten Hauptarbeitsschritt bildet (s. Kap. 2). Im Anschluß daran wird der zweite Arbeitsschritt, die Auszeichnung der Daten nach den internationalen Standards SGML/TEI beschrieben (3.1). Die Arbeiten an der sukzessiven Auszeichnung der Daten des Wörterbuchs (3.2) und des Quellenverzeichnisses (3.3) und die dabei aufgetretenen und zu berücksichtigenden Probleme werden ausführlich geschildert. In der Darstellung des dritten Arbeitsschritts wird berichtet, wie die elektronische Publikation erfolgt, d. h. welche Rolle dabei der Einsatz der Datenbanktechnologie spielt (4.1), wie diese für die Erstellung der Internetversion (4.2) und der CD-ROM-Version (4.3) genutzt wird und wie die entsprechenden Benutzeroberflächen erstellt wurden und funktionieren. Schließlich wird anhand der Publikationen, Präsentationen und Veranstaltungen des Projektteams die Öff-

fentlichkeitsarbeit dokumentiert (5). Gesondert wird über den großen Erfolg des digitalen DWB auf der *Documenta11* berichtet, der das Projekt national und international bekannt machte (6). Die Pressereaktionen werden aufgeführt (7) und zum Schluß die Perspektiven aufgezeigt (8), die sich für die weitere Bearbeitung der Daten des digitalen DWB und seine mögliche Verbindung mit anderen abgeschlossenen oder geplanten Projekten im Rahmen einer attraktiven „Verteilten Digitalen Bibliothek“ bieten. Die SGML-Dokumentation im Anhang soll noch einmal vor Augen führen, welche Herausforderungen in dem Projekt zu bewältigen waren, das nicht auf die allgemein bevorzugte, aber nur wenig mehr als die herkömmliche Buchform bietende Imagedigitalisierung setzte, sondern auf die Volltextdigitalisierung, um so die Möglichkeiten der neuen Technologie voll zu nutzen und eine wissenschaftlich anspruchsvolle und zugleich populäre digitale Ressource zu schaffen.

2 Datenerfassung und Fehlerkorrektur

Am 1. November 1998 wurde mit der Digitalisierung des DWB begonnen. Eine Imagedigitalisierung kam für das überaus komplexe DWB nicht in Frage, da sie lediglich die Möglichkeit bietet, über Metadaten auf das digitale Bild, d.h. auf ein elektronisches Faksimile etwa einer Wörterbuchseite, zuzugreifen, nicht aber auf einzelne Zeichenfolgen. Wir entschieden uns daher für eine Volltextdigitalisierung, die den Zugriff auf jedes einzelne Zeichen erlaubt und somit Ausgangsbasis für einen umfassenden Zugriff auf die Wörterbuchdaten ist.

Für die Volltextdigitalisierung mußte zunächst die Art der Erfassung geklärt werden. Ein naheliegender Weg könnte z. B. darin bestehen, den Text mittels eines Scanners erfassen zu lassen, der in einem ersten Schritt den Text abtastet und die einzelnen Druckzeichen in Form eines aus Punkten (Pixels) bestehenden Bildes abspeichert; in einem zweiten Schritt vergleicht eine spezielle Software für optische Zeichenerkennung (*Optical Character Recognition* = OCR) die einzelnen Buchstabenbilder mit einem Katalog von Mustern. Durch den Vergleich werden die Bilder durch eine digitale Repräsentation ersetzt, die der Computer weiterverarbeiten kann.

Entgegen landläufigen Ansichten und entgegen nachdrücklichen Empfehlungen kam das OCR-Verfahren für die Erfassung des DWB nicht in Frage, denn aufgrund der mangelhaften Druckqualität vieler DWB-Bände, der vielen Sonderzeichen und der wechselnden Schriftarten und Schriftgrößen war ein sehr hoher Prozentsatz an nicht erkannten und nachträglich einzufügenden oder zu verbessernden Zeichen zu erwarten. Zudem hätte man alle 67.744 Spalten auf herkömmliche Weise gegen die Druckfassung korrekturlesen müssen; da aber Korrekturleser bei einem Durchgang in der Regel lediglich 70% der Fehler erkennen, wären zum Errei-

chen einer akzeptablen Qualität von 99,99% mindestens vier Korrekturgänge erforderlich gewesen. Aus diesen Gründen entschieden wir uns für einen zweiten, scheinbar aufwendigeren Weg, nämlich die direkte Texterfassung im sogenannten *Double-Keying*-Verfahren. Dabei geben zwei Teams von Datentypisten unabhängig voneinander den gesamten Wörterbuchtext von Hand in den Computer ein. Zur Kontrolle der Eingabe wurde für einzelne Strecken auch das *Triple-Keying*-Verfahren angewandt.

Die Eingabe wurde von einem Erfassungsbüro in China durchgeführt. Aufgrund der bereits im Vorgängerprojekt „Mittelhochdeutsche Wörterbücher auf CD-ROM und im Internet“ gemachten guten Erfahrungen und der günstigen Konditionen erhielt ein weiteres Mal die Eingabefirma „TQY Double Key“ in Nanjing den Auftrag. Für die Erfassung in China sprachen neben finanziellen Gründen vor allem ganz praktische Überlegungen: Muttersprachler lesen „verstehend“ und korrigieren dabei das, was sie als „falsche“, d. h. von der normierten Standardsprache abweichende Schreibung wahrnehmen. Dies ist für eine zeichengetreue Wiedergabe der vielen historischen Belege im DWB natürlich nicht gewünscht. Es erwies sich daher als Vorteil, daß in China Nichtmuttersprachler am Werk waren, die tatsächlich nur das abschrieben, was sie sahen. Ein weiterer Vorteil war, daß die chinesischen Datentypisten von ihrer eigenen Schrift her gewohnt sind, auf jedes noch so geringe Detail zu achten, weil es eine bedeutungsunterscheidende Funktion haben kann. Sie registrieren daher kleinste Abweichungen im Druckbild eines Buchstabens sehr genau. Das betrifft vor allem die zahlreichen Sonderzeichen in den Etymologieteilen des DWB mit allen möglichen Akzenten über und unter den Buchstaben und die Wörter in griechischer Schrift, aber auch durchbrochene, verschmierte oder verstümmelte Buchstaben. Die Eingabe erfolgte also äußerst zeichensensibel, und es wurde jedes Diakritikum, das einem muttersprachlichen Erfasser unter Umständen aufgrund des schlechten Drucks oder des gemeinsamen Auftretens mit anderen Akzenten entgangen wäre, miterfaßt.

Die manuelle Eingabe hatte noch einen weiteren großen Vorteil: Alle für die spätere Datenauszeichnung relevanten und aufgrund der Typographie erkennbaren Merkmale konnten bei der Texterfassung bereits mit eingegeben werden, wodurch der späteren Datenauszeichnung in erheblichem Umfang vorgearbeitet werden konnte. In Trier war nach einer eingehenden Datenanalyse eine ausführliche Anweisung für die Eingabe des Wörterbuchs in China ausgearbeitet worden, anhand derer alle wesentlichen typographischen Merkmale schon bei der Texterfassung kodiert werden konnten (vgl. Abb. 1).

有关格林 (GRIMM) 词典输入工作的规定
样板 6

义项举例

6597 GEWÖHNLICH GEWÖHNLICH (I, 1 älteste belege) 6598

例 6. 1
义项第一级
大写罗马数字

例 6. 2
义项第三级
阿拉伯数字

普通段落

例 6. 3
义项第四级
小写拉丁字母

例 6. 4
义项第五级
小写希腊字母

GEWÖHNLICH, *adjectivum* und *adverbium*, erfolgreichste der *adjectivablationes* unserer *sippe*, namentlich für die *prosa* und *umgangssprache*. sie führt auf *gewon* (vgl. *ko-* *wohn* sp. 6470; sp. 6480) zurück. zu den *nebenformen* mit *dental* s. u.

I. älteste belege, sonderentwicklung in der *rechtssprache*, *statistik*, *formen*.

1) auf *deutschem* boden ist die *ableitung* spät bezeugt (vgl. *GRAFF* I, 870); erst die *mittelhochdeutsche dichtung* zeigt fortgeschrittene *entwicklung*, in der ihr das *angel-sächsische* zuvorkommt (vgl. *BOSWORTH-TOLLER* 471). dort sind *prädicative*, *attributive* und *adverbiale verbindungen* — mit und ohne *anlehnung* an *lateinischen gebrauch* — beobachtet: *þam folce was gewunelic* (it was usual with the people) *richter* 7, 8; *eall daet was gewunelic an dīsan lande* *Saxon chronicles* 238 *Barle* (aus 1100); *gewunellere mildtheortnyse, solita clementia*. *lat. hymnen* 11, 26; *gewunellce cōwōde, solita sit dicere* *Beda* 4, 19; *de gewunellce is nemned, consueta vocatur* 4, 4.

fürs *deutsche* bietet die *negirte form* den ersten *anhaltspunkt*: *insolite, ungewonlichho* *Hrabaniſche glossen* *STEINMEYER-SIEVERS* 1, 187 (anders *Keronische glossen*). hier waren aber *entsprechende formen* mit *gawon* zur hand, die das *nächste bedürfnis* völlig deckten: *ungawon, insolens, inusitatus; ungewon, inusitato*, sind *entsprechungen*, die von den *Hrabaniſchen glossen* bis in die *mittelhochdeutsche zeit* hinüberführen.

a) anders ist es mit der *positiven fassung* des *begriffs*, für die das *alte adjectiv* *giwon* nicht in allen *verwendungen* ausreicht. die *prädicative verbindung* mit dem *verbum substantivum* sietle hier durchaus auf ein *persönliches, runderst handelndes subject* und ist erst spät und ganz *vereinzelt* auf *unpersönliche subjects* bezogen, die das *object eines brauches* einführen: mit allen *demo* *daz gewon* ist s. oben sp. 6475. dagegen bilden solche *verwendungen* einen *hauptausgangspunkt* des *mittelhochdeutschen gebrauches* von *gewonlich* (s. u.). noch *deutlicher* wird das für die *attributiven verbindungen*, die bei *gewon* *spärrlich* belegt sind und die einen *wechsel* der *actionsart* dort bedingen, vgl.: und *nohten gewonag ritterspil* (var.: *gewöhnlichen*) *HEINRICH V. D. TURLIN krona* 29150 *Scholl*.

a) hierher weisen *nozi zeugnisse* des 12. *jahrhunderts*, die *ältesten*, die *GRAFF* a. a. o. für unsere *bildung* anführt: *dar nach ubir gewonlich zit wart swanger daz selbe wip. Miltler exodus* 124, 35 *Diemer* (s. auch unter c); vgl. auch: ein *ander tier* ... *heizet idris*, *daun zellet phisiologus daz iz ſlant si dem nikhuse* (nicks, *krokodil*) *unt ist digge sin gewonlich geslahte prosabearbeitung des Phisiologus* s. *Diutisca* 2, 25.

β) im *gegensatz* zum *substantiv* sind beim *abgeleiteten adjectiv* die *formen ohne präfix* auf *niederdeutsche belege* beschränkt: *wontlich* (*Hildegard* 1478) *weith*. 4, 606; vgl. *wontlik, wontlik* *SCHILLER-LÜBBEN* 5, 767. an der *substantivischen weiterbildung* greifen solche *formen weiter* aus: vgl. *wonlicheit* *LEXER* 3, 978; s. unter *gewöhnlichkeit*, *andererseits* ist eine *substantivform* zu unserem *adjectiv* mit *präfix*, auch in der *allgemeineren bedeutung* belegt, die an unserer *gansen gruppe* sonst nicht mehr zu beobachten ist (vgl. *dagegen* sp. 6471): *benedictus man ich dich* (*Maria*) *das diu gotlich frucht dines libes geseget was vor aller zit in sines vater ewiger gewonlichit* *spätmittelhochd. tractat* s. *Alemannia* 3, 106.

γ) *formen mit dental*, die aus *gewont* (s. sp. 6496) *erklärt*, aber auch *phonetisch gedeutet* werden können, begegnen nicht nur in *niederdeutschen*, sondern auch in *oberdeutschen denkmälern*: vgl. *gewondlich* *bayrische urkunde* von 1275 s. *deutsche städtechroniken* 4, 181; *Teuerdanc* 128; *gewandlich* *beter. weith*. 2, 146; *gewendlich* *Iglauer stadtbuch* 3, 174; *andererseits* vgl.: *unde han ... dīt vurgenantē teotament in eyne uffinbar instrument gemachet unde han daz geseigent myt myne gewontlichen zeichen ... Limburger urk. v. 1271 bei WYSS *Limd. chron.* 122; *komen de meyhayde to Brunswyk tohope uppe allen vyf radhusen in dem schyne gewontlike hovellude to kesende**

IV. 1. (4.)

HENNING-BRANDIS (1488) s. *dtische städtechron.* 16, 261; als it *gewontlich* was u. a. s. *SCHILLER-LÜBBEN* 2, 106; als byt her *gewontlich* ist *gewest urkundenformel* (1284) s. *BAUER-COLLITZ* 301; *dan idt is gewantlik, dat ... jūngere glosse zu reike d. v. 87 Brandes; sine wanunge unde gewontlike* 96; vgl. auch *gewontlichen* unter *gewöhnlichen*.

b) in der *mittelhochdeutschen dichtung*

α) löst die *prädicative verbindung* mit dem *verbum substantivum* den *gegensatz* von *gewonlich* gegen *gewone* am *sichersten* zu tage treten: *gewonlich giebt die möglichkeit unmittelbarer beziehung* auf den *geübten brauch*, während bei *gewone* erst der *umweg* über das *betheiligte persönliche subject* zu nehmen ist. nur in *einzelnen ausnahmefällen* greift auch *gewonlich* hier in den *bereich* von *gewone* über, wie *später gewone* in den von *gewonlich* (sp. 6479/80): *dō twang er die burger fūrer denn si gewon wrint: dā von wārent in die vrent. dō muocent im die gebūren mēr quotes gē dann si gewonlich wārent* aus *CHRISTIAN des küchenmeisters neuen casus sancti Galli* s. *Wackernagel lesebuch* 837, 16; *dazu vergleiche* aus den *niederdeutschen dialogen* *GREGORS*: *als he gewontlik was, als se gew. was* *SCHILLER-LÜBBEN* 3, 106.

1)) sonst *erfordern diese verbindungen ein neutrales subjectpronomen* und *berühren sich auf das engste* mit *entsprechenden verbindungen* des *substantivs gewonheit*; vgl.:

bi den *selben ziten* *was daz gewonlich, si wāre arme ode rich, daz si mohte durch ir māt riten swar si duhte gūt, unbesprochen und ane leit daz was dō gewonheit, swa man dāheine riten sach, daz ir nieman niht ansprach.* *WIRNT Wigolots* 2359 ff.:

wie bei den *rechtssprachlichen verbindungen* von *gewonheit*, so *stehen sich auch hier hauptstats- und nebensatzformen* gegenüber:

a)) *ouch ist daz gewonlich daz man dem sündigen man nach riuwen sünde vergebe.* *HARTMANN Iwein* 8104;

sō ist ot gewonlich, daz der bābest und sin rāt solher gehe niht enhāt, dāmit er iht verlie. *OTTOKAR bētr. reimechron.* 5408 *Seemüller*:

für die *abgrenzung der adjectivverbindung* gegen die des *substantivs* ist *bemerkenswerth*, dass *infinitive* an *stelle* des *substantivs* *steht*, wie sie bei *gewonheit* hier so *beliebt* sind, nur in *niederdeutschen belegen* beobachtet wurden: *gewontlik is ... kopluden over mere to varende* *Hieronymus* s. *SCHILLER-LÜBBEN* 2, 106.

b)) zu den *nebensatzformen* vgl.:

dō sprach der marcgrāve 'ir edeln künige rich, als ir nu wider ritet (daz ist gewonlich) heim ze Burgonden daz ir si mit ir fēret'. *Nidelingen* 1624, 2 *Lachmann*;

gegen:

vil manic ritter künne lac dō jamerliche erlagen. der wart gedunnen und getragen ab dem gevilde sē gehant und von den Kriechen dā gebrant, als ez dō was gewonlich. die burger ouch dō fliggen sīb, daz si des selben pflagen. *KONRAD V. WÜRZBURG trof. krieg* 37893;

die *besitzer der padstaben* ... *sullen 13 arme lent, die mit ouchen als gewontlich ist, darkomen, alle wochen frei paden* *Iglauer stadtbuch* III 174⁴ s. *JELINEK* 2, 318; als it *gewontlik* was (*dial. Gregors* 166) *SCHILLER-LÜBBEN* a. a. o.; *dat nicht gewontlik noch bequeme was* (*urkunden des hauses Stahlhof zu London* 121) *ebenda*.

3)) bei diesen *verbindungen*, die *ausschließlich eine mehrheit von personen als träger des brauches voraussetzen*, ist die *abgrenzung* gegen die *rechtssprachlichen verbindungen* (s. u.) oft *strittig*; vgl.: *ez was gewonlich in der alten e. swelh wib ein degenkint gebere, daz si sich darbete der kirchen vierzech tage spec. ecclesiae* 29 *Kelle*;

414

例 6. 5 非引文

Abb. 1: Anweisungen für die Eingabe des Wörterbuchs in China

Die Vorschriften der Eingabeanweisung betrafen die Wiedergabe der Band-, Spalten- und Zeilenzählung, die Ersatzkodierungen der in MS WORD nicht darstellbaren Zeichen, die Kennzeichnung von Zweifelsfällen, die Markierung der Lemmata und anderer ausdrucksseitig erkennbarer lexikographischer Informationspositionen.

Die Verszitate konnten dadurch, daß sie im Druckwörterbuch vom übrigen Text abgesetzt sind, eindeutig als solche erkannt und bereits bei der Erfassung in ## und ##{ eingeschlossen werden. Auch die den Versziten beigefügten Stellenangaben, die aus Kapitälchen für den Autornamen, Kursive für den Werktitel und/oder Buchstaben-Ziffern-Kombinationen für die Referenz bestanden und meist vom eigentlichen Zitat abgesetzt waren, konnten erkannt und mit #CC+ bzw. #CC– gekennzeichnet werden. Einzüge, die bestimmte Gliederungsabschnitte kennzeichneten, wurden ebenso miterfaßt und kodiert wie unterschiedlich weite Einrückungen bei Versziten. Später konnten diese Merkmale ohne weiteres in SGML-Kodierungen umgesetzt werden (vgl. Punkt 3.2).

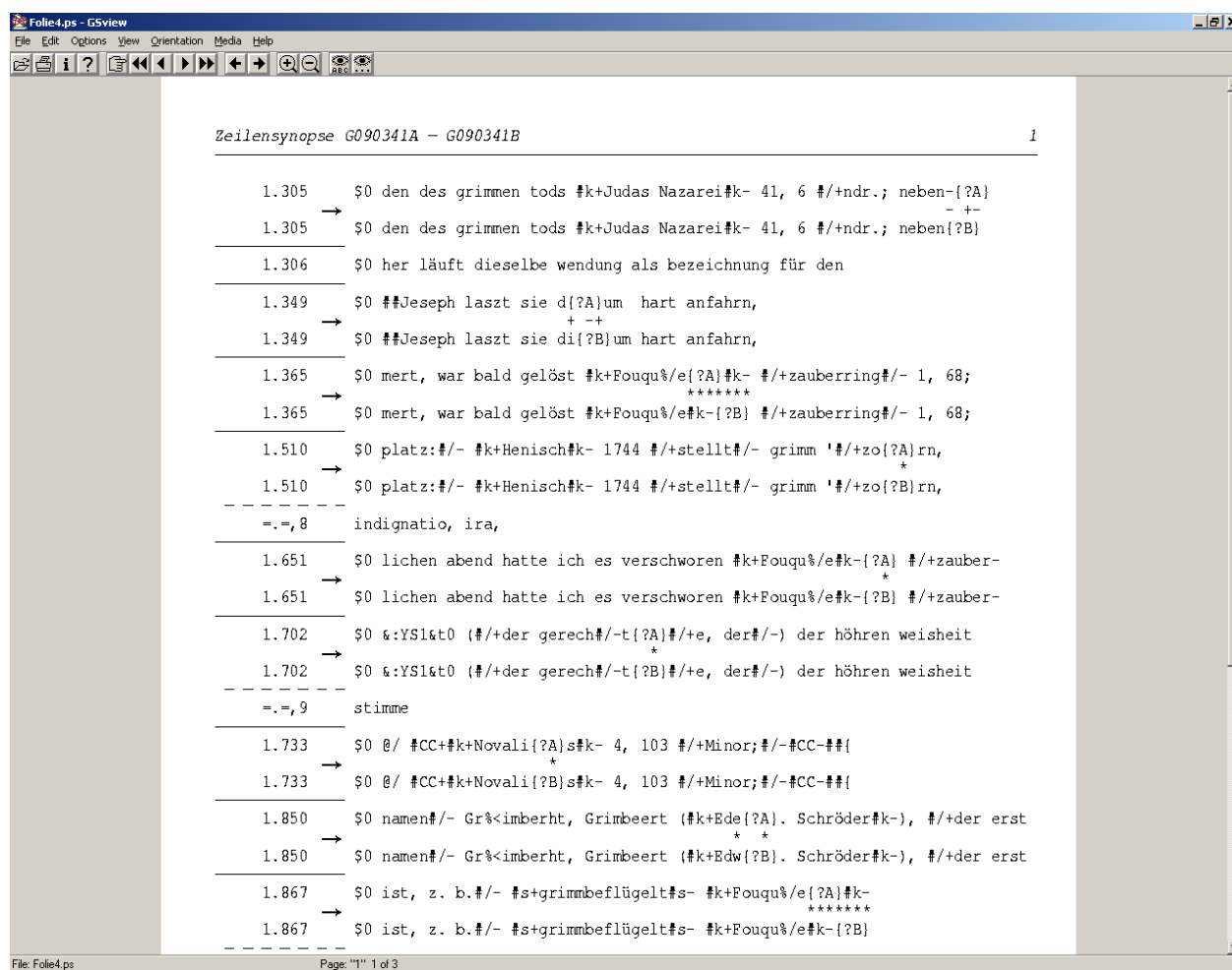


Abb. 2: Ausschnitt aus einem Vergleichsprotokoll

Die in China mit dem Textverarbeitungsprogramm MS Word für Windows erfaßten Textdaten wurden in Trier nach TUSTEP konvertiert, dem an der Universität Tübingen speziell für die Zwecke geisteswissenschaftlicher Forschung entwickelten *TUebinger System von TExtverarbeitungsProgrammen* (Informationen über TUSTEP unter www.uni-tuebingen.de/zdv/tustep/). Mit Hilfe des TUSTEP-Moduls #VERGLEICHE wurden die beiden Eingabeversionen automatisch miteinander verglichen. Sämtliche Abweichungen der beiden Versionen wurden dabei in einem automatisch generierten Protokoll festgehalten (vgl. Abb. 2), das anhand des Vergleichs mit dem Druckwerk manuell abgearbeitet wurde.

Das Vergleichsprotokoll enthält Zeilensynopsen von allen Stellen, an denen sich Eingabeversion A und Eingabeversion B voneinander unterscheiden. Die Kopfzeile des Vergleichsprotokolls in Abb. 2 gibt Aufschluß über den Band und die Spalte innerhalb des Wörterbuchs, mit der der verglichene Textabschnitt beginnt; hier wurde ab Spalte 341 im 9. Band des DWB verglichen (die Bandzählung bezieht sich auf die bei dtv erschienene Taschenbuchausgabe in 32 Bänden mit einem Quellenband). Ein Minuszeichen zeigt an, daß Version B desselben Textes an der entsprechenden Stelle ein Zeichen weniger enthält als Version A, ein Pluszeichen weist darauf hin, daß in Version B ein Zeichen mehr enthalten ist als in Version A. Asteriske kennzeichnen Abweichungen beider Versionen voneinander, etwa dort, wo die Datentypisten Zweifelsfälle markiert haben; {?A} bzw. {?B} war die verabredete Kennung für Fälle, bei denen man in China auf Probleme stieß: Zum Teil waren Zeichen nicht in den vorliegenden Kodierungsanweisungen enthalten – dies betraf vor allem Runenzeichen, hebräische und kyrillische Wörter, entlegene Währungsangaben oder ungewöhnliche Akzentkombinationen –, oder die Datentypisten waren sich generell unsicher, was an den betreffenden Stellen zu lesen war. Diese Fälle wie auch die anderen im Protokoll festgehaltenen Abweichungen wurden in Trier von Hilfskräften und Mitarbeitern im gedruckten Werk nachgeschlagen und nachbearbeitet. Dabei wurden Eingabefehler korrigiert, fehlende Zeichen nachgetragen, Kodierungen für bestimmte typographisch hervortretende Merkmale eingegeben und bestimmte Phänomene wie etwa unterschiedlicher Zeileneinzug bei gleicher inhaltlicher Funktion vereinheitlicht.

In einem weiteren Schritt wurden darüber hinaus eigens entwickelte Fehler-suchroutinen eingesetzt (vgl. Abb. 3), um Druckfehler und doppelte Eingabefehler – d. h. Falscheingaben, die von beiden Teams an der gleichen Stelle gemacht wurden – aufzuspüren. Die automatische Fehlersuche förderte Eingabefehler, in größerem Umfang aber Fehler der Druckausgabe zutage.

```

TUSTEP 47 DATENDS DATEN
Datei: DATEN+KOR.GRI GERMA127
2.0/8 |#DA,zwe2,seq-ap
2.0/9 |#K0,?1,zwe2,+,*
2.1 |>1z aeiouäöüy
2.1/4 |*****
2.1/5 |*2. Pruefung auf drei hintereinander vork. gleiche Zeichen:
2.1/6 |
2.1/7 |ZF+ |aaa|bbb|ccc|ddd|eee|fff|ggg|hhh|jjj|kkk|lll|mmm|nnn|
2.1/8 |ZF+ |ooo|ppp|qqq|rrr|sss|ttt|uuu|vvv|www|xxx|yyy|zzz|äää|ööö|
2.1/9 |ZF+ |üüü|
2.3/6 |*Sprungtabelle:
2.3/7 |SPN 0
2.3/8 |SPJ 3
2.4/9 |*eof
3.0 |#=: KOR03 #- 09.11.99 15:35:27
3.0/1 |
3.0/2 |#DA,zwe3,seq-ap
3.0/3 |#K0,?1,zwe3,+,*
3.0/4 |>1z aeiouäöüy#%
3.0/5 |>2z .,:;!?)
3.0/6 |*****
3.0/7 |*3. Pruefung auf E-C-Verwechslung:
3.0/8 |*In Zieldatei auf "C" achten. Die Umsetzung
3.0/9 |*von Suchroutinen fuer falsche c-e ist schwierig.
3.1 |*daher sind relativ viele Saetze in Zieldatei.
3.1/1 |
3.1/2 |ZF+ |fc|fch|fcz|gc|gchen|jc|jch|kc|kch|lc|llch|lck|lc>1|
3.1/3 |ZF+ |mc|mch|mc |
3.1/4 |ZF+ |nch|nc>1|nck|>1ncl.|>1nc|ncz|nc>2|nc ||nc|
3.1/5 |ZF+ |pc|pch|pck|qc|
3.1/6 |ZF+ |sch|scr|sc|sc>1|sc.|sc ||sc|
3.1/7 |ZF+ |tch|tck|tc.|tc>1|tc|vc|wc|
3.1/8 |ZF+ |exc|xc|
3.1/9 |ZF+ |zch|zcw|zc>1|zc|
3.2 |ZF+ |cd|cj|cp|cs|cv|cw| schen |
3.2/1 |ZF+ |>erc|arcg|cgm.|cgl.|cg|
3.2/2 |ZF+ |rcma| cm |cme|cm|
3.2/3 |ZF+ |cni||cn|
3.2/4 |SPN 0
3.2/5 |SPJ 3
3.2/6 |*eof
4.0 |#=: KOR04 #- 10.11.99 14:57:55
4.0/8 |#DA,zwe4,seq-ap
4.0/9 |#K0,g05.tus,zwe4,+,*
4.1 |>1z aeiouäöüy
4.3/2 |*4. (Nahezu) unmögliche Doppelkonsonanzen:
4.3/3 |ZF+ |jj|ww|yy|
4.3/4 |ZF+ |thh|chh|>1hh||hh|
*=2.0/8 Gib Anweisung >
** :01 ENTER DELETE INSERT 10:54

```

Abb. 3: Ausschnitt aus einer TUSTEP-Datei mit Fehlersuchroutinen

Die Kontrollroutinen trugen verschiedenen möglichen Fehlerquellen Rechnung: Durchaus ergebnisreich war das Aufspüren von doppelten Eingabefehlern bei bestimmten Buchstabenkombinationen; es wurden z. B. drei hintereinander vorkommende gleiche Zeichen oder auch für das Neuhochdeutsche eher ungewöhnliche Zeichenkombinationen wie yy gesucht (vgl. hierzu und zum folgenden die durch senkrechte Striche voneinander abgetrennten sogenannten Suchzeichenfolgen in Abb. 3). Dabei war im einzelnen zu prüfen, ob ein (Eingabe- oder Druck-)Fehler vorlag oder nicht. Häufig entstanden Fehler auch durch die Falscheingabe von *c* statt *e* aufgrund schlechter Druckqualität oder die Verwechslung von *fi* und Ligatur *fl*. Neben der Korrektur der Eingabefehler und der Bearbeitung der Zweifelsfälle, die sich im Durchschnitt auf etwa ein Vorkommen innerhalb einer Spalte beliefen, wurden die im DWB gefundenen Druckfehler in einem Druckfehlerverzeichnis zusammengestellt, das in der Zusammenstellung der Korrigenda einem Neudruck des DWB beigegeben werden könnte. Jeder eindeutige Druckfehler wurde für die elektronische Version korrigiert, doch ist in den Ausgangsdaten noch der ursprüngliche Text mit dem Druckfehler erhalten.

Der in China erfaßte Wörterbuchtext wurde in Portionen von jeweils 20 Spalten per E-Mail nach Trier übermittelt und dann in der geschilderten Weise weiterbearbeitet. Nach 18 Monaten lag das Wörterbuch digitalisiert, abgeglichen und korrigiert im TUSTEP-Format vor (vgl. Abb. 4). Die für das Projekt erstellten Routinen für die Konvertierung und den Vergleich der eingegebenen Daten können inzwischen mit kaum nennenswerten Modifizierungen in anderen Digitalisierungsprojekten des Kompetenzzentrums eingesetzt werden.

```

TUSTEP 47 DATENDS DATEN
Datei DATEN=GRIMMALL
GERMA127
90340.7 $0 _#F+#H:1#G:grimm.#F- #/+adj.#/-
90340.8 $0 _A. #/+#s+form.#/-#s- #/+dem west- und nordgerm. gemeinsames wort:
90340.9 $0 ahd. #/- grim, grimm, #/+asächs. ags. afries. #/- grim, #/+anord. #/-
90340.10 $0 grimr; #/+übereinstimmende bedeutung ist#/- ^#/#wütend, wild#/-,
90340.11 $0 #/+vielfach weiter entwickelt zu#/- ^#/#böse, furchtbar#/-#/+; im got. nur,
90340.12 $0 mit anderer ablautstufe, #/- gramjan ^#/#reizen#/-, #/+von einer idg.
90340.13 $0 wurzel#/- #H:#G:ghrem- ^#/#reiben, knarren, knirschen#/-, #/+die in gr. #/-
90340.14 $0 #G+xremizv#G-, #G+xreméuv#G- ^#/#mit den zähnen knirschen, wiehern#/-,
90340.15 $0 #G+xromow#G- #/+u. a. wiederkehrt, vgl. #/- #k+Fick#k- #H:3#G:3, 142; #k+Falk-Torp#k-
90340.16 $0 1, 340, 347. #/+die grundbedeutung tritt rein zu tage im ahd.
90340.17 $0 subst. #/- zano gagrim #/+stridor dentium Mons. fragm. #/- 10, 6, 24;
90340.18 $0 16, 2 #/-Hench. das adj. hat in mhd. zeit noch die vorhand
90340.19 $0 vor#/- grimmig. (#/+s. d. #/-), #/+in der nhd. zeit wird es allmählich
90340.20 $0 von dem abgeleiteten adj. zurückgedrängt, und in#/- 18. #/+jh.
90340.21 $0 ist es in gefahr, der concurrenz zu erliegen#/- (grimm, #/+olim
90340.22 $0 für#/- grimmig #k+Frisch#k- [1741] 373#H:b#G:; ^#/#im hochdeutschen ver-
90340.23 $0 altet, im oberdeutschen aber noch für grimmig, wütend,
90340.24 $0 grausam gebraucht#/- ^#/#Adelung#k-), #/+als es sich in die poe-
90340.25 $0 tische sprache rettet, um auf diesem gebiet noch ein kräf-
90340.26 $0 tiges leben zu entfalten. #/-
90340.27 $0 _B. #/+#s-bedeutung. #/-#s- #/+als ausgangspunkt der bedeutungs-
90340.28 $0 entwicklung ist der begriff der zähneknirschenden wut an-
90340.29 $0 zusetzen; und vielleicht steckt in der bei#/- #k+Murner#k- (#/+nach#/-
90340.30 $0 #k+Schmidt#k- #/+elsäss. #/- 155#H:a#G:.) #/+überlieferten redensart#/- ein grimen
90340.31 $0 für#/- grimmig #k+Frisch#k- [1741] 373#H:b#G:; ^#/#im hochdeutschen ver-
90340.32 $0 1) #/+am nächsten heran an diesen ausgangspunkt führt
90340.33 $0 die bedeutung#/- ^#/#in wut geratend, zornwütig, ergrimmt#/-#/+:#/-
90340.34 $0 so der grimme rihtare an der jüngsten urteile zorn-
90340.35 $0 lichen chumet #/+spec. eccl. #/- 143 #/+Kelle; #/-
90340.36 $0 #Hdaz zurnder sô sêre, _daz er mich dô sluoc
90340.37 $0 mit einer starken schalten: _des wart ich grimme genuoc
90340.38 $0 @/ #CC+#Nib. #/- 1545, 4:#CC-##{
90340.39 $0 #/+doch tritt diese bedeutung im mhd. gegenüber#/- 3 #/+zurück,
90340.40 $0 wie man etwa am Nib.-lied studieren kann; später häu-
90340.41 $0 figer: efferus#/- grim und schier als ertaubt #k+Frisius#k-
90340.42 $0 (1556) 460#H:b#G:; soll einem das nicht das hertz grim machen,
90340.43 $0 quod dicit ^dominus' #k+Luther#k- 29, 666 #/+Weim.; #/- der lew
90340.44 $0 schmecket usz dem geruch des pardi den eebruch der
90340.45 $0 lewin. darumb würt er grim das selbig z#;ou rechen
90340.46 $0 #k+Eppendorff#k- #/+Plinius#/- (1543) 50; #/+mit präpos.: #/- das pöfel
90340.47 $0 .. wider die Carthaginenser grim machen #k+Xylander#k-
90340.48 $0 #/+Polybius#/- (1574) 59; ist er in seinem gewüt erzürt und
90340.49 $0 gantz grim über Lewfriden worden #k+Wickram#k- 2, 351
90340.50 $0 #/+lit. ver.; und so bis in die neue zeit:#/-
90340.51 $0 #Hmein Tilly kam nicht aus dem hain,
90340.52 $0 er war erhitzt und grim genug
90340.53 $0 @/ #CC+#k+Droste-Hülshoff#k- 2, 158:#CC-##{
1 Zchflg. gefunden 1 Satz gefunden 1314714 durchsucht
*90340.7 Gib Anweisung >
**;01 ENTER DELETE INSERT 14:55

```

Abb. 4: Artikel „¹grimm“ im TUSTEP-Format

Obwohl nach dem Abarbeiten der Vergleichsprotokolle und der weiteren (halb-) automatischen Fehlerkorrektur schon eine weitestgehend fehlerfreie Version des Wörterbuchs vorlag, wurden Ausdrücke der aktuellen digitalen Fassung stichprobenartig gegen das Druckwörterbuch korrekturgelesen. Ergebnis dieses letzten Korrekturschritts war eine akzeptable Fehlerfreiheit von 99,997%, was etwa 3 Fehlern auf 10 Wörterbuchspalten entspricht. Dabei muß berücksichtigt werden, daß es sich bei diesen sogenannten „Fehlern“ nicht unbedingt um echte, d. h. auf falsches Abschreiben zurückzuführende Fehler handelt, sondern daß sich die Unstimmigkeiten

meist aus der schlechten Druckqualität des DWB ergeben haben, für die die chinesischen Datentypisten nicht verantwortlich zeichnen. Insgesamt war das Ergebnis der Eingabe äußerst zufriedenstellend, so daß die durch eine Trierer Initiative gegründete Erfassungsfirma auch anderen Projekten und Kooperationspartnern empfohlen und mit Erfolg eingesetzt werden konnte.

Von der korrigierten TUSTEP-Datei ausgehend wurden zum einen Programmroutinen für die Auszeichnung der Daten mit den internationalen Standards SGML und TEI entwickelt, zum anderen wurden auf der Basis der korrigierten Eingabedaten mittels einer TUSTEP-Satzroutine PDF-Files hergestellt, die das Wörterbuch so nachbilden, wie es im Druck vorliegt. Die PDF-Files des gesamten Wörterbuchs einschließlich des Quellenverzeichnisses sind fertiggestellt, und ein „print on demand“ ist möglich. Sobald die Sonderzeichen eingebunden sind, kann das gesamte Wörterbuch auch im PDF-Format mit dem digitalen DWB mitgeliefert werden.

Die Sonderzeichendarstellung bereitet allerdings generell noch Probleme, weil Sonderzeichen in verschiedenen Schriftarten und Schriftgrößen vorkommen können, viele dieser Zeichen noch nicht in den Zeichensätzen der elektronischen Medien zur Verfügung stehen und es auch keine standardisierten Abbildungsmöglichkeiten für diese Zeichen gibt. Für die Darstellung von annähernd 1200 Zeichen mußten Lösungen gefunden werden; soweit es möglich war, wurde auf Ersatzdarstellungen zurückgegriffen, etwa bei den in früheren Sprachstufen häufig vorkommenden übergestellten Buchstaben. Die seltener auftretenden Sonderzeichen, z. B. Runen, alte Gewichts- und Währungsbezeichnungen oder musikalische Notenzeichen, wurden durch Umschreibungen in eckigen Klammern ersetzt. Langfristig ist eine Lösung angestrebt, die möglichst umfassend den UNICODE-Standard berücksichtigt. Dagegen wurde die Einbringung von Zeichenimages, wie sie derzeit noch das OED verfolgt, aus mehreren Gründen ausgeschlossen: Zunächst ist das Einbinden von Zeichenimages sehr aufwendig, da diese in allen verwendeten Schriftgrößen vorgehalten werden müssen. Außerdem entstehen dadurch, daß mit den Images der Textfluß unterbrochen wird, größere Probleme bei der Durchführung von Suchanfragen und bei der zeilenweisen Darstellung der Wörterbuchartikel am Bildschirm. Die rasch fortschreitende Entwicklung von UNICODE läßt eine zufriedenstellendere Lösung in absehbarer Zeit erwarten, daher wurden alle Zeichen, die derzeit in den digitalen Versionen des DWB noch nicht exakt darstellbar sind, so kodiert, daß sie sofort umgesetzt werden können, wenn die entsprechenden Fonts zur Verfügung stehen.

3 Kodierung mit SGML nach den Richtlinien der TEI

3.1 Auszeichnung nach internationalen Standards: SGML/TEI

Die TUSTEP-Daten, die nach dem ersten Projektabschnitt vorliegen, bieten zwar ein genaues elektronisches Abbild der gedruckten Bände, dieses basiert jedoch auf rein typographischen Merkmalen und gibt nicht die eigentliche Mikro- und Makrostruktur des Wörterbuchs wieder. Deshalb war es notwendig, die Textdaten im Hinblick auf die Strukturinformationen eigens zu kodieren, ferner sollten sie unabhängig von spezifischen Programmsystemen zur weiteren elektronischen Publikation in Online- und Offline-Medien gleichermaßen genutzt werden können.

SGML (*Standard Generalized Markup Language*) bzw. XML (*eXtensible Markup Language*) sind Methoden, mit denen Textdokumente beschrieben werden. In einem SGML/XML-Dokument werden Informationen zu Inhalt und Struktur, nicht aber zum Layout mitgeliefert. Ein und dasselbe Dokument kann daher als gemeinsamer Ausgangspunkt für die Publikation in den gängigen Medien (offline, online, print) dienen. Das SGML/XML-kodierte Dokument selbst ist dabei unabhängig von Hard- und Software und kann langfristig genutzt werden.

Kernstück einer Kodierung nach SGML/XML stellt die Beschreibung der Dokumentstruktur dar: Durch eine bestimmte Menge von Regeln wird eine Klasse von Texten als ein bestimmter Dokumenttyp charakterisiert. Diese sogenannte DTD (*Document Type Definition*) wird für verschiedene Dokumentgruppen wie etwa Briefe, technische Dokumentationen oder Wörterbücher eigens spezifiziert und außerhalb der eigentlichen SGML-Daten gespeichert. Damit kann sie jederzeit auch für die Umsetzung anderer Projekte genutzt werden.

Wie schon im Vorgängerprojekt orientierte sich die Auszeichnung der Daten des DWB an den SGML-basierten Regeln, die von der TEI (*Text Encoding Initiative*) zur Kodierung von Wörterbüchern vorgesehen wurden (vgl. dazu den Abschlußbericht, <http://germa83.uni-trier.de/MWV-online/Docs/finale-mwv.pdf>). Da diese Regeln bewußt sehr allgemein gehalten sind, konnten sie ohne Probleme für die Auszeichnung des DWB verwendet werden. Unter „SGML-Tags“ oder „SGML-Kodierungen“ sind im folgenden stets die von der TEI vorgeschlagenen Tags zu verstehen.

3.2 Sukzessive Datenkodierung

Die Erfassung des DWB und die Fehlerkorrektur war der erste Hauptarbeitsschritt, dem weitere Bearbeitungsgänge folgen mußten. Parallel zum Eintreffen der ersten Daten in Trier und der abgeschlossenen Korrektur dieser Daten wurde damit be-

gonnen, Programmroutinen zu erstellen, die die von der TEI vorgeschlagenen metasprachlichen Kodierungen sukzessive in die TUSTEP-Dateien einbrachten.

Auszeichnung mit Hilfe von TUSTEP

Besonders vorteilhaft für die Auszeichnungsarbeit war, daß TUSTEP sich durch die Bevorzugung eines ökonomischen, expliziten Markups bereits einer Kodierung bedient, die die zu kennzeichnenden Elemente umschließt. Dementsprechend steht #k+ am Beginn, #k– am Ende von Text in Kapitälchen oder #/+ am Beginn, #/– am Ende einer kursiv zu setzenden Passage. HTML oder SGML verwenden zur Markierung einzelner Einheiten ebenfalls öffnende und schließende Tags; die TUSTEP-spezifischen Anfangs- und Endekennungen können also in bestimmten Fällen sofort durch SGML-Tags ersetzt werden, wie etwa bei der Markierung von Autornamen im DWB: Aus #k+Falk-Torp#k– wurde `<author rend="prose">Falk-Torp</author>`.

Außerdem waren bereits bei der Erfassung des Wörterbuchtexts bestimmte Kodierungen (Anfang und Ende der Verszitate und unterschiedlich weite Einzüge) in die Daten eingebracht worden, die nun umgesetzt werden konnten (vgl. Kap. 2).

Datenkodierung im *topdown*-Verfahren

Wie im Vorgängerprojekt wurde bei der Auszeichnung des DWB ein *topdown*-Verfahren angewendet, d. h. die obersten Ebenen der Artikelstruktur wurden zuerst markiert und von hier aus die weiteren Elemente in Angriff genommen (vgl. zur Kodierung der Daten Abb. 5).

Hauptlemmata, die im Wörterbuch in Versalien gesetzt sind, wurden bereits in China mit der TUSTEP-Kodierung #F+ und #F– markiert und konnten somit sofort als Lemmata „erkannt“ und in entsprechender Weise mit der SGML-Kodierung `<form> <form type="lemma"> LEMMA </form> </form>` markiert werden.

In dem Beispiel in Abb. 5 tritt ein Homographentrenner hinzu, der in einem `rend`-Attribut enthalten ist. Durch die Kennzeichnung der Lemmata konnten auch die Artikelanfänge eindeutig bestimmt und die einzelnen Artikel automatisch in `<entry>` und `</entry>` eingeschlossen werden. Die `<entry>`-Tags enthalten weitere Informationen zu den jeweiligen Artikeln: Zum einen wird die laufende Nummer des Eintrags zusammen mit der Buchstabenstrecke, in der sich der Artikel befindet, wiedergegeben: Im Beispiel ist dies im Ausdruck `id="GG25306"` enthalten, der anzeigt, daß es sich um Lemma Nr. 25.306 innerhalb der G-Strecke handelt. Zum anderen gibt das `n`-Attribut des `<entry>`-Tags Aufschluß über die Stelle, an der es im Druckwörterbuch zu finden ist, nämlich in Band 9 in Spalte 340 und Zeile 7.

legen zum Wortgebrauch. Ein Beispiel für ein neueres nationalsprachliches Wörterbuch, das von seinem Umfang und Anspruch her als mit dem DWB vergleichbar eingeschätzt werden kann, ist das *Oxford English Dictionary* (OED). Hier wurden jedoch von Beginn an bestimmte Vereinbarungen zur Artikelgestaltung befolgt, und auch bei der Digitalisierung der zweiten Ausgabe war es von großem Vorteil, daß es eine weitgehend einheitliche Artikelstruktur gab, an die man die Routinen zur Generierung der Datenkodierung schnell anpassen konnte. Das DWB hingegen weist eine enorme Heterogenität in der Artikelgestaltung auf, die sich zum einen aus der langen Entstehungszeit des Wörterbuchs, zum anderen aber auch aus der Vorliebe einzelner Bearbeiter für bestimmte Gliederungssysteme ergibt. Im 19. Jahrhundert existierten noch keinerlei Vorgaben für die Erstellung von Wörterbüchern, so daß die Bearbeiter bei der Abfassung ihrer Artikel nach Gutdünken verfahren. Erst in den dreißiger Jahren des 20. Jahrhunderts wurden Anweisungen für das Verfassen der DWB-Artikel erstellt. Von den Anfängen des DWB mit den Brüdern Grimm bis zum Ende seiner Ausarbeitung schlugen immer wieder zum Teil gegenläufige Bearbeiterpräferenzen durch.

Um die leichtere Orientierung im Wörterbuch zu ermöglichen, wird daher in der elektronischen Fassung ein Überblick über den Aufbau der Artikel in einem gesonderten Fenster angeboten, dessen Inhalte mit dem eigentlichen Wörterbuchtext im benachbarten Fenster verknüpft sind (vgl. unter 4.2 und 4.3). Voraussetzung für die Programmierung dieses Fensters war eine SGML-Kodierung, der entnommen werden konnte, welche Hierarchiestufe des jeweiligen Artikels gerade erreicht war, und die zudem Schachtelungen ihrer selbst erlaubte, damit die komplexe Struktur der Artikel korrekt nachgebildet werden konnte. Mit dem `<sense>`-Tag, der solche Schachtelungen erlaubt, konnte diese Kodierung realisiert werden. Die Kodierung `<sense n="B." level="2" value="9.0340.27">` enthält also die Information, daß der folgende Abschnitt mit der Gliederungsmarke *B.* eingeleitet wird, sich auf der zweiten Hierarchieebene des Artikels befindet und in Band 9 auf Spalte 340 in der 27. Zeile beginnt (vgl. Abb. 6; halbfetter Satz kennzeichnet in den Abbildungen die jeweils relevanten Merkmale).

Die `<sense>`-Tags wurden in die Daten eingesetzt, indem bestimmte Schemata der Artikelgliederung aus den Daten isoliert und den ProgrammROUTINEN zugrundegelegt wurden. Verschiedene Uneinheitlichkeiten mußten berücksichtigt werden: „1)“ etwa kann die dritte, zweite oder oberste Gliederungsebene bezeichnen, je nachdem, ob der Artikel mit „I. A. 1)“, „A. 1) a)“ oder „1) a)“ beginnt. Auch die Abfolge der Gliederungsmarken variiert: Neben „I. A. 1)“ kommt z. B. auch „A. I. 1)“ vor. Die Programme prüften, ob die Gliederung des jeweiligen Artikels einem der Schemata entsprach oder nicht. Zum Teil wurde daraufhin deren Bandbreite erweitert, zum Teil mußte manuell nachgebessert werden, etwa in Fällen, wo Gliederungsmarken durch Setzerfehler unterschlagen oder wo falsche Gliederungsmarken gesetzt waren. Bestimmte interne Kodierungen ermöglichten dann

eine „Korrektur“ der Gliederungsmarken, die in den Routinen berücksichtigt wurde.

```

<entry id="GG25306" n="9.0340.7">
  <form>
    <form type="lemma" rend="1">grimm
    </form>,
  </form>
  <gramgrp>
    <gram type="adj">adj.
    </gram>
  </gramgrp>
  <sense>
    <sense n="A." level="2" value="9.0340.8"> &kursiv;&Spaced;form
    [...]
      <emph rend="greek"> &khgr;&xrgr;&egr;&mgr;&igrac;&zgr;&ohgr;
      </emph>
      [...]
      <title type="sigle">
        <bibl>
          <author rend="prose">Fick
          </author>
          <ref>&Super;3&super;3, 142
          </ref>
        </bibl>
      </title>
    [...]
  </sense>
  <sense n="B." level="2" value="9.0340.27">
  [...]
    <sense n="1)" level="3" value="9.0340.32">
    [...]
      <add rend="vers">
        <q type="line">mein Tilly kam nicht aus dem hain,
        </q>
        <q type="line">er war erhitzt und grimm genug
        </q>
        <title type="sigle">
          <bibl>
            <author rend="verse">Droste-Hülshoff
            </author>
            <ref>2, 158
            </ref>;
          </bibl>
        </title>
      </add>
    [...]
  </sense>
</sense>
</entry>

```

Abb. 6: Artikel „¹grimm“ mit hierarchisch angeordneter SGML-Kodierung

Die Nachbildung der Hierarchie durch die oben erwähnte Schachtelung der `<sense>`-Tags wurde zum Aufbau des Artikelgliederungsfensters genutzt.

Unterscheidung der Trenn- und Bindestriche

Ein weiteres Beispiel für den Einsatz halbautomatischer Verfahren ist die notwendige Unterscheidung der Trenn- und Bindestriche: Da der ursprüngliche Zeilenfall der zweispaltig gesetzten Druckausgabe nicht beibehalten wurde, mußten zunächst die Zeilengrenzen beseitigt werden. Außerdem sind Reihungen im DWB nicht durch Kommata getrennt; statt eines eindeutigen

eichel-, grün-,

am Zeilenende mußte

eichel- grün-

aufgelöst werden. Zur effektiven Lösung dieses Problems wurde ein Programm erstellt, das die in einer Datei vorkommenden und aufgrund ihrer Umgebung als „problematisch“ erkannten Kurzstriche am Zeilenende nacheinander am Bildschirm aufrief und dem Bearbeiter erlaubte, je nach der Funktion dieser Striche unterschiedliche Kodierungen einzusetzen, die im späteren Programmablauf in bestimmter Weise „gedeutet“ wurden. Die Markierung `&-` stand dabei für solche Fälle, bei denen der Kurzstrich erhalten bleiben, die folgenden Zeichenketten aber sofort an diesen angeschlossen werden sollten. `&&-` wurde für solche Fälle verwendet, bei denen nach dem Bindestrich ein Leerzeichen stehen sollte. Auf diese Weise war z. B. die Unterscheidung von feststehenden Ausdrücken wie

dettes`&-`

à-dettes

und Reihungen in der Form

eichel- grün`&&-`

schellen- `#/+und#/-` herzdaus

gewährleistet; die übrigen Kurzstriche konnten aufgelöst und der nachfolgende Text direkt mit dem vorhergehenden verbunden werden. Die neu zusammengefügte Zeichenfolgen wurden danach in eine gesonderte Datei ausgegeben und nochmals überprüft.

Auszeichnung der Stichwörter in Reihen- oder Nestartikeln

Ähnlich aufwendig war die Identifizierung und Auszeichnung der in Reihen- oder Nestartikeln angesetzten Stichwörter. Aus Platzgründen sind Lemmata in Kompositareihen innerhalb eines Artikels vielfach so gedruckt, daß lediglich die Grundwörter gesperrt angegeben werden, während für das Bestimmungswort ein Bindestrich

steht (vgl. Abb. 7). Wird ein neues Bestimmungswort angesetzt und die Vollform nicht als Neueintrag, sondern im Fließtext angeführt, entfällt dieser Bindestrich und wird erst mit dem nächsten Grundwort, das nun mit dem vorangegangenen Bestimmungswort zusammengedacht werden muß, wieder gesetzt.

Schwierig war also zum einen das Auffinden des Bestimmungsworts, das teils mit dem Hauptlemma identisch war, teils aber auch erst im Artikelinnern auftrat. Zum anderen konnte das mit Spiegelstrich eingeleitete Grundwort nicht ohne weiteres um das Bestimmungswort ergänzt werden, sondern es mußten ggf. Fugens-, -es oder -en eingesetzt werden. Nach der automatischen Verarbeitung fielen umfangreiche Nachkorrekturen an. Außerdem wurden im nachhinein solche Fälle von der Auszeichnung ausgenommen, bei denen Wörter nur genannt, nicht aber durch grammatische Angaben, Bedeutungserläuterungen oder Belegzitate näher beschrieben waren (in Abb. 7 z. B. zum Bestimmungswort *giesz-* die Grundwörter *-becken*; *-brenke*).

TUSTEP-Datei:

gieszbach; #s+-bad#s- #/+heil-
kräftiges tropfbad#/- #k+Jacobsson#k- 5, 676#H:a#G;; #k+Adelung#k- 2, 684;
diese hatten ihm auch geraten, den wasserfall zu einem
gieszbad zu verwenden #k+L. Steub#k- #/+wanderungen i. bayr.
gebirge#/- 113; -becken; -brenke; **#s+-brunnen#s-** #/+im garten zum
begieszen der pflanzen#/-

SGML-Datei:

```
gieszbach;
</sense>
<sense>
<form>
<abbr>gieszbad</abbr>
<form type="subl" n="7.7411.43">-bad</form></form>
<sense> &kursiv;heilk&auml;ftiges tropfbad&recte;
<title type="sigle"><bibl>
<author rend="prose">Jacobsson</author>
<ref>5, 676&Super;a&super;</ref></bibl></title>;
<title type="sigle"><bibl>
<author rend="prose">Adelung</author>
<ref>2, 684</ref></bibl></title>; diese hatten ihm auch geraten, den wasserfall zu einem gieszbad zu verwenden
<title type="sigle"><bibl>
<author rend="prose">L. Steub</author>
<title>wanderungen i. bayr. gebirge</title>
<ref>113</ref></bibl></title>; -becken; -brenke; </sense>
<form>
<abbr>gieszbrunnen</abbr>
<form type="subl" n="7.7411.47">-brunnen</form></form>
<sense> &kursiv;im garten zum begieszen der pflanzen&recte;
```

Abb. 7: Sublemmata im DWB und ihre Umsetzung

Die Vollform des Sublemmas wurde in `<abbr>`-Tags eingeschlossen und für die Anzeige der Sublemmata in der Lemmaliste verwendet (vgl. Abb. 11 und Abb. 13). Die Korrekturen der Formen wurden mit Hilfe von Listen durchgeführt, die die Hilfskräfte mit den jeweiligen Vorkommen im Druckwörterbuch verglichen. Die aufgefundenen Sublemmata wurden, ähnlich dem Verfahren bei der Lemmaauszeichnung, mit Band-, Spalten- und Zeilenangabe versehen. In der aktuellen elektronischen Version sind etwa 80% der vorkommenden Sublemmata ausgezeichnet, für die restlichen Fälle müssen die übrig gebliebenen Sperrungen einzeln überprüft und die ggf. neu hinzukommenden Sublemmata der Lemmaliste hinzugefügt werden.

Durch das Einsetzen einer Identifikationsnummer auch für die Sublemmata kann die Verlinkung der Lemmaverweise im Wörterbuch erfolgen, die bereits weitgehend vorbereitet wurde: Zunächst wurden diejenigen Zeichenketten gesucht, die als Lemmaverweiszeichen gelten könnten, etwa „vgl.“ oder „s.“. Natürlich mußten diese Fälle daraufhin überprüft werden, ob sich daran tatsächlich ein Lemma anschloß bzw. ob es sich bei den aufgefundenen Zeichenfolgen wirklich um Verweiseinleitungen handelte. In einem zweiten Schritt wurden die Verweise ausgezeichnet; das Ergebnis wurde von Hilfskräften nachgeprüft bzw. korrigiert, denn vielfach gingen Verweise ins Leere, d. h. für das Verweiswort gab es keinen Lemmaansatz oder das Verweiswort war durch mehrere andere Wörter von der Verweiseinleitung getrennt oder ging dieser gar voraus. Zur Zeit wird an der Auszeichnung der Lemmata, auf die verwiesen wird, noch gearbeitet, denn hier sollen die Identifikationsnummern isoliert und bei den Verweisen mit

```
<ref target="IDENTIFIKATIONSNUMMER DES LEMMAS">
```

eingesetzt werden.

Der `<ref>`-Tag wird auch für die Verlinkung mit dem Quellenverzeichnis genutzt: Nachdem neben den Belegstellenangaben zu Verszitaten auch diejenigen zu Prosazitaten im Wörterbuch zum größten Teil ausgezeichnet sind, wird nun intensiv an der Verknüpfung der Belegstellenangaben mit den Einträgen im Quellenverzeichnis gearbeitet. Dabei werden die Identifikationsnummern des Quellenverzeichnisses in die Wörterbuchdaten eingebracht und sollen so umgesetzt werden, daß mit dem Anklicken einer Belegstellenangabe im Wörterbuch die jeweils zitierte Quelle im Quellenverzeichnis aufgeschlagen wird. Sowohl die Realisierung der Binnenverweise innerhalb des Wörterbuchs als auch die Verknüpfung mit dem Quellenverzeichnis werden Bestandteile des für Sommer 2005 geplanten Updates sein.

Als eine weitere Auszeichnungskomponente wurden ergänzend zu den im Wörterbuch bereits vorgegebenen Artikelpositionen Informationen über den Autor eines Artikels und die Lieferung, der der Artikel zuzuordnen ist, eingefügt. Diese Informationen waren jeweils in den Wörterbuchbänden und im Quellenband enthal-

ten; sie wurden digitalisiert und in einer Tabelle zusammengefaßt, auf die die Programmroutinen zugreifen.

3.3 Auszeichnung des Quellenverzeichnisses

Die Auszeichnung des Quellenverzeichnisses (Band 33 der dtv-Ausgabe) ging in ähnlicher Weise vonstatten wie diejenige des Wörterbuchs (vgl. Abb. 8): Auch hier mußte die Strecke, der ein Eintrag zugeordnet war, genau benannt und jeder Eintrag mit einer eindeutigen Identifikationsnummer versehen werden. Ebenfalls waren Spalten- und Zeilennummer mitzukodieren, und die Positionen der einzelnen Einträge mußten überdies so markiert werden, daß eine Reproduktion des Layouts und – zu einem späteren Zeitpunkt – eine Verknüpfung mit den Wörterbuchdaten ermöglicht wurde.

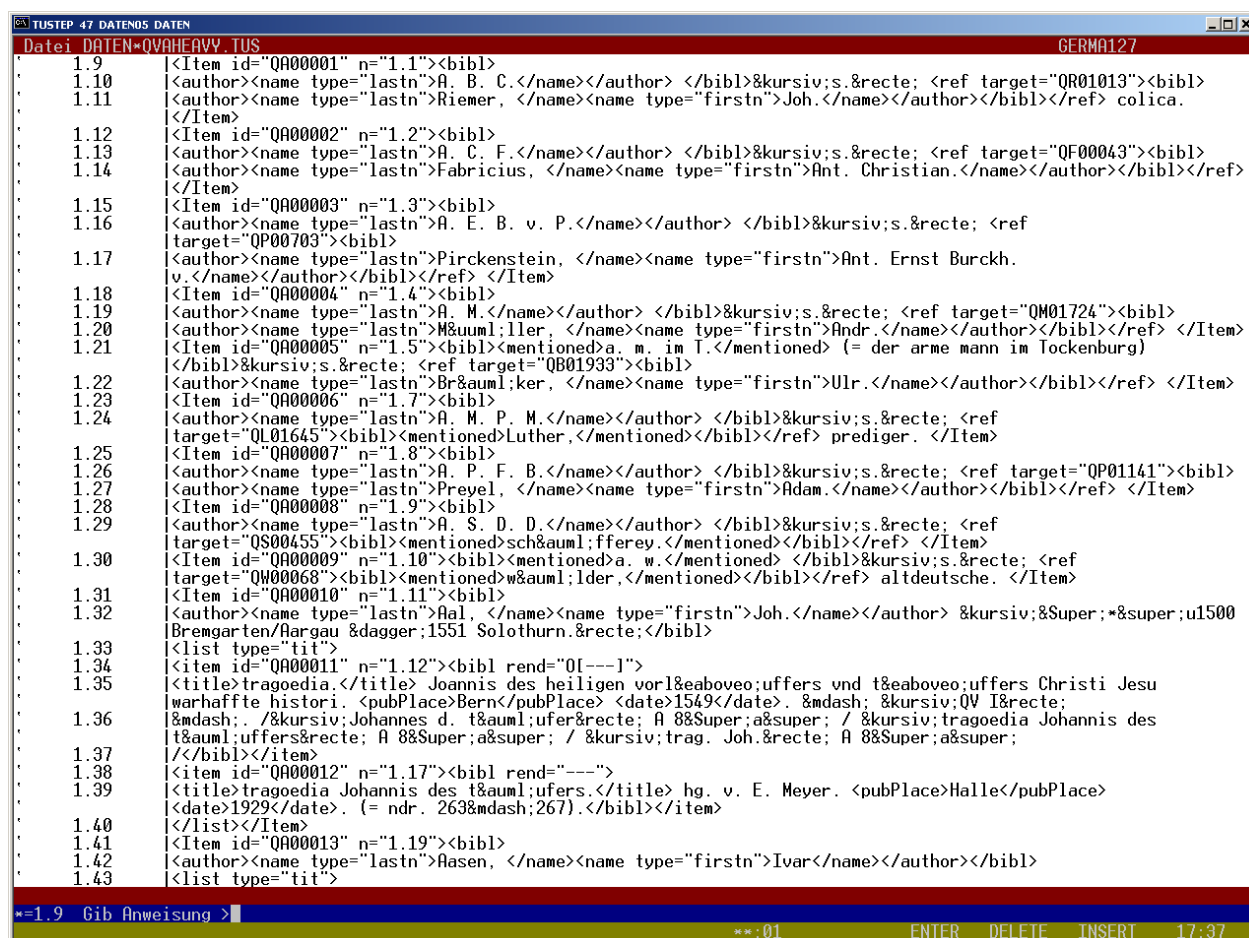


Abb. 8: Quellenverzeichnis im TUSTEP-Format mit SGML-Kodierungen

Da die Erstellung des 1971 erschienenen Quellenverzeichnisses sich auf einheitliche Konventionen stützte, ging die Auszeichnung relativ rasch vonstatten. Dabei konzentrierten sich die Arbeiten auf die Kodierung der Nach- und Vornamen der

Autoren und bestimmte Namenszusätze, die mit `<mentioned>`-Tags versehen wurden. Außerdem wurden die Quellentitel und Publikationsorte sowie das jeweilige Erscheinungsjahr ausgezeichnet und Haupt- und Nebeneinträge markiert.

Schon jetzt können Autoren und Quellen mit einer bestimmten Datierung innerhalb des Quellenverzeichnisses über die in der CD-ROM-Version eingebundene Suchmaske gesucht werden. Zusätzlich zu den Informationen, die das Quellenverzeichnis schon in gedrucktem Zustand liefert, wurden jedoch Eintragungen gemacht, die das Auffinden von Quellen aus einem festgelegten Zeitraum weiter erleichtern, indem bei Quellen, zu denen keine Datierung angegeben war, die ungefähre Entstehungszeit eingefügt wurde. Für das Update ist eine Zeitleiste vorbereitet, anhand derer der Benutzer Quellen und Belege aus bestimmten Zeiträumen auffinden kann.

4 Die elektronische Publikation

Die SGML/TEI-konforme Version dient als gemeinsame Basis für die verschiedenen Publikationsplattformen. Für die Bereitstellung im Internet werden die SGML-Daten nach HTML konvertiert, so daß sie mit üblichen Browsern angezeigt werden können. Es erfolgt zunächst eine typographische Aufbereitung, indem die SGML-Markierungen über entsprechende Style-Angaben in HTML-Anweisungen umgesetzt werden. Darüber hinaus werden die Verweisstrukturen innerhalb des Wörterbuchs in Hyperlinks abgebildet, die zu einem späteren Zeitpunkt verfolgt werden können. In ähnlicher Weise werden die Daten für die CD-ROM-Version aus den SGML-Dokumenten erzeugt. Für beide Plattformen wird eine Suchmaschine aufgebaut, die sowohl über die graphische Oberfläche der CD-ROM als auch über einen Web-Server angesprochen werden kann. Sie erlaubt dem Benutzer einen gezielten Zugriff auf den Wörterbuchinhalt: Zusätzlich zu einer herkömmlichen Volltextrecherche werden aufgrund der SGML-Kodierung detaillierte Abfragen einzelner Artikelpositionen ermöglicht, wie z. B. die Suche nach grammatischen Angaben, nach Stichwörtern, einzelnen Autoren sowie Ausdrücken aus nicht kursiven Textteilen wie z. B. den Belegziten oder kursiven Textteilen wie z. B. den Bedeutungsangaben.

Für das Benutzerhandbuch und den Umgang mit der elektronischen Publikation wurde eine konsistente, durchschaubare und im Hinblick auf ihre Aufgabe leicht verständliche Terminologie entwickelt (s. unter 4.3 und 4.4 die Termini *Buchstabenleiste*, *Stichwortsuchfenster*, *Lemmalaufleiste*, *Artikelgliederungsfenster* usw.). Einen Überblick über die Termini bietet das Register des Benutzerhandbuchs zur CD-ROM.

4.1 Datenbanken

Ausgangsbasis für den Aufbau der Datenbanken sind die SGML-Daten, deren formale Korrektheit von einem SGML-Parser kontrolliert wurde. Bei fehlerhaften Kodierungen erfolgte ein Korrekturschritt, der anschließend nochmals validiert wurde. Der Einsatz des Parsers gewährleistet, daß der SGML-Datenpool nur SGML-Dokumente enthält, die korrekt nach den TEI-Richtlinien kodiert sind.

Der Import der Daten in die Datenbanken für die Zielplattformen CD-ROM und Internet (vgl. Abb. 9) erfolgt über mehrere *Tool Command Language* (Tcl)-Skripte, die auf Basis des *Copenhagener SGML Tools* (CoST) entwickelt wurden. CoST ist als Tcl-Extension implementiert und arbeitet zusammen mit dem von James Clark entwickelten SGML-Parser *nsgmls*. Es setzt auf der durch den SGML-Parser ausgegebenen *Element Structure Information Stream* (ESIS)-Kodierung für SGML-Dokumente auf.

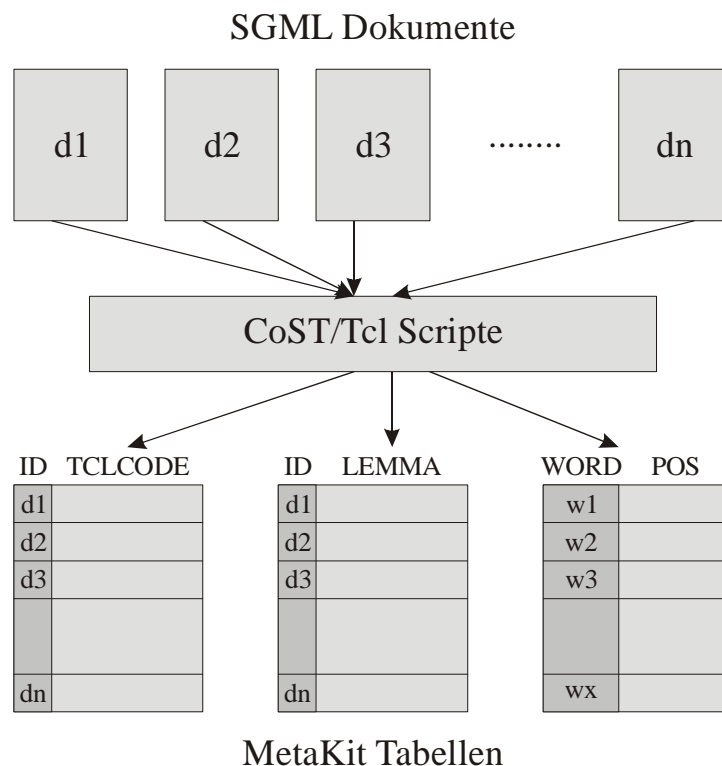


Abb. 9: Import der Daten in die Datenbanken

In Abb. 6 (vgl. 3.2) ist ein Artikel des Wörterbuchs den TEI-Richtlinien entsprechend kodiert dargestellt. Diese Kodierung verdeutlicht – durch die explizite Markierung der Artikelteile –, daß der Artikel auf seiner ersten Stufe in einen Formteil, grammatische Angaben und einen Bedeutungsteil unterteilt ist. Die Artikelteile

sind selbst wiederum in weitere Komponenten wie beispielsweise untergeordnete Bedeutungsebenen gegliedert.

Die SGML-Hierarchie des Wörterbuchs wird innerhalb der CoST-Skripte benutzt, um den betreffenden Elementen Aktionen zuzuordnen. Dazu wird eine CoST-Spezifikation definiert, d. h. ein Prozeß, der die SGML-Daten zu bestimmten Zwecken konvertiert, etwa zur Erzeugung einer Internet- oder einer CD-ROM-Version. Ein Prozeß besteht aus einzelnen Ereignissen, die an die SGML-Elemente gebunden werden können. Dabei kann man unterscheiden, ob die Aktion beim Antreffen eines Anfangstags, eines Endetags oder für den Inhalt eines Elements ausgeführt werden soll.

Ein Ausschnitt aus dem CoST-Skript zum Import der Wörterbuchartikel ist in Abb. 10 gezeigt. Durch dieses Skript werden Tcl-Statements für ein *Tool Kit* (Tk)-Textwidget erzeugt (vgl. fett markierte Anweisungen), um den Wörterbuchtext mit Hilfe des Taggingmechanismus der Tk-Textwidgets visuell aufzubereiten und eine möglichst originalgetreue Darstellung zu erzielen. Als Ausgabe der CoST-Skripte werden mehrere Datentabellen zur Verwaltung des Wörterbuchinhalts aufgebaut. Im wesentlichen werden dabei mit Hilfe des Softwarepakets MetaKit Tabellen zum Zugriff auf die Wörterbuchartikel, die Artikelhierarchie und Wortindizes für die Suchmaschine erstellt und verwaltet. Über die Tcl-Extension *Mk4tcl* werden die Daten direkt aus den CoST-Skripten in die betreffenden Tabellen importiert. In den in Abb. 10 gezeigten fett gesetzten und unterstrichenen Anweisungen werden die vorberechneten Tcl-Statement-Sequenzen in die zugehörigen MetaKit-Tabellen eingefügt. Über eindeutige Identifikationsnummern – sogenannte *Ids* – sind die Informationen zu den Wörterbuchartikeln später abrufbar. Interaktive Komponenten (z. B. Hyperlinks zu anderen Wörterbuchartikeln) werden über entsprechende Bindings auf den Widget-Tags definiert.

Das erste Ereignis der in Abb. 10 gezeigten Spezifikation bezieht sich auf ein `<entry>`-Element. Die Spezifikation beinhaltet eine Aktion für den Inhalt dieses Tags. Dort wird das Attribut `ID` zu `<entry>` – die aktuelle interne Numerierung des Wörterbuchartikels – abgefragt. Diese ID wird dazu verwendet, eine Marke auf den Artikelanfang in einem Tk-Textwidget zu setzen (mit `$w` wird das Textwidget angesprochen, in dem der Artikel später angezeigt werden soll). Diese Marke wird später genutzt, um den Artikelanfang in den sichtbaren Fensterbereich zu bewegen.

Eine zentrale Anweisung im gesamten CoST-Skript stellt die Zeile `do-Children dwb2tcl` dar, über die von einer SGML-Hierarchieebene in die nächsttiefere abgestiegen wird. In unserem Beispiel wird aus der `<entry>`-Ebene in die darin geschachtelte `<form>`-Ebene gewechselt. Für das Stichwort des Artikels wird durch die Anweisung `{element FORM withattval TYPE lemma}` ein Tcl-Statement zur Darstellung des Texts in einem Textwidget erzeugt. Das CoST-Statement besagt, daß die dort zugeordneten Anweisungen für alle SGML-

Elemente namens FORM ausgeführt werden sollen, falls zusätzlich das Attribut TYPE den Wert lemma hat (das Kriterium für das Stichwort des Artikels). Der Wortlaut des Stichworts wird zum einen über die Funktion ConvertTaggedText für das Textwidget eingetragen und zum anderen über die darauffolgende MetaKit-Anweisung in die Tabelle für die Stichwörter aufgenommen. Auf diese Weise werden in einem *depth-first-search*-Prozeß alle SGML-Elemente des aktuellen Dokuments behandelt.

```
specification dwb2tcl {
  {element ENTRY} {
    content {
      set tclcode ""
      set curr_lemid [query attval ID]
      AddToTCLcode "\$w mark set node\$curr_lemid insert-1c"
      AddToTCLcode "set ai \[\$w index insert\]"

      doChildren dwb2tcl

      AddToTCLcode "\$w ins e \{\n\} \{ 12 \}"
      AddToTCLcode "set ei \[\$w index insert\]"
      AddToTCLcode "\$w tag add lem\$curr_lemid \$ai \$ei"

      mk::set TCLDB.TCLCODE!\$maxtclid ID \$curr_lemid CODE \$tclcode
      incr maxtclid
    }
  }

  {element FORM withattval TYPE lemma} {
    content {
      foreachNode child {
        if {[string compare [query nodetype] "PEL"] == 0} {
          set text [string trim [content]]
          regsub -all "\n" $text "" text
          ConvertTaggedText $text w l

          mk::set DB.LEMMA!\$maxlemid ID \$curr_lemid LEMMA \$text
          incr maxlemid
        } else {
          doNode dwb2tcl
        }
      }
    }
  }
}
```

Abb. 10: CoST-Skript zur Wiedergabe von Wörterbuchartikeln

Nach der Bearbeitung der dem <entry>-Tag untergeordneten SGML-Elemente wird der für den Artikel vollständig generierte Tcl-Code in die betreffende MetaKit-Tabelle TCLDB.TCLCODE eingetragen. Dieser Prozeß wird während des Datenimports in die Datenbank für alle Artikel des Wörterbuchs ausgeführt, so daß im

Anschluß daran der Zugriff auf jeden beliebigen Abschnitt des Wörterbuchs möglich ist.

4.2 Publikation im Internet

Parallel zur Erstellung der CD-ROM-Version wurde, ebenfalls basierend auf dem SGML/TEI-Datenpool, die Bereitstellung des Wörterbuchs auf dem Server und die Entwicklung der graphischen Web-Oberfläche durchgeführt (vgl. Abb. 11).

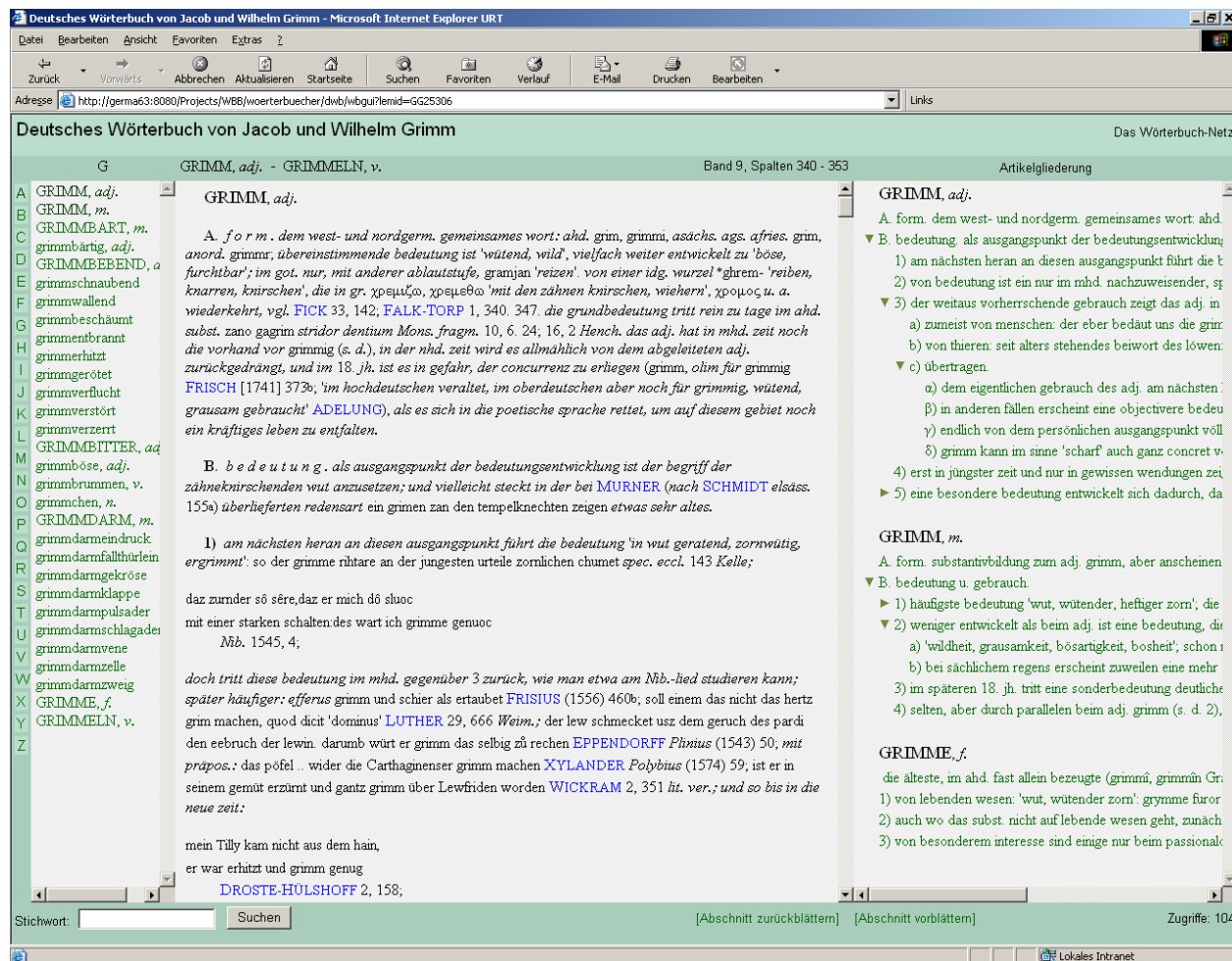


Abb. 11: Oberfläche der Internetversion

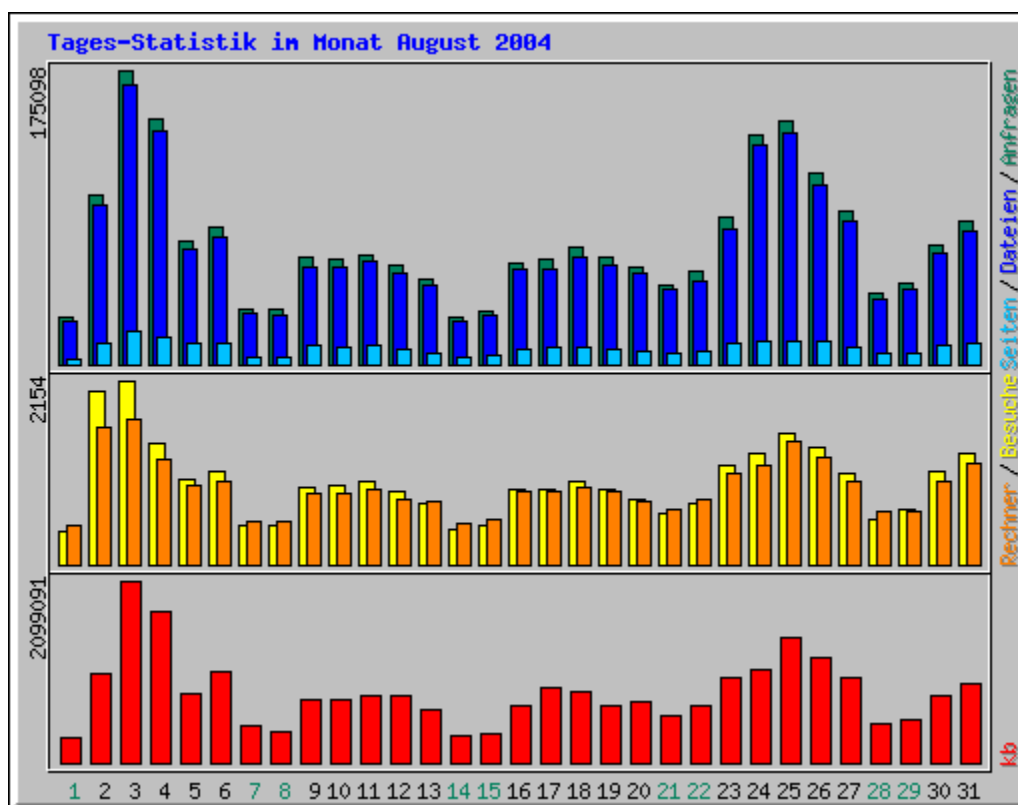
Im Aufbau orientiert sich die graphische Web-Oberfläche mit ihrer Einteilung in die *Buchstabenleiste*, die *Lemmalaufleiste* mit dem *Stichwortsuchfenster*, das *Artikelfenster* und das *Artikelgliederungsfenster* an der Oberfläche der CD-ROM (vgl. unter 4.3).

Ebenso wurde die Suchmaske, die voraussichtlich im Herbst 2004 eingesetzt werden wird, analog zur CD-ROM konzipiert. Sie wird in einem zusätzlichen Fens-

ter präsentiert werden und die grundlegenden Funktionen einer Volltextsuche, die mittels UND-, ODER- oder NICHT-Verknüpfung und Nachbarschaftssuche genauer spezifiziert werden kann, sowie eine Stichwortsuche umfassen.

Seit Juni 2002 ist das gesamte Wörterbuch über die Adresse www.DWB.uni-trier.de verfügbar. Diese Testversion wurde für die Browser Netscape 4.7 und Internet Explorer 5.5 optimiert. Die ständige Weiterentwicklung der Browser sowie ihre uneinheitliche Unterstützung von JavaScript lassen allerdings eine browserunabhängige Darstellung des Wörterbuchs im Internet geboten erscheinen. Deshalb wurde zur Darstellung der aktuellen Internetversion, die seit September 2004 online ist, ein *Content Management System* auf *open source*-Basis eingesetzt. Das System ZOPE kann auch für andere Wörterbücher, die u. U. mit dem DWB verbunden werden sollen (vgl. unter 8 Perspektiven), leicht angepaßt werden.

Bereits nach der Installation der ersten Internetversion waren Zugriffszahlen im Bereich von ca. 10.000 Anfragen pro Tag zu verzeichnen, die nach den Berichten in überregionalen Zeitungen wie der *Frankfurter Allgemeinen Zeitung* am 29.10.2002 (vgl. unter 7 Presseberichte) auf ca. 50.000 pro Tag anstiegen. Mit Erscheinen der CD-ROM beim Verlag Zweitausendeins und der damit ebenfalls verbundenen Werbung für die Online-Version stiegen die Anfragen an das Wörterbuch auf über 100.000 (die Spitzenwerte liegen nach der Aufnahme einer Pressemitteilung über das Online-DWB durch die dpa am 3.8.2004 bei über 175.000 Anfragen an einzelnen Tagen, vgl. Abb. 12).



Monats-Statistik für August 2004		
Summe Anfragen	2266883	
Summe Dateien	2114995	
Summe Seiten	299207	
Summe Besuche	29795	
Summe kb	25257484	
Summe unterschiedlicher Rechner (IP-Adressen)	21148	
Summe unterschiedlicher URLs	1003	
Summe unterschiedlicher Verweise	903	
Summe unterschiedlicher Anwenderprogramme	2001	
	Schnitt	Maximum
Anfragen pro Stunde	3046	14876
Anfragen pro Tag	73125	175098
Dateien pro Tag	68225	165864
Seiten pro Tag	9651	19154
Besuche pro Tag	961	2154
kb pro Tag	814758	2099091

Abb. 12: Zugriffsstatistiken für den Monat August 2004

4.3 Publikation auf CD-ROM

Die graphische Oberfläche für die CD-ROM ist auf der Basis von Tcl/Tk implementiert unter Verwendung der Extensions Tix und BLT. Sie ist von links nach rechts in verschiedene Navigationsbereiche untergliedert (vgl. Abb. 13). Über ein BLT-tabset-widget kann zunächst eine Wortstrecke aufgeschlagen werden. Ein Mausklick auf eine der Indexkarten der *Buchstabenleiste* am rechten Rand des Begrüßungsbildschirms öffnet das elektronische Wörterbuch, das sich dem Nutzer folgendermaßen präsentiert: In der *Lemmalaufleiste* im linken Fenster wird eine Liste aller Lemmata der ausgewählten Buchstabenstrecke abgebildet. Die Stichwortsuche ist möglich über diese *Lemmalaufleiste* oder direkt über das entry-widget für das *Stichwort* in der Fußleiste. Auf die in dieses Feld eingegebenen Zeichen reagiert die Lemmalaufleiste und zeigt die Stichwörter an, die mit der eingegebenen Zeichenfolge beginnen. Per Mausklick auf das gesuchte Lemma in der Lemmalaufleiste wird der gewünschte Artikel im *Artikelfenster* (rechts neben der Lemmalaufleiste) aufgeschlagen. Die *Kopfzeile* bietet Informationen zur aufge-

schlagenen Buchstabenstrecke und zum geladenen Abschnitt (hier GRIMM bis GRIMMBITTER). In der ersten Zeile des Artikelfensters (linke Hälfte des Hauptbereichs) finden sich Angaben zum aufgeschlagenen Lemma: Lemmavarianten, grammatische Angabe, exakte Position im gedruckten Wörterbuch (Band, Spalte und Zeile), außerdem lassen sich das Erscheinungsjahr der gedruckten Lieferung und der Artikelverfasser einblenden.

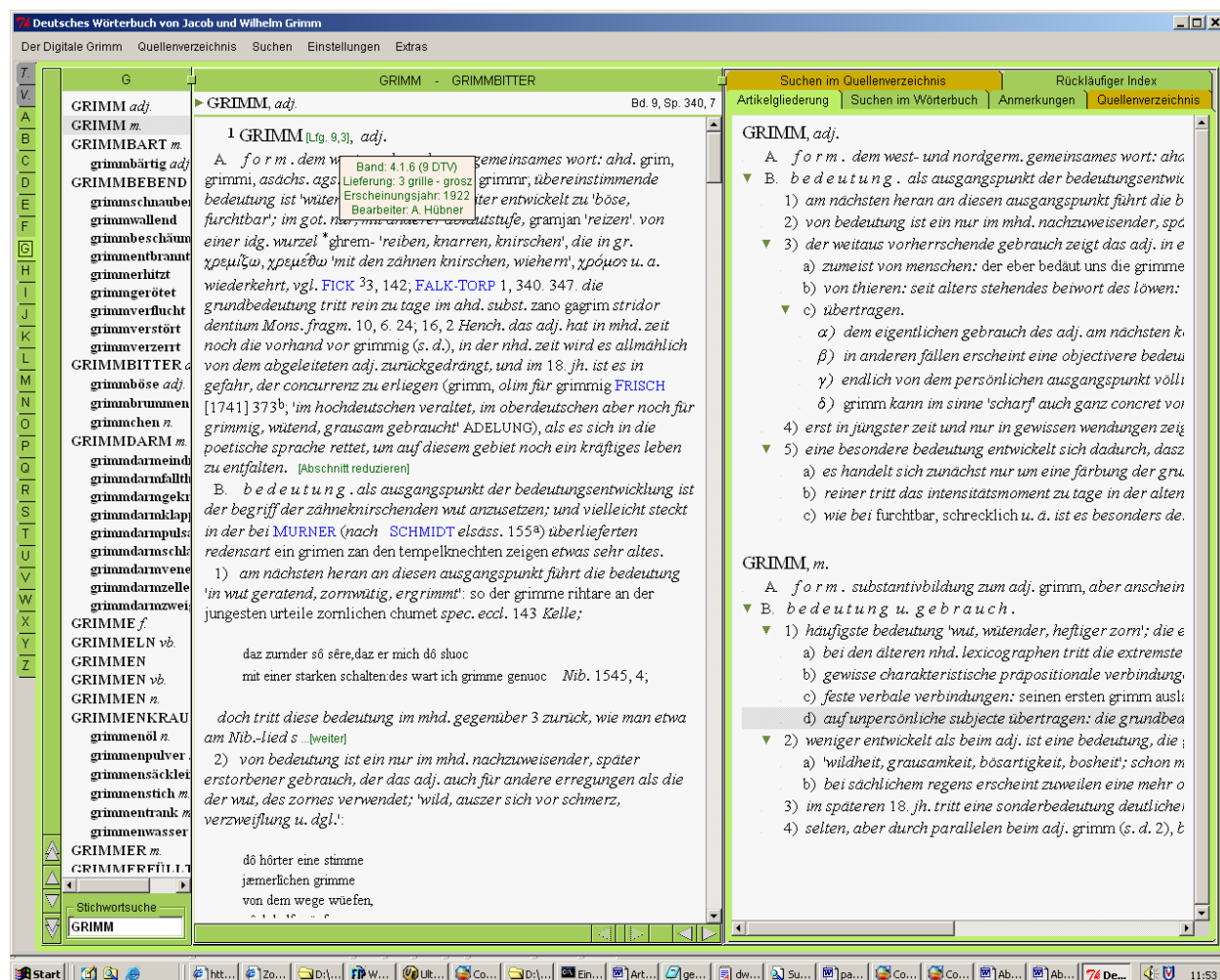


Abb. 13: Oberfläche der CD-ROM

Im Artikelfenster selbst wird der Wörterbuchtext unter Beibehaltung aller typographischen Eigenschaften abgebildet. Um bei den Benutzern die Akzeptanz der elektronischen Version des DWB im Vergleich zum Druckwörterbuch zu steigern, orientiert sich das Erscheinungsbild stark an der gedruckten Fassung. Es wird nicht nur der ausgewählte Artikel dargestellt, sondern eine größere Wortstrecke vorgehalten, so daß auch der Artikelkontext mit angezeigt werden kann.

Der Text längerer Artikel wird zur besseren Übersichtlichkeit zunächst verkürzt angezeigt. Dazu werden die Artikelabschnitte zu Beginn der Darstellung auf einen Anfangsteil reduziert und können durch die Markierungen [weiter] am Ende des Abschnitts ergänzt und durch einen Mausklick auf [Abschnitt reduzieren] entsprechend auch wieder verkürzt werden.

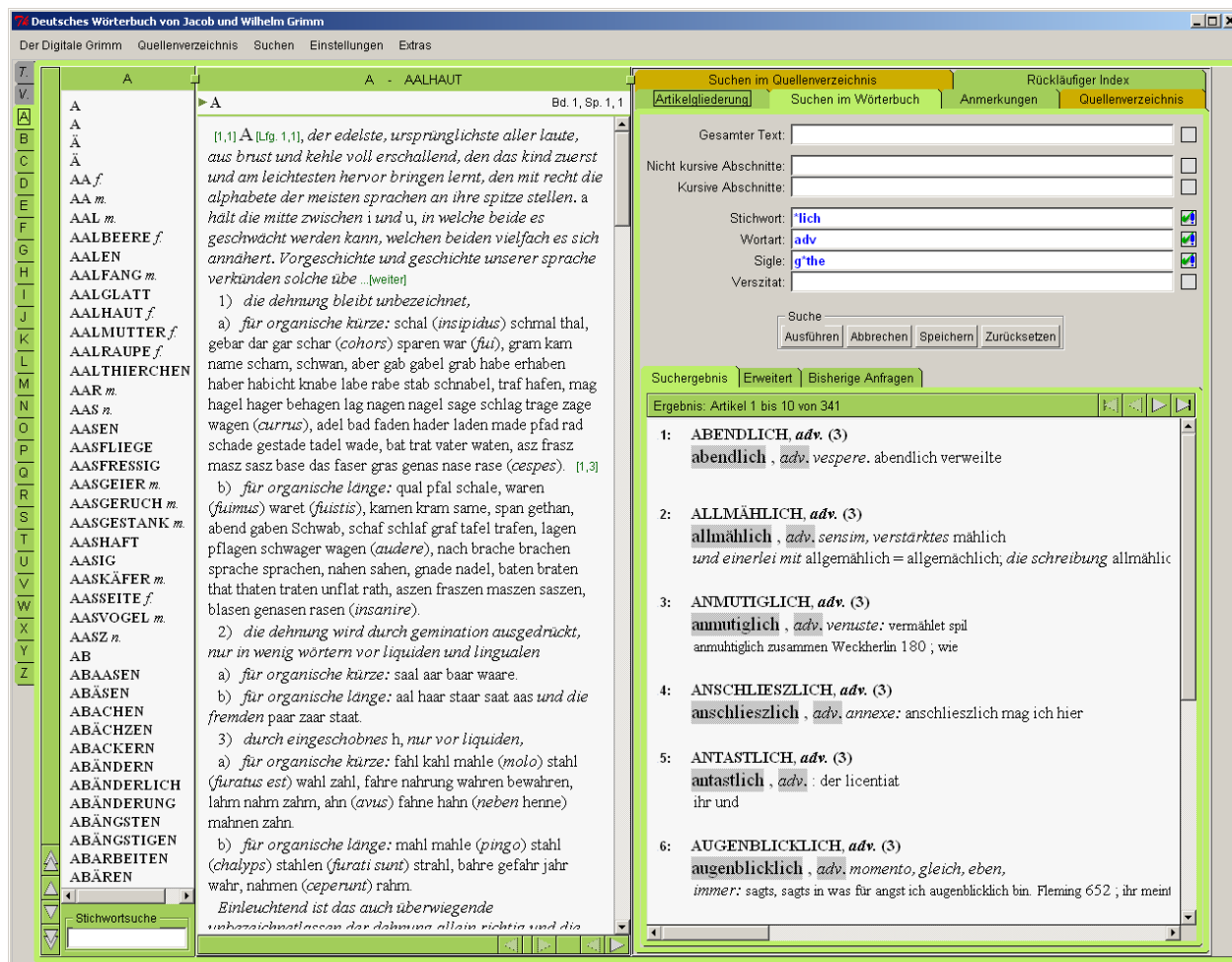


Abb. 14: Suchmaske mit Suchmodi (rechtes Fenster)

Zur Suche im Wörterbuchttext steht dem Nutzer eine elaborierte Suchmaske zur Verfügung, die eine Anzahl verschiedener Suchmodi bereithält (Abb. 14). Neben der einfachen Volltextsuche mit den üblichen Trunkierungsmöglichkeiten erlaubt die CD-ROM über verschiedene Eingabefelder auch Suchanfragen, die auf bestimmte Artikelteile beschränkt werden können: So kann gesondert in den Stichwörtern, den Wortarten (grammatischen Angaben), den Siglen und den Versziten und in den kursiven und den nicht kursiven Abschnitten der Wörterbuchartikel gesucht werden, wobei die Suchanfragen über logische Operatoren wie UND, ODER, NICHT kombinierbar sind.

Das Suchergebnis wird in Form eines KWIC (*Keyword in Context*)-Index unterhalb der Eingabefelder angezeigt und gibt auch Aufschluß über die Vorkommenshäufigkeit der Suchzeichenfolge. Es kann alphabetisch von A bis Z oder Z bis A, oder aber chronologisch von der ersten bis zur letzten oder der letzten bis zur ersten Lieferung sortiert werden. Die Möglichkeit der kombinierbaren Suchanfragen läßt komplexe Recherchen zu, die im gedruckten Wörterbuch nicht durchführbar sind: Ein Nutzer, der sich für das Thema „Wortbildung bei Goethe“ interessiert, kann z. B. eine kombinierte Suchanfrage formulieren, die alle Adverbien mit dem Derivativsuffix „-lich“, zu denen das DWB einen Goethe-Beleg beinhaltet, ausgibt. Hierzu gibt er im Eingabefeld *Stichwort* die Zeichenfolge „*lich“ ein, im Eingabefeld *Wortart* „adv“ und im Eingabefeld *Siglen* „G*the“ (der Asterisk steht für kein oder beliebig viele Zeichen und berücksichtigt die verschiedenen Schreibungen des Autornamens „Goethe“) und verbindet diese Elemente über die UND-Bedingung (Checkboxes am rechten Rand der Eingabefelder).

Mit einem Klick auf einen Treffer der Ergebnisliste im Suchfenster läßt sich die entsprechende Stelle im Artikelfenster aufschlagen (vgl. Abb. 14). Die Suchzeichenfolgen sind dort farbig unterlegt. Über die Pfeiltasten am rechten unteren Rand des Artikelfensters kann zum nächsten oder vorhergehenden Treffer der Ergebnisliste vor- bzw. zurückgeblättert werden. Auch das Quellenverzeichnis des DWB ist in die Suchmöglichkeiten mit eingebunden; die Verlinkung der darin enthaltenen Einträge mit dem Wörterbuchtext wird zur Zeit erprobt.

5 Publikationen und Öffentlichkeitsarbeit

Die Präsentationen und Publikationen des Projekts gingen zum einen aus einer regen Tagungs- und Konferenzteilnahme hervor, die sich sowohl im nationalen als auch im internationalen Rahmen bewegte, zum anderen verdanken sie sich dem großen Interesse der öffentlichen Medien, die immer wieder in Presseartikeln, aber auch in Rundfunk- und Fernsehbeiträgen über die Digitalisierung des DWB berichteten (vgl. unter 7).

Ein Überblick über die Präsentationen und Publikationen der Projektmitarbeiterinnen und -mitarbeiter sowie ein kurzer Bericht über das in der letzten Förderphase in der Akademie der Wissenschaften und der Literatur Mainz durchgeführte und die Projektarbeiten bilanzierende Kolloquium sollen im folgenden die Öffentlichkeitsarbeit des Projekts dokumentieren.

5.1 Projektpräsentationen

Die Entwicklung des Projekts wurde von Beginn an sowohl im In- als auch im Ausland mit regem Interesse verfolgt. Durch das Vorgängerprojekt, den Mittel-

hochdeutschen Wörterbuchverbund, waren bereits zahlreiche Kontakte zu nationalen und internationalen Forschergruppen, etwa zur TEI oder zu den Digitalisierungszentren in München und Göttingen geknüpft worden, die genutzt werden konnten, um zum einen Rückmeldung zur Internetversion und dem Prototyp der CD-ROM-Version zu erhalten und zum anderen Verbindungen zu anderen Digitalisierungsprojekten herzustellen mit dem Ziel, Informationen und Knowhow auszutauschen.

Besonders intensiv waren die Beziehungen zu den Institutionen, die sich mit der Weiterentwicklung des Standards SGML/TEI beschäftigten: So wurde das Projekt insgesamt dreimal zu Präsentationen im Rahmen der Gemeinsamen Internationalen Jahrestagung der *Association for Literary and Linguistic Computing* und der *Association for Computers and the Humanities* in Glasgow (2000), Tübingen (2002) und Athens (Georgia/USA) (2003) eingeladen. Mit der Anwendung der TEI-Guidelines bei der Digitalisierung des DWB befaßte sich ein Vortrag anlässlich des *TEI Members Meeting* in Nancy (2003).

Im Rahmen eher informatischer Veranstaltungen beteiligte sich das Projekt an einer Tcl/Tk-Konferenz in Vancouver/Kanada (2002). Im Rahmen eines Kolloquiums wurde es am Zentrum für Informatik der Universität Halle (2004) vorgestellt.

Gelegenheiten zur Dokumentation der Projektfortschritte ergaben sich auf nationalem Gebiet zunächst im Rahmen des 4. gemeinsamen Kolloquiums der Digitalisierungszentren München und Göttingen „Forum Retrodigitalisierung“ im November 2000. Hier konnte bereits ein erster Prototyp der CD-ROM-Version des DWB vorgeführt werden.

Mit der Verwendung von TUSTEP bei der Realisierung der elektronischen Version des DWB im Projekt gingen seitens der Tübinger Forschergruppe Bemühungen einher, TUSTEP durch neuere Programmodule im Hinblick auf automatisches Tagging weiterzuentwickeln. Gerade das DWB mit seinem großen Umfang und der Heterogenität der Typographie und Struktur stellte an die TUSTEP-Module und deren Stabilität und Flexibilität besondere Anforderungen; mit der Teilnahme an den TUSTEP-Workshops in Tübingen und Blaubeuren etwa zum Thema „Kommandomakros und interaktive Benutzeroberflächen“, „Praxis der SGML-konformen Textauszeichnung nach den Richtlinien der TEI“ oder „WWW – CGI – TUSTEP“ konnten Probleme beim Einsatz des Softwarepakets besprochen und Anregungen an die für die Weiterentwicklung Verantwortlichen herangetragen werden, die direkt in die Programmierung der Module einfließen. Dabei begegnete die Tübinger Entwicklergruppe den Wünschen der Trierer Projektmitarbeiter während des gesamten Projektverlaufs stets mit großer Offenheit und Kooperationsbereitschaft. Insbesondere die Jahrestagungen der International TUSTEP User Group, bei denen das Projekt 1999 und 2003 in Burgos vorgestellt wurde, waren für die Projektarbeit förderlich.

Auch bei stärker philologisch ausgerichteten Veranstaltungen konnte sich das Team stets mit großem Erfolg und Gewinn beteiligen, so etwa auf der Tagung „Edition und Übersetzung“ der Arbeitsgemeinschaft für Germanistische Edition 2000 in Osnabrück, auf der 3. Arbeitstagung deutschsprachiger Akademiewörterbücher 2000 in Zürich, im Rahmen des Deutschen Germanistentags 2001 in Erlangen, auf der Internationalen intradisziplinären Tagung zur Lexikographie in der Berlin-Brandenburgischen Akademie der Wissenschaften 2002 in Berlin und auf dem 4. Arbeitstreffen deutschsprachiger Akademiewörterbücher 2003 in Göttingen.

Der Höhepunkt der bisherigen Projektvorstellungen war zweifellos die Präsentation der CD-ROM-Publikation im Rahmen der Feiern anlässlich des 150jährigen Jubiläums des DWB im Juli 2004 in Berlin.

Die Wichtigkeit des direkten Austauschs mit ähnlich gelagerten Projekten und die Möglichkeit der gegenseitigen Information über angewandte Standards und jüngste Entwicklungen auf dem Gebiet der plattformunabhängigen Digitalisierung kann nicht genug betont werden, und es bleibt für die Zukunft zu wünschen, daß der beschrittene Weg der Orientierung an internationalen Standards auch innerhalb Deutschlands weiter fortgesetzt wird.

5.2 Publikationen

Publikationen, die sich mit dem Projektvorhaben beschäftigen, wurden in Fachzeitschriften, in Universitäts- und Akademiejournalen und in Tagungsbänden aufgenommen. Dazu trugen zum einen die oben bereits dokumentierten Konferenz- und Tagungsteilnahmen der Mitglieder des Projektteams bei, zum anderen aber auch konkrete Anfragen durch die Herausgeber einschlägiger Zeitschriften.

Die Publikationstätigkeit des Projektleiters und der Projektmitarbeiterinnen und -mitarbeiter soll durch die folgende Zusammenstellung auszugsweise dokumentiert werden, dabei sind die bibliographischen Angaben zur Online- und Offline-Version des DWB vorangestellt:

- Deutsches Wörterbuch von Jacob und Wilhelm Grimm im Internet, entwickelt im Fach Germanistik der Universität Trier in Zusammenarbeit mit dem Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften, bearbeitet von Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner, Vera Hildenbrandt, Thomas Schares, Klaudia Wegge. Gefördert durch die DFG im Rahmen des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ 2002. www.DWB.uni-trier.de.
- 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.

Herausgegeben 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. 1. Auflage, Juli 2004. Frankfurt am Main: Zweitausendeins 2004.

[Benutzerhandbuch von Vera Hildenbrandt und Tanja Wagner; 44 S.

Begleitbuch mit einer Einleitung von Kurt Gärtner und Beiträgen von Hartmut Schmidt: Das Deutsche Wörterbuch. Gebrauchsanleitung; Ecke Bonk: random walk/random reading; Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner, Vera Hildenbrandt, Thomas Schares, Klaudia Wegge: Wie das Deutsche Wörterbuch in den Computer kam; Martin Weinmann: Grimm-Bilderbogen; Wilhelm Grimm: Bericht über das Deutsche Wörterbuch; Jacob Grimm: Über das pedantische in der deutschen sprache; 131 S.]

- Bartz, Hans-Werner/Christmann, Ruth/Hildenbrandt, Vera/Schares, Thomas: Beyond the Scope of the Printed Dictionary. Digitizing the „Deutsche Wörterbuch“ by Jacob and Wilhelm Grimm. In: Standards und Methoden der Volltextdigitalisierung. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8./9. Oktober 2001. Hg. v. Thomas Burch, Johannes Fournier, Kurt Gärtner und Andrea Rapp (Akademie der Wissenschaften und der Literatur Mainz. Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse; Einzelveröffentlichung Nr. 9). Stuttgart 2003, S. 291-295.
- Bartz, Hans-Werner/Hildenbrandt, Vera: Das Deutsche Wörterbuch von Jacob Grimm und Wilhelm Grimm auf CD-ROM und im Internet. In: Wissenschaftliche Lexikographie im deutschsprachigen Raum. Beiträge zur Internationalen intradisziplinären Tagung zur Lexikographie in der Berlin-Brandenburgischen Akademie der Wissenschaften, 13.-16. Oktober 2002. Hg. v. Thomas Städtler. Heidelberg 2003, S. 445-449.
- Burch, Thomas/Christmann, Ruth/Hildenbrandt, Vera/Schares, Thomas : Ein „Hausbuch“ für alle? – Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. In: Jahrbuch für Computerphilologie 2, Paderborn 2000, S. 11-34. Außerdem in: Jahrbuch für Computerphilologie – online:
<http://computerphilologie.uni-muenchen.de/jg00/christma/christma.html>.
- Burch, Thomas/Gärtner, Kurt: Arbeiten des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier. In: Akademie der Wissenschaften und der Literatur Mainz. Jahrbuch 2000. Jahrgang 51. Stuttgart 2001, S. 253-261, hier S. 255-257.
- Burch, Thomas/Gärtner, Kurt/Schares, Thomas: Das digitale Deutsche Wörterbuch der Brüder Grimm. In: www.germanistik2001.de. Vorträge und Referate des Erlanger Germanistentages 2001. In Zusammenarbeit mit Petra Boden u.a. hg. v. Hartmut Kugler. Bielefeld 2003, Bd. 2, S. 713-728.
- Burch, Thomas/Gärtner, Kurt/Fournier, Johannes: Arbeiten des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier. In: Akademie der Wissenschaften und der Literatur Mainz. Jahrbuch 2001. Jahrgang 52. Stuttgart 2002, S. 251-264, hier S. 252-254.
- Burch, Thomas: Tcl/Tk-Based Electronic Publication of the Deutsche Wörterbuch. In: 9th Annual Tcl-Conference 2002 in Vancouver/Canada. Proceedings electronically available on CD-ROM, edited by Jeff Hobbs and Kevin Kenny. Online unter:

<http://aspn.activestate.com/ASPN/Tcl/TclConferencePapers2002/Tcl2002papers/burch-epub/dwbtcltk.doc>.

- Burch, Thomas/Hildenbrandt, Vera: „Random Reading im Book of Words“. Das an der Universität Trier digitalisierte Deutsche Wörterbuch auf der Documenta11. In: Unijournal. Zeitschrift der Universität Trier 28/2002, S. 25f.
- Burch, Thomas/Gärtner, Kurt/Fournier, Johannes: Arbeiten des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier. In: Akademie der Wissenschaften und der Literatur Mainz. Jahrbuch 2002. Jahrgang 53. Stuttgart 2003, S. 258-268, hier S. 258-261.
- Burch, Thomas/Gärtner, Kurt/Hildenbrandt, Vera: Das digitale *Deutsche Wörterbuch* der Brüder Grimm: vom Buchformat zur elektronischen Publikation. In: Bibliothek und Wissenschaft 36 (2003), S. 163-177.
- Christmann, Ruth/Hildenbrandt, Vera/Schares, Thomas: Ein „heiligthum der sprache“ digitalisiert: Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. In: Nicolás Castrillo Benito/Peter Stahl (Hg.): TUSTEP educa. Actas de Congreso del International TUSTEP User Group. Peñaranda de Duero (Burgos) Octubre 1999. Burgos 2001, S. 13-37.
- Christmann, Ruth: Books into Bytes: Jacob and Wilhelm Grimm's *Deutsches Wörterbuch* on CD-ROM and on the Internet. In: Literary and Linguistic Computing 16 (2001), S. 121-133.
- Christmann, Ruth/Schares, Thomas: Towards the User: The digital Edition of the „Deutsche Wörterbuch“ by J. and W. Grimm. In: Literary and Linguistic Computing 18 (2003), S. 11-22.
- Das digitale Deutsche Wörterbuch der Brüder Grimm: ein „Hausbuch“ für alle. Akten des Kolloquiums vom 15. und 16. Mai 2003 in Mainz. Hg. v. Thomas Burch, Kurt Gärtner und Vera Hildenbrandt. (Akademie der Wissenschaften und der Literatur Mainz: Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse). Stuttgart 2004 [in Vorbereitung].
- Falterbaum, Beate/Hildenbrandt, Vera: „Das digitale Deutsche Wörterbuch der Brüder Grimm: ein ‚Hausbuch‘ für alle“. Bericht zum Kolloquium am 15. und 16. Mai 2003. In: Unijournal. Zeitschrift der Universität Trier 29/3 (2003), S. 33f.
- Falterbaum, Beate/Hildenbrandt, Vera: Das digitale Deutsche Wörterbuch der Brüder Grimm: ein „Hausbuch“ für alle. Bericht über das Kolloquium an der Akademie der Wissenschaften und der Literatur Mainz am 15./16. Mai 2003. In: Lexicographica. Internationales Jahrbuch für Lexicographie 19 (2003), S. 320-322.
- Gärtner, Kurt/Wolz, Ralf: Aus der Arbeitsgemeinschaft Elektronisches Publizieren. In: Akademie-Journal. Magazin der Union der deutschen Akademien der Wissenschaften 1/2004, S. 67/68.
- Hildenbrandt, Vera/Schares, Thomas: Das Grimmsche Wörterbuch geht ins 21. Jahrhundert: Präsentation eines Prototyps des digitalen Deutschen Wörterbuchs von Jacob und Wilhelm Grimm. Vortrag im Rahmen des 4. Gemeinsamen Kolloquiums „Forum Retrodigitalisierung“ am 28./29. November 2000 in München.
Online unter: www.bsb-muenchen.de/mdz/forum/hildenbrandt/index.htm.

- Hildenbrandt, Vera: Ecke Bonk zeigt deutsche Worte: 100 Tage, im 28 Sekundentakt. In: FAZ.NET 1.7.2002.
- Kersting, Ruth/Rapp, Andrea: Mein schönes Fräulein ... Bedeutungs- und Bezeichnungswandel in Wortfeldern anhand des Grimmschen Wörterbuchs. In: Praxis Deutsch 165 (2001), S. 54-59.

Der folgende Bericht über das Projekt geht fast ausschließlich auf die einschlägigen Publikationen der Projektmitarbeiter zurück:

- Mreschar, Renate I.: Der „elektronische Grimm“. Das Deutsche Wörterbuch wird digitalisiert. In: Einblicke. Geistes- und Sozialwissenschaften in der DFG. Hg. v. Jutta Rateike (Deutsche Forschungsgemeinschaft). Bonn 2003, S. 63-73.

5.3 Kolloquium in der Akademie der Wissenschaften und der Literatur Mainz am 15./16. Mai 2003

Das Kolloquium zum Thema „Das digitale Deutsche Wörterbuch der Brüder Grimm: ein ‚Hausbuch‘ für alle“, das am 15. und 16. Mai 2003 in der Akademie der Wissenschaften und der Literatur Mainz stattfand und von der DFG in großzügiger Weise gefördert wurde, trug in besonderem Maße zu einem guten Projektabschluß bei. Hier wurden Offline- und Online-Version vorgeführt und verschiedene Aspekte der Projektarbeit vorgestellt, die von namhaften Fachwissenschaftlerinnen und Fachwissenschaftlern aus Retrodigitalisierungs- und Wörterbuchprojekten und Vertretern aus Bibliotheken und Verlagen aus Deutschland, den Niederlanden, der Schweiz und Großbritannien mit Statements begleitet wurden.

Einen Höhepunkt des Kolloquiums stellte der Abendvortrag Edmund Weiners dar, der Einblick in die Arbeit am digitalen *Oxford English Dictionary* gewährte. Die Gelegenheit, die eigene Arbeit im internationalen Vergleich zu sehen, war äußerst gewinnbringend für das Projektteam, wie überhaupt der Austausch der einzelnen Projekte von den Besuchern gern angenommen und die Demonstrationen verschiedener anderer Projekte aus Trier sehr begrüßt wurden.

Die Veranstaltung demonstrierte beispielhaft, wie aus der Zusammenarbeit von Philologie und Informatik und durch die Kooperation von Universitäten, Bibliotheken und Akademien neue Möglichkeiten für den Einsatz internationaler Standards und innovativer Technologien zur weiteren Erforschung, aber auch zur Erhaltung und umfassenden Nutzung geisteswissenschaftlicher Grundlagenwerke geschaffen werden können.

Ein Tagungsband, der die einzelnen Beiträge und Statements versammeln wird, ist in Vorbereitung und soll noch im Verlauf des Jahres 2004 in den „Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse“ der Akademie der Wissenschaften und der Literatur Mainz erscheinen.

6 Der digitale Grimm auf der *Documenta11* in Kassel, 8. Juni bis 15. September 2002

Von besonderer Bedeutung für das Projekt war die Zusammenarbeit mit dem Künstler und Typosophen Ecke Bonk, der sich mit einem Exponat zum DWB an der *Documenta11* beteiligte. Von Juni bis September 2002 war in Kassel eine in zwei Räumen untergebrachte Installation zu besichtigen, die die Entwicklung des DWB und seine noch nicht vollendete Geschichte demonstrierte. In einem ersten Raum waren die 380 Lieferungen der Erstbearbeitung des DWB in der chronologischen Reihenfolge ihres Erscheinens ausgestellt. Zusätzlich waren die 47 bereits erschienenen Lieferungen der Neubearbeitung mit in die Ausstellung hineingenommen. Als logische Ergänzung bot sich schließlich die Aufnahme der weiteren Entwicklung des DWB in seiner digitalisierten Form an: Durch verschiedene Projektionen der digitalen Version wurden dem Besucher jeweils andere Ansichten des Wörterbuchs präsentiert.

Im Durchgang aus dem ersten in den zweiten Raum der Installation, gleichsam den Übergang des gedruckten DWB zu einem neuen, digitalen Zeitalter veranschaulichend, wurde im Verlauf der 100 Ausstellungstage jedes einzelne Stichwort in einer vom Eintrag „A“ bis zu „Zypressenzweig“ durchlaufenden „Endlosen Liste“ angezeigt. In einem zweiten Raum wurden dann die Wörterbuchartikel selbst auf drei Wände projiziert; dafür wurden die Artikel per Zufallsgenerator aus dem gesamten Datenbestand des DWB ausgewählt. Nach einer gewissen Zeit, die je nach Artikellänge variierte, wurden – wiederum durch den Zufallsgenerator ausgewählt – drei neue Artikel präsentiert, so daß der Besucher sich mit immer neuen und unerwarteten Wörtern konfrontiert sah. Nach den Vorstellungen und Ideen von Ecke Bonk wurde die für das Exponat benötigte Software in Trier entwickelt und zusammen mit den Daten des digitalen Grimm zur Verfügung gestellt.

Für die Projektmitarbeiterinnen und -mitarbeiter bot die Installation Bonks zum einen eine willkommene Gelegenheit, ihre Arbeit einem großen Publikum zu präsentieren, zum anderen war die Begeisterung, mit der das Projekt Bonks aufgenommen wurde, Ansporn für die weiteren Projektarbeiten. Der Erfolg der Zusammenarbeit mit Ecke Bonk zeigte sich auch darin, daß die Stadt Kassel im September 2003 eine weitere Ausstellung der Arbeit Ecke Bonks und des Trierer Projekts unter dem Titel „Random Reading 2“ durchführte und sich um eine dauerhafte Installation des *Documenta*-Exponats bemüht.

Neben der Öffentlichkeitswirkung für das Projekt hatte die Zusammenarbeit mit Ecke Bonk aber auch direkten Einfluß auf das Erscheinungsbild des „Digitalen Grimm“: In der CD-ROM-Version findet sich unter dem Menüpunkt *Extras* die

von der *Documenta11* inspirierte Zusatzfunktion *Random Reading*: Klickt man diese an, so wird per Zufallsgenerator ein beliebiges Stichwort aus dem digitalen Grimm ausgewählt, dessen Bedeutung man weiter nachgehen kann. Auf der CD wird außerdem die in der *Documenta11*-Präsentation enthaltene „Endlose Liste“ aller Stichwörter des digitalen Grimm als Bildschirmschoner zum Herunterladen angeboten.

7 Berichte in Presse, Internet, Rundfunk und Fernsehen

Die Presse reagierte mit kleineren und größeren Beiträgen auf die Entwicklung der elektronischen Version des DWB. Durch die Zusammenarbeit mit dem Künstler Ecke Bonk im Rahmen der *Documenta11* wurde das Projekt zunächst einem größeren Publikum bekannt (vgl. unter 6). Mit der Publikation der Internetversion des DWB wuchs das Interesse an dem Digitalisierungsprojekt, das sich mit dem Erscheinen der Beta-Version noch weiter verstärkte. Das Erscheinen der vorläufigen und nun der endgültigen Version der CD-ROM-Fassung wurde ausführlich in den verschiedensten Organen besprochen.

Durch die ausführliche Berichterstattung über die Digitalisierung des DWB wurde nicht nur die Voraussetzung für einen immer höheren Bekanntheitsgrad des digitalen DWB selbst geschaffen, sondern auch die Vermittlung der wissenschaftlichen Inhalte und ihre Kommunikation in die Öffentlichkeit aktiv betrieben. Daß die Digitalisierung und damit die breite Verfügbarkeit des DWB von großem Interesse für alle Teile der Gesellschaft ist, zeigt sich in der Vielfalt der Organe, in denen über das DWB berichtet wurde und von denen hier eine Auswahl gegeben wird:

a) Zeitungs- und Zeitschriftenartikel, Internetbeiträge:

- Bocksrocker, Simone/Lenz, Daniel/Uebelhöde, Rainer: Die Wort-Welt ist rund – oder doch eine Scheibe? In: buchreport.magazin, Februar 2004, S. 86f.
- China macht's möglich. LEXIKON: Deutschland erhält digitales Grimm-Wörterbuch. In: Mannheimer Morgen, 3. August 2004, S. 28.
- China sei Dank: Deutschland erhält digitales Grimm-Wörterbuch. In: dpa KORR-Inland/Literatur/Sprache/Online/, August 2004.
- China sei Dank: Grimm-Wörterbuch auch digital verfügbar. Chinesische Gymnasiasten tippeten das „Deutsche Wörterbuch“ in den Computer. In: Der Standard, 10. August 2004. <http://derstandard.at/?url=/?id=1746728>.
- Chinesische Schüler tippten Grimms Wörterbuch ab. In: Braunschweiger Zeitung, 3. August 2004, S. 24.
- Der 600-Millionen-Zeichen-Deal. Was die Uni Trier in China mit den Gebrüdern Grimm treibt. In: Saarbrücker Zeitung, 3. August 2004, S. C5.

- Deutschland erhält digitales Grimm-Wörterbuch.
http://portale.web.de/Internet/?msg_id=5226937.
- Deutschland erhält digitales Grimm-Wörterbuch. Grundlagenwerk der deutschen Sprache in China in Computer eingegeben.
www.auswaertiges-amt.de/www/de/willkommen/deutschlandinfo/nachrichten_dtl/nachrichten_archiv_html?wb=2 (4. August 2004).
- Dirksen, Jens: Der Grimm für alle. Das Wörterbuch der Märchenbrüder liegt jetzt im Internet und auf CD-Rom vor. In: Neue Ruhr Zeitung, 14. August 2004.
- Dornröschen. In: Die Zeit, Nr. 33, 5. August 2004, S. 38.
- 300 Millionen gedruckter Zeichen in den Computer eingegeben. China sei Dank: Deutschland erhält das Wörterbuch der Gebrüder Grimm digitalisiert als Kauf-CD. In: Göttinger Tageblatt, 3. August 2004, S. 15.
- 300 Millionen Zeichen – in China abgetippt. Grimmsches Wörterbuch in digitaler Form. In: Wiesbadener Kurier, 7. August 2004, S. 14.
- Edinger, Jutta: „Digitaler Grimm“ erscheint. DFG-Projekt der historischen Sprachwissenschaft. In: Trierischer Volksfreund, Nr. 129, 4. Juni 2004, S. 11.
- Fleißarbeit in Fernost. Das Grimmsche Wörterbuch auf zwei CD-ROMs.
www.hr-online.de/website/rubriken/kultur/index.jsp?rubrik=5982&key=standard_document_1822090.
- Geisel, Sieglinde: Im Wortgestöber der Jahrhunderte. Das Grimm'sche Wörterbuch wird hundertfünfzig Jahre alt. In: Neue Zürcher Zeitung, Nr. 143, 23. Juni 2004, S. 35.
- GI Paris: Le dictionnaire de la langue allemande des frères Grimm - en ligne. L'immense dictionnaire des frères Jacob und Wilhelm Grimm accessible en ligne et bientôt sur CD-ROM. In: FplusD. Das deutsch-französische Sprachenportal.
www.fplusd.de/main/homepage/bereich/rubrik/artikel/index_html?mpId=62&article_id=360&mySelected=2.
- Giersberg, Dagmar: Das Deutsche Wörterbuch der Brüder Grimm wird digitalisiert. Beitrag für das Internetportal „Kultur und Gesellschaft“ des Goethe-Instituts.
www.goethe.de/kug/mui/int/thm/de60015.htm (Oktober 2003).
- Grimm digital. Großtat: Das Deutsche Wörterbuch als CD-ROM. In: Frankfurter Allgemeine Zeitung, Nr. 159, 12. Juli 2004, S. 30.
- Grimm digital. In: Mitteldeutsche Zeitung, 24. Juli 2004, S. 26.
- Grimm digital. Wörterbuch made in China. In: Badische Zeitung, 3. August 2004, S. 18.
- Grimms Deutsches Wörterbuch auf 2 CD-ROMs. Das „allerwichtigste Buch in deutscher Sprache“ endlich in elektronischer Version – der Digitale Grimm. In: Languagelab.at.
www.katzenmusik.com/index.php?id=36&backPID=36&swords=digitale&tt_news=155 (29. Juli 2004).
- Grimm-Werk aus Nanjing. In: Aachener Zeitung, 3. August 2004, S. 19.
- Grimm-Wörterbuch als deutsch-chinesisches Joint Venture. www.mediaculture-online.de/Details.305+M555ea2db7d8.0.html.

- Grimm-Wörterbuch auf Silberscheiben. In: Freie Presse, 30. Juli 2004.
- Heine, Matthias: Blüdel Glimm. In: Die Welt, 3. August 2004, S. 27.
- Heine, Matthias: Wie die Blüdel Glimm an uns vorbeigingen. In: Berliner Morgenpost, 4. August 2004. <http://morgenpost.berlin1.de/archiv2004/040804/feuilleton/story694980.html>.
- Hübl, Michael: Deutsch made in China. Das Grimmsche Wörterbuch ist jetzt digitalisiert. In: Badische Neueste Nachrichten, 11. August 2004.
- Kroener, Wolfgang: Jetzt dozieren die Grimms digital. Das 32 Bände umfassende Wörterbuch auf CD-ROM – Trierer Wissenschaftler erarbeiteten die elektronische Fassung. In: Rheinzeitung, Nr. 188, 14./15. August 2004, S. 108.
- Lingnau, Fee Isabelle: Warum Chinesen besser abtippen. Sprachschatz. Das Deutsche Wörterbuch der Brüder Grimm gibt es jetzt auf CD-ROM, eingetippt in Nanjing. In: Hamburger Abendblatt, 16. Juli 2004.
- Link, Susanne: „Das wichtigste Buch in deutscher Sprache“. Die digitalen Brüder Grimm. In: Frankenpost, 7. August 2004, S. 20.
- Maier, Thomas (dpa): China sei Dank. Deutschland bekommt digitales Grimm-Wörterbuch. In: Sächsische Zeitung, 3. August 2004, S. 11.
- Maier, Thomas (dpa): China sei Dank: Deutschland erhält digitales Grimm Wörterbuch. In: dpa Landesdienst Hessen, 2. August 2004.
- Maier, Thomas (dpa): China sei Dank: Digitaler Grimm. Wörterbuch – Gymnasiasten tippen in Nanjing das Grundlagenwerk der deutschen Sprache per Hand in den Computer. In: Darmstädter Echo, 3. August 2004, S. 8.
- Maier, Thomas: China und die „frouwe“. In: Fuldaer Zeitung, 7. August 2004.
- Maier, Thomas: Chinesen geben deutsches Wörterbuch in den Computer ein. In: Badisches Tagblatt, 3. August 2004.
- Maier, Thomas: Dank China richtiges Deutsch. Grimm-Wörterbuch auf CD-ROM erschienen. In: Trierischer Volksfreund, 3. August 2004, S. 4.
- Maier, Thomas: Flinke Finger vollenden Forschungsprojekt. In: Die Glocke, 3. August 2004.
- Maier, Thomas (dpa): Grimms aus China. In: Wiesbadener Tagblatt vom 3. August 2004. www.main-rheiner.de/region/objekt.php3?artikel_id=1564360.
- Maier, Thomas: Gymnasiasten tippen 300 Millionen Zeichen. China sei Dank: Deutschland erhält ein digitales Grimm-Wörterbuch. In: General-Anzeiger. 18. August 2004, S. 13.
- Maier, Thomas: Meterweise Platz im Bücherregal gespart. Die 32 Bände des Grimmschen Wörterbuchs gibt es nun auf zwei CD-Roms – Bequeme Recherchen. In: Thüringische Landeszeitung, 7. August 2004, S. 3.
- Paterno, Wolfgang: Hasenherz. Eselsscherz. 600 Millionen von Hand eingetippte Zeichen: die digitalisierte Form des gigantischen „Deutschen Wörterbuchs“. In: Profil, 31/04, 26. Juli 2004, S. 105.
- Patzer, Georg: Der edelste, ursprünglichste aller laute, aus brust und kehle voll erschallend. In: Stuttgarter Zeitung, Nr. 192, 20. August 2004, S. 8.

- Saltzwedel, Johannes: Ein Massiv wird zur Scheibe. In: SPIEGEL SPECIAL 3/2003, S. 64/65.
- Schiffhauer, Nils: Ein Nationalheiligtum der deutschen Sprache wird zu Bits. In: Frankfurter Allgemeine Zeitung, Nr. 251, 29. Oktober 2002, S. T1f.
- Schiffhauer, Nils: Grimm im Netz. In: Augsburger Allgemeine Zeitung, Nr. 3, 4. Januar 2003, S. V3.
- Schlag nach auf der CD. Grimms Wörterbuch erschienen. In: Bremer Nachrichten, 6. August 2004, S. 18.
- Schlag nach auf der CD. Grimms Wörterbuch erschienen. In: Weser Kurier, 6. August 2004, S. 18.
- Schröder, Lothar: Der ganze „Grimm“ auf nur zwei CD-Roms. In: Rheinische Post, 29. Juli 2004, S. A7.
- Schwarze, Dirk: Chinesen schrieben das Wörterbuch ab. „Von A bis Zypressenzweig“ – Das Grimmsche Wörterbuch gibt es auf zwei CD-Roms. In: Hessische/Niedersächsische Allgemeine, Nr. 165, 17. Juli 2004.
- Suhl, Nicole: Surfen durch 550 Jahre deutscher Sprachgeschichte. In: Lübecker Nachrichten, Nr. 285, 6. Dezember 2002, S. 21.
- Thormählen, Andrea: Surfen im Meer der Wörter. 300 Millionen Zeichen abgetippt: In China entstand der digitale Grimm – „Nationalheiligtum der deutschen Sprache“ auch auf CD. In: Pforzheimer Zeitung, Magazin, Nr. 67, 20. März 2004, S. 51.
- Vogel, Nikolai: Auf den unabsehbaren Strecken der Sprache. Dank fleißiger Typistinnen aus China: Das Grimmsche Wörterbuch gibt es jetzt auch digital. In: Süddeutsche Zeitung, Nr. 202, 1. September 2004, S. 14.
- Voigt, Anne: Jacob und Wilhelm Grimm: Grimms Deutsches Wörterbuch auf 2 CD-ROMs. Das „allerwichtigste Buch in deutscher Sprache“ endlich in elektronischer Version – der Digitale Grimm. www.daf-portal.de/ticker/detail.php?id=1226 (20. Juli 2004).
- Weinmann, Martin: Der jüngste Grimm. In: 150 Jahre Deutsches Wörterbuch. Kolloquium, veranstaltet von der Arbeitsstelle Deutsches Wörterbuch an der Berlin-Brandenburgischen Akademie der Wissenschaften am 8. Juli 2004. <http://150-grimm.bbaw.de/links/weinmann.pdf>.
- Widmann, Arno: Selten ist Globalisierung so schön. In: Berliner Zeitung vom 3. August 2004, <http://BerlinOnline.de/berliner-zeitung/politik/363548.html>.
- Wort-Schatz. www.dwb.uni-trier.de. In: C'T Magazin für Computertechnik 18 vom 23. August 2004.

b) **Rundfunkbeiträge:**

- Beitrag von Nils Schiffhauer im WDR 5 am 30. September 2002.
- Beitrag „Der digitalisierte Grimm“ von Nils Schiffhauer in der Deutschen Welle am 4. November 2002.

- Interview mit Thomas Burch zur Digitalisierung des DWB und zur *Documenta11*-Installation im SWR2 am 7. Oktober 2002.
- Interview mit Vera Hildenbrandt im SWR2 am 13. Mai 2003.
- Interview mit Thomas Burch zur Publikation des „Digitalen Grimm“ im Bayerischen Rundfunk am 9. Juli 2004.
- Beitrag im Schweizer Radio DRS am 21. Juli 2004. www.drs.ch/indexnetz.html, unter der Rubrik „im trend“.
- Interview mit Hans-Werner Bartz, „Das Grimmsche Wörterbuch auf 2 CD-Roms; abgetippt von fleißigen Chinesinnen“ in „Radio EINS – der Tag“ am 3. August 2004. www.radioeins.de/_/suche/gefunden/themen_jsp/key=themen_125849.html.
- Interview mit Kurt Gärtner und Jingning Tao fuer die Deutsche Welle, Redaktion China, am 9. August 2004.

Außerdem wurde der „Digitale Grimm“ am 24. August 2004 als besondere Empfehlung der Monate August und September in die Sachbücherliste der Süddeutschen Zeitung und des Norddeutschen Rundfunks aufgenommen.

c) Fernsehbeiträge:

- Beitrag von Ansgar Zender in Südwest3 am 10. August 2004.
- Beitrag im Lokalsender TrierPlus am 23. August 2004.

8 Perspektiven

Mit der elektronischen Publikation des DWB – online seit Juni 2002 und offline seit Juli 2004 – ist die Arbeit an den geschaffenen digitalen Ressourcen nicht zu Ende. Die Dynamik des Mediums erfordert Weiterentwicklungen und Verbesserungen, die vor allem in zwei Richtungen laufen: die Optimierung der Auszeichnung, um die Recherchemöglichkeiten erheblich zu erweitern, und der Aufbau bzw. die Erweiterung des digitalen Wörterbuchverbunds, um die Verknüpfungsmöglichkeiten mit dem DWB als zentraler Komponente umfassend zu nutzen.

Das digitale DWB ist prinzipiell offen für weitere Kodierungen, doch war schon bei der Projektplanung zu fragen und umsichtig zu erörtern, wie weit die Auszeichnung gehen sollte und ob der aus dem eingebrachten Markup erzielte Nutzen den zu betreibenden Aufwand rechtfertigt.

Die Optimierung für bereits jetzt ausgezeichnete Elemente kann beliebig verbessert werden, wenn die personellen Ressourcen zur Verfügung stehen. Denn trotz eingehender Datenanalyse und möglichst umsichtiger Programmierung ist wegen des enormen Textvolumens und der bekannten Heterogenität des DWB ein „perfektes“ Markup unerreichbar. Zwar können durch weitere sorgfältige Korrek-

turgänge fehlerhafte Kodierungen, die erst im Laufe intensiver Benutzung entdeckt werden, bereinigt werden, und so können bislang noch nicht oder nur teilweise ausgezeichnete Elemente mit den vorgesehenen TEI-Tags versehen werden. Doch beim DWB können in noch geringerem Umfang als bei dem Vorgängerprojekt, dem Mittelhochdeutschen Wörterbuchverbund, automatische Auszeichnungsroutinen, die auf typographisch-ausdrucksseitigen Merkmalen aufsetzen, entwickelt werden.

Wie schon bei dem Vorgängerprojekt wurden die Formteile der Artikel nicht weiter ausgezeichnet; die TEI hat keine entsprechenden Richtlinien dafür vorgesehen. Dennoch wäre eine Auszeichnung der in den Formteilen belegten Formvarianten für sprachwissenschaftliche Untersuchungen sehr nützlich, vor allem regionale Varietäten des Deutschen wie das Niederdeutsche würden dadurch gezielt recherchierbar gemacht. Eine Erweiterung der an gegenwartssprachlichen Wörterbüchern entwickelten Richtlinien im Hinblick auf die historischen und die Dialektwörterbücher wäre daher ein dringendes Desiderat. Eine stärkere deutsche Mitwirkung im TEI-Konsortium wäre im Hinblick auf das nach wie vor weltweite Interesse an der deutschen Sprache, ihrer Geschichte und ihrer Verwandtschaft mit den anderen germanischen Sprachen sehr willkommen.

Die Formteile des DWB enthalten außerdem eine Fülle von Informationen zur Etymologie und zahlreiche Belege aus den indogermanischen Sprachen. Eine weitergehende Auszeichnung wäre auch hier vor allem deshalb wünschenswert, weil das DWB gerade in seinen jüngsten Teilen als das umfangreichste etymologische Wörterbuch des Deutschen angesehen werden kann. Zur Zeit ist die perfekte Darstellung aller Sonderzeichen und Zeichen mit Diakritika allerdings noch ein Problem, auf das in diesem Bericht bereits eingegangen wurde. Die über 1200 Zeichen im DWB weitestgehend korrekt darzustellen und auch in der Online-Version zuverlässig abzubilden, gehört zu den Aufgaben, die in absehbarer Zeit durchzuführen sind, um die Perspektiven für die etymologischen Recherchen zu verbessern. Dazu gehörte vor allem – wie bereits in dem Vorgängerprojekt, dem Mittelhochdeutschen Wörterbuchverbund geschehen – die Auszeichnung der Sprachangaben und der zugehörigen Belege.

Immer wieder wird der Wunsch geäußert, den Wortschatz der Zitate in den Belegstellen der Wortartikel zu lemmatisieren. Das ist durchaus möglich, und es gibt bereits Lemmatisierungswerkzeuge für historische Sprachstufen (z. B. die für das Mittelhochdeutsche in Trier entwickelten), mit denen der hier auftretenden, enormen Varianz zum Teil Rechnung getragen werden könnte. Zur vollständigen Bewältigung dieser Aufgabe bedürfte es eines eigenen Projekts. Die für das digitale DWB entwickelten ausgefeilten Suchmöglichkeiten bieten jedoch fürs erste einen akzeptablen Ersatz und unterstützen das Auffinden der historischen Varianten eines Wortes in den Belegteilen recht effektiv.

Für die künftige Planung von Verbesserungen ist nicht allein die weitergehende Auszeichnung von Bedeutung, sondern die Verknüpfung des DWB mit anderen Wörterbüchern und die Ausgestaltung eines Wörterbuchverbundes, in dem das DWB eine Zentralfunktion hat, entscheidend. Die große Attraktivität des digitalen DWB, die durch die hohen Zugriffszahlen auf die Internetversion und die Verkaufszahlen der CD-ROM-Version nachdrücklich unterstrichen wird, bildet eine erfolgversprechende Grundlage dafür, das Interesse an der Geschichte des Deutschen und an seinen Dialekten zu wecken und zu fördern. Wer sich für die Geschichte eines Wortes und die damit verbundene Sache oder Vorstellung interessiert, kann sich am besten durch die Konsultation von Wörterbüchern darüber informieren. Ein digitaler Verbund von Wörterbüchern bietet Perspektiven, deren Reichweite und Faszination nach allem, was wir von den bisherigen Ansätzen wissen, nicht hoch genug eingeschätzt werden können. In Trier sind besonders gute Bedingungen für den Aufbau eines solchen Verbundsystems von Wörterbüchern gegeben: Zum einen wurden bereits Vorarbeiten für die Verknüpfung des DWB mit dem Mittelhochdeutschen Wörterbuchverbund geleistet, zum anderen werden im Rahmen des DFG-Projekts „Dialektwörterbuchverbund“ (www.dwv.uni-trier.de) zur Zeit die wichtigsten Dialektwörterbücher zum Westmitteldeutschen digitalisiert und miteinander verknüpft. Die Verknüpfung eines Teils des modernsten deutschen Dialektwörterbuchs, des *Pfälzischen Wörterbuchs*, mit den elektronischen Versionen des DWB und des Mittelhochdeutschen Handwörterbuchs von Matthias Lexer wurde bereits erprobt. Die Verknüpfung mit den weiteren Wörterbüchern (*Rheinisches Wörterbuch*, *Lothringisches Wörterbuch*, *Elsässisches Wörterbuch*) soll zu gegebener Zeit realisiert werden.

Die Arbeiten zum DFG-Projekt „Digitalisierung und elektronische Publikation des Goethe-Wörterbuchs im Internet“ sind ebenfalls in diesem Jahr angelaufen. Auch mit dieser Ressource würde sich, nicht zuletzt aufgrund der zahlreichen Goethe-Belege im DWB, eine Verlinkung anbieten. Weitere Perspektiven eröffnen sich aber auch im Hinblick auf Verknüpfungen mit den großen Enzyklopädien wie der *Oekonomischen Encyclopädie* von Johann Georg Krünitz (www.kruenitz.uni-trier.de) oder auch mit anderen Wörterbüchern, die, wie etwa das *Grammatisch-kritische Wörterbuch* von Adelung (BSB München), im Rahmen des Retrodigitalisierungsprogramms der DFG digitalisiert wurden.

Innerhalb der vom „Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier“ betreuten Projekte werden alle Wörterbücher unter einer einheitlichen Benutzeroberfläche zugänglich gemacht und mit offenen und dokumentierten Schnittstellen ausgestattet, ähnlich wie dies bei einem Großteil der digitalen Ressourcen des *Electronic Text Center* der University of Virginia der Fall ist. Doch ist die Anbindung anderer Wörterbücher, die unter ihren eigenen Benutzeroberflächen und ihren spezifischen Benutzungsaspekten zugänglich bleiben können, ebenso möglich, wie z. B.

die des *Deutschen Rechtswörterbuchs*, das bereits durchgängig auf die entsprechenden Artikel im DWB und im Mittelhochdeutschen Wörterbuchverbund verweist.

Von entscheidender Bedeutung wäre aber die Einbringung der abgeschlossenen Teile der Neubearbeitung des DWB in den Verbund. Die viel zu wenig bekannte und daher auch viel zu wenig benutzte Neubearbeitung umfaßt die am meisten veralteten Buchstabenstrecken A bis F und wird in wenigen Jahren abgeschlossen sein. Noch vor dem Abschluß sollten die bereits vorliegenden Bände mit der Erstbearbeitung verknüpft werden, ähnlich wie die Neubearbeitung des *Oxford English Dictionary* mit der Vorgängerauflage verbunden ist. Dieser elektronische Wörterbuchverbund aus dem alten und dem neuen DWB würde auf lange Zeit das maßgebende wordhistorische Grundlagenwerk für die deutsche Sprache bleiben. Mit der Digitalisierung der Erstbearbeitung auf der Basis modernster Datenstandards sind jetzt die optimalen Voraussetzungen geschaffen, um die Kernkomponente eines noch umfassenderen Verbundes, so wie er oben skizziert wurde, zu verwirklichen.