

URN: urn:nbn:de:kobv:b4-opus-24400

KURT GÄRTNER & RALF PLATE,  
Die Doppelfunktion des digitalen Textarchivs als Wörterbuchbasis  
und als Komponente der Online-Publikation. Am Beispiel des Mittel-  
hochdeutschen Wörterbuchs,

in:

*Perspektiven einer corpusbasierten historischen Linguistik und Philologie.  
Internationale Tagung des Akademienvorhabens „Altägyptisches Wörter-  
buch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften,  
12. – 13. Dezember 2011*, herausgegeben von Ingelore Hafemann,  
Berlin 2013, S. 193-204.

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Thesaurus Linguae Aegyptiae 4

Perspektiven einer corpusbasierten historischen Linguistik und  
Philologie. Internationale Tagung des Akademienvorhabens  
„Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen  
Akademie der Wissenschaften, 12. – 13. Dezember 2011

herausgegeben von Ingelore Hafemann

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

**Thesaurus Linguae Aegyptiae**

4

BERLIN 2013

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Perspektiven einer corpusbasierten historischen Linguistik  
und Philologie

Internationale Tagung des Akademienvorhabens „Altägyptisches  
Wörterbuch“ an der Berlin-Brandenburgischen Akademie der  
Wissenschaften, 12. – 13. Dezember 2011

herausgegeben von Ingelore Hafemann

BERLIN

2013

Dieser Band wurde durch die gemeinsame Wissenschaftskonferenz im Akademienprogramm mit Mitteln des Bundes (Bundesministerium für Bildung und Forschung) und des Landes Berlin (Senatsverwaltung für Wirtschaft, Technologie und Forschung) gefördert

Die Publikation unterliegt folgender Creative-Commons-Lizenz:  
„Namensnennung – Keine kommerzielle Nutzung – Weitergabe unter  
gleichen Bedingungen 3.0 Deutschland“

<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



URN: urn:nbn:de:kobv:b4-opus-24310

## INHALTSVERZEICHNIS

VORWORT	7
GREGORY CRANE & ALISON BABEU Global Editions and the Dialogue among Civilizations	11
<b>HISTORISCHE CORPUS-PROJEKTE – SYNCHRON UND DIACHRON</b>	
STÉPHANE POLIS & JEAN WINAND The Ramses project. Methodology and practices in the annotation of Late Egyptian Texts	81
SERGE ROSMORDUC The Ramses project in perspective. Managing evolving linguistic data	109
DIETER KURTH Das Edfu-Projekt. Ziel, Methode und Verarbeitung der lexikographischen Ergebnisse	121
INGELORE HAFEMANN & PETER DILS Der Thesaurus Linguae Aegyptiae – Konzepte und Perspektiven	127
GÜNTER VITTMANN Zur Arbeit an der Demotischen Textdatenbank: Textauswahl	145
GERNOT WILHELM Das Hethitologie Portal Mainz	155
JOST GIPPERT The TITUS Project. 25 years of corpus building in ancient languages	169
KURT GÄRTNER & RALF PLATE Die Doppelfunktion des digitalen Textarchivs als Wörterbuchbasis und als Komponente der Online-Publikation. Am Beispiel des Mittelhochdeutschen Wörterbuchs	193
HANS-CHRISTIAN SCHMITZ, BERNHARD SCHRÖDER & KLAUS-PETER WEGERA Das Bonner Frühneuhochdeutsch-Korpus und das Referenzkorpus ,Frühneuhochdeutsch‘	205

ALEXANDER GEYKEN Wege zu einem historischen Referenzkorpus des Deutschen: das Projekt Deutsches Textarchiv	221
BRYAN JURISH Canonicalizing the Deutsches Textarchiv	235
<b>WORTGESCHICHTE - TEXTGESCHICHTE - SPRACHGESCHICHTE: TRADITION UND INNOVATION BEI DER TEXTPRODUKTION</b>	
FRANK FEDER & SIMON D. SCHWEITZER Auf dem Weg zu einem integrierten Lexikon des Ägyptisch- Koptischen	245
FRIEDHELM HOFFMANN Die Demotische Wortliste – virtuell erweitert	263
GÜNTER VITTMANN Kursivhieratische Texte aus sprachlicher und onomastischer Sicht	269
MATHEW ALMOND, JOOST HAGEN, KATRIN JOHN, TONIO SEBASTIAN RICHTER & VINCENT WALTER Kontaktinduzierter Sprachwandel des Ägyptisch-Koptischen: Lehnwort-Lexikographie im Projekt Database and Dictionary of Greek Loanwords in Coptic (DDGLC)	283
THOMAS GLONING Historischer Wortgebrauch und Themengeschichte. Grundfragen, Corpora, Dokumentationsformen	317
LOUISE GESTERMANN Die altägyptischen Sargtexte in diachroner Überlieferung	371
THOMAS STÄDTLER Überlegungen zu Textsorte und Diskurstradition bei der Beschreibung von Textcorpora und ihr Bezug zur lexikographischen Forschung	385

## VORWORT

Die internationale Tagung „Perspektiven einer corpusbasierten historischen Linguistik und Philologie“ vom 12. – 13. Dezember 2011 am Akademienvorhaben „Altägyptisches Wörterbuch“ der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) war dem Thema des Aufbaus und der Nutzungsperspektiven elektronischer Textcorpora und Wörterbücher in den historischen Sprachen gewidmet. Die Teilnehmer, Vertreter der Ägyptologie, der Hethitologie, Indogermanistik sowie Referenten aus der historischen Lexikographie des Mittel- und Frühneuhochdeutschen und des Altfranzösischen diskutierten vor allem über die Veränderungen, die mit dem Einsatz elektronischer Erfassungs- und Verarbeitungsprozeduren einhergehen. Vertreter der Computerlinguistik vom „Zentrum Sprache“ der BBAW wurden in die Diskussionen einbezogen. Dort beschäftigt man sich seit Jahren mit dem Aufbau großer elektronischer Textcorpora (DWDS), darunter auch solcher, die historische Texte (DTA) für die elektronische Nutzung ermöglichen.

Die größte Herausforderung dieser neuen elektronischen Corpora und Wörterbücher ist es, sowohl den Methoden und damit den wissenschaftlichen Ansprüchen der traditionellen Philologie und Lexikographie unbedingt verpflichtet zu bleiben als auch neue Gebiete wie die Corpus- und Computerlinguistik für die historischen Sprachen zu öffnen. Die Teilnehmer haben gemeinsam und disziplinenübergreifend die Möglichkeiten und Grenzen der Datenerfassung, ihrer Präsentation und den Nutzen neuer Auswertungsprozeduren diskutiert.

Unter dem ersten Thema „Historische Corpusprojekte – synchron und diachron“ wurden elektronische Corpora vorgestellt und ein intensiver Austausch darüber geführt, welche Datenstrukturen die linguistischen Inhalte in adäquater Weise abbilden. Wichtig war die Frage, auf welche Resonanz diese elektronischen Corpora bei den Nutzern gestoßen sind und welche Erwartungen und Anforderungen aus den verschiedenen Fachdisziplinen an die Projekte herangetragen werden. Der Austausch über Nutzungsperspektiven elektronischer Corpora schloss auch die Diskussion über die Erarbeitung projektübergreifend einsetzbarer Standards der Codierung und Strukturierung historischer Textdaten mit ein. Hinsichtlich einer mittel- und langfristigen Nutzbarkeit sowie einer langfristigen Datensicherheit stehen solche Fragen zunehmend im Focus und einige aktuelle Initiativen dazu wurden vorgestellt. Spezielle technische Aspekte



elektronischer Datenerfassung und automatischer Analyse- und Speicherungsverfahren elektronischer Textdaten konnten am letzten Tag als ein Themenschwerpunkt mit den Programmierern diskutiert werden.

Ein zweiter Schwerpunkt waren konkrete Fragstellungen aus der historischen Lexikographie und diachronen Textanalyse. Für das Ägyptische ist der diachrone Ansatz auf Grund der über vier-tausendjährigen Textüberlieferung von großer Relevanz. Themen wie historischer und/oder textgattungsspezifischer Wortgebrauch, die Erarbeitung diachroner Wortlisten und Aspekte des kontaktindizierten Sprachwandels konnten disziplinübergreifend zwischen den Ägyptologen und den Kollegen der historischen Lexikographie des Mittel- und Frühneuhochdeutschen und des Altfranzösischen behandelt werden.

Mit dem Abendreferenten Gregory Crane, dem Begründer der „Perseus Digital Library“, wurde ein breites Publikum angesprochen. In seinem Vortrag hat er noch einmal die hohe Relevanz und die neuen Möglichkeiten der Einbeziehung zahlreicher Wissenschaftler und einer interessierten Öffentlichkeit in die Projektarbeit demonstriert, die das Internet auf völlig neue Weise eröffnet hat. Die Herausgeberin ist sehr froh, seinen programmatischen Beitrag zu diesem Thema, dessen schriftliche Form er gemeinsam mit Alison Babeu erarbeitet hat, ebenfalls in diesem Band präsentieren zu können.

Wir danken der Berlin-Brandenburgischen Akademie der Wissenschaften für die umfassende Unterstützung unserer Projektarbeit und ganz speziell der Vorbereitung dieser Konferenz sowie der Möglichkeit, die Akten auf dem E-Doc-Server der Akademie veröffentlichen zu können.

Der Hermann und Elise geborene Heckmann Wentzel-Stiftung sei hiermit ausdrücklich für die unbürokratische und großzügige finanzielle Unterstützung dieser erfolgreichen Tagung gedankt.

Das Akademienvorhaben „Altägyptisches Wörterbuch“ konnte sich als aktives Mitglied des Weiteren auf das „Zentrum Grundlagenforschung Alte Welt“ stützen, dem alle altertumswissenschaftlichen Vorhaben der BBAW angehören. Dem Zentrum ist es zu danken, dass der Abendvortrag von Gregory Crane einem breiteren Publikum dargeboten werden konnte.

Allen Autoren dankt die Herausgeberin für ihre anregenden Diskussionen und die qualitätvollen Beiträge in diesem Band.

Auf eine Gesamtbibliographie wurde verzichtet und die Abkürzungen der in den ägyptologischen Beiträgen erwähnten Zeitschriften und Reihen folgen dem Lexikon der Ägyptologie, herausgegeben von Wolfgang Helck und Wolfhart Westendorf, Band VII: Nachträge, Korrekturen, Indices, Wiesbaden 1992, XIV-XIX.

Ganz besonders sei schließlich Frau Angela Böhme für die gewissenhafte redaktionelle Bearbeitung der Manuskripte gedankt sowie Dr. Simon Schweitzer für seine Hilfe beim Erstellen des Layouts.

Berlin, Mai 2013

Ingelore Hafemann



DIE DOPPELFUNKTION DES DIGITALEN TEXTARCHIVS ALS  
WÖRTERBUCHBASIS UND ALS KOMPONENTE DER ONLINE-  
PUBLIKATION  
AM BEISPIEL DES MITTELHOCHDEUTSCHEN WÖRTERBUCHS

KURT GÄRTNER & RALF PLATE

Abstract

Auf dem Hintergrund des umfassenden EDV-Einsatzes, der die vorbereitende Materialbereitstellung seit 1986 (*Findebuch zum mittelhochdeutschen Wortschatz*) bzw. 1994 (für das neue *Mittelhochdeutsche Wörterbuch* selbst) ebenso wie die Ausarbeitung und Publikation des neuen *Mittelhochdeutschen Wörterbuchs* seit 2006 kennzeichnet, beleuchtet der Beitrag den Gewinn von digitalen Textcorpora für Macher und Nutzer von historischen Belegwörterbüchern: Den Lexikographen verschafft ein umfangreiches digitales Textarchiv und daraus durch halbautomatische Lemmatisierung gewonnenes Belegarchiv, das in einem Redaktionssystem für die Artikelarbeit bereitgestellt wird, größere und leichtere Übersicht über den historischen Sprachgebrauch und entlastet sie von zeitraubenden Exzerptions- und Korrekturarbeiten; den Wörterbuchbenutzern wird es durch die Verknüpfung der Belegzitate bzw. Belegstellenangaben mit den digitalisierten und im Online-Angebot zur Verfügung gestellten Volltexten der Wörterbuchquellen ermöglicht, den Quellenbezug der lexikographischen Befunde zu rekontextualisieren. Aus der Erfahrung eines Vierteljahrhunderts der Vorbereitung und Ausarbeitung eines großen digitalen Belegwörterbuchs werben die Autoren für die Bereitstellung von umfassenden digitalisierten Textcorpora (einschließlich der Retrodigitalisierung der älteren lexikographischen Hilfsmittel) für die historische Sprachforschung selbst wie für die Nutzer ihrer Forschungsergebnisse. Eine umfassende Textdigitalisierung von historischen Sprachquellen des Deutschen steckt aber leider (im Gegensatz zur Bilddigitalisierung von Handschriften und gedruckten Büchern) trotz großer Anstrengungen und überzeugender Ergebnisse in einzelnen Vorhaben immer noch in den Anfängen oder wird zum Teil unter zu engen Gesichtspunkten betrieben.

1

Das *Mittelhochdeutsche Wörterbuch* (MWB) soll den mittelhochdeutschen Wortschatz in seinen zeitlichen Grenzen von ca. 1050 bis ca. 1350 beschreiben. Die Beschreibung des deutschen Wortschatzes in den Zeiträumen davor und danach erfolgt durch das von Theodor Frings und Elisabeth Karg-Gasterstädt begründete *Althochdeutsche Wörterbuch* (AWB) und das von Oskar Reichmann begründete *Frühneuhochdeutsche Wörterbuch* (FWB). Wie AWB und FWB ist das MWB

ein Belegwörterbuch;<sup>1</sup> sein Ziel ist eine repräsentative Darstellung des mittelhochdeutschen Wortschatzes in seiner sprachsystematischen Gliederung und seiner zeitlichen, räumlichen und textsortenspezifischen Erstreckung. Die Vorarbeiten zum MWB bis zum Einsetzen der Artikelarbeit wurden von 1994 bis 1999 durch die DFG gefördert, zum 1.1.2000 erfolgte die Weiterförderung im Rahmen des Akademienprogramms.

Bei der Planung des MWB war von vornherein vorgesehen, dass auf allen Stufen der Wörterbucharbeit die EDV eingesetzt werden sollte.<sup>2</sup> Zum Wörterbuchplan seien zunächst einleitend einige kurze Erläuterungen gegeben. Dieser sah als zentrale Einheit eine digitalisierte Quellensammlung bzw. ein digitales Textarchiv mit einem auf Zuwachs angelegten Corpus von retro-digitalisierten Volltexten vor, das die Basis für ein digitales Belegarchiv bildete, das aus dem Textarchiv durch Verfahren der halbautomatischen Lemmatisierung gewonnen wurde. Unter den weiteren digital nutzbaren Hilfsmitteln sind als wichtigste Ergänzung des Quellencorpus und des Belegarchivs die Vorgängerwörterbücher zu nennen. Jedes neue Wörterbuch hat lexikographische Vorfahren, auf denen es aufbaut und die daher einen wesentlichen Teil der Wörterbuchbasis bilden. Das gilt für Wörterbücher vergangener Sprachepochen in viel höherem Maße als für Wörterbücher zur Gegenwartssprache. Mit DFG-Unterstützung haben wir daher in Trier von 1997 an die alten mittelhochdeutschen Wörterbücher zusammen mit ihren Quellenverzeichnissen maschinenlesbar gemacht und miteinander verknüpft. Es entstand so ein digitaler Wörterbuchverbund, der außer den Quellenverzeichnissen vier lexikographische Hilfsmittel vereinigt:

1) Das erste und bis heute nicht ersetzte Wörterbuch zum Mittelhochdeutschen wurde von 1854 bis 1866 mit Benutzung des Nachlasses von Georg Friedrich Benecke durch Wilhelm Müller und Friedrich Zarncke ausgearbeitet. Der Benecke-Müller-Zarncke (= BMZ) ist ein exzellentes Wörterbuch, aber nicht leicht zu benutzen,

---

<sup>1</sup> Vgl. dazu ausführlich PLATE, R., Das Mittelhochdeutsche Wörterbuch: Beleglexikographische Konzeption, EDV, Vernetzungspotentiale, in: *Lexicographica* 23 (2007), 77-95.

<sup>2</sup> Zur Erprobung des EDV-Konzepts vgl. ausführlich PLATE, R. & U. RECKER, Elektronische Materialgrundlage und computergestützte Ausarbeitung eines historischen Belegwörterbuchs. Erfahrungen und Perspektiven am Beispiel des neuen Mittelhochdeutschen Wörterbuchs, in: LEMBERG, I. et. al. (Hrsg.), *Chancen und Perspektiven computergestützter Lexikographie*, *Lexicographica*; Series Maior 107, Tübingen 2001, 155-177.

denn er ist nach Wortstämmen angeordnet, unter denen alle Glieder einer Wortfamilie erscheinen. 2) Um den BMZ besser benutzbar zu machen, arbeitete Matthias Lexer von 1872 bis 1878 einen alphabetischen Index aus, der zugleich als Supplement und Handwörterbuch auf der Basis des BMZ fungieren sollte. 3) Da schon während der kurzen Ausarbeitungszeit des Mittelhochdeutschen Handwörterbuchs zahlreiche neue Quellen zum Mittelhochdeutschen erschlossen wurden, ließ Lexer 1878 zusammen mit dem dritten Band seines Wörterbuchs umfangreiche Nachträge erscheinen, die das dritte Wörterbuch des digitalen Verbunds bilden. 4) Das vierte Werk ist das sog. *Findebuch zum mittelhochdeutschen Wortschatz*, das an der Universität Trier zwischen 1986 und 1992 ausgearbeitet wurde. Es vereinigt 106 Glossare zu den wichtigsten nach Lexer, d.h. nach 1878 erschienen Textausgaben des Mittelhochdeutschen. Die vier Komponenten wurden miteinander und mit dem Quellenverzeichnis von Eberhard Nellmann zu BMZ und Lexer sowie dem Quellenverzeichnis zum *Findebuch* verknüpft und vielfältig recherchierbar gemacht. Man kann nun also ein Stichwort in allen vier Komponenten, wie Abb. 1 am Beispiel von *abbet* zeigt, nebeneinander „aufschlagen“ und vergleichen.

Der Ausbau des digitalen mittelhochdeutschen Textarchivs, auf dessen Bedeutung für die Ausarbeitung des MWB wie für seine Nutzung im Internet im Folgenden genauer eingegangen wird, wurde 2001 bis 2003 in einem deutsch-amerikanischen Kooperationsprojekt mit DFG/NSF-Förderung vorangetrieben; erfasst wurden vor allem die über 100 *Findebuch*-Quellen, die den größten Zuwachs an neuem Wortschatz (rund 8000 Lexeme) lieferten. Wieder mit DFG-Förderung wurde 2002 bis 2005 ein Artikelredaktionssystem entwickelt, das mit Beginn der Artikelarbeit zur Verfügung stand.



auch für elektronische Nutzung bereitgestellt werden sollte; die Publikation des eben vorgestellten Verbunds der digitalisierten Vorgängerwörterbücher auf CD und im Internet war der erste Schritt dazu, die Bereitstellung der wachsenden digitalen Belegsammlung für das neue Wörterbuch in einem Online-Angebot seit dem Jahr 2000 der zweite.

Seit 2006 wird nun das neue Wörterbuch selbst lieferungsweise im Druck und nach einer kurzen Schutzfrist auch elektronisch publiziert. Bisher liegen vier Doppellieferungen vor, die 2006, 2007, 2009 und Ende 2011 erschienen sind und auf 2224 Spalten die Wortstrecke *a* bis *évrouwe* enthalten. In einer um Druckfehler bereinigten Fassung werden sie zusammen mit einem zusammengefassten und überarbeiteten Quellenverzeichnis (das das Verzeichnis der ersten Lieferung mit den drei Nachträgen der Lieferungen 2-4 vereinigt) und mit einem Vorwort als Band 1 voraussichtlich Ende 2012 erscheinen.

Aus den Erfahrungen bei der Vorbereitung, der Ausarbeitung und der Publikation des neuen Wörterbuchs soll heute zu der Leitfrage dieser Tagung Stellung genommen werden, also zur Frage nach den Nutzungsperspektiven elektronischer Textcorpora und Wörterbücher, und zwar gerade in ihrem wechselseitigen Verhältnis: Welche Rolle spielt für uns das elektronische Textcorpus bei der Ausarbeitung des Wörterbuchs, und welche Rolle spielt es dann anschließend für die Nutzer der elektronischen Publikation des Wörterbuchs?

Zur Frage der Rolle des elektronischen Textcorpus für die Ausarbeitung des Wörterbuchs seien zunächst einige Zahlen zum Stand der elektronischen Materialbasis vorgelegt, die nach Erscheinen der dritten Doppellieferung erhoben wurden.<sup>4</sup> Die Stichwortliste für das Wörterbuch umfasst insgesamt rd. 84.000 Artikelkandidaten, die wir zu Projektbeginn aus den Vorgängerwörterbüchern des 19. Jahrhunderts und dem *Findebuch* kompiliert haben; diese Liste bildete dann später auch die Grundlage für die Vernetzung der Vorgängerwörterbücher untereinander, die oben unter 1 beschrieben worden ist, und ebenso auch für die Anbindung dieses Wörterbuchverbunds an das neue Wörterbuch und seine Materialien im Online-Angebot des MWB, das unten vorgestellt wird. Tatsächlich artikelwertig dürften von diesen 84.000 Kandidaten für das neue Wörterbuch nach den bisherigen Erfahrungen nur etwa zwei Drittel sein, das MWB

---

<sup>4</sup> PLATE, R., Das Mittelhochdeutsche Wörterbuch. Bearbeitungsstand 2009, Erfahrungen und Perspektiven, in: SCHMID, H. U. (Hrsg.), *Perspektiven der germanistischen Sprachgeschichtsforschung*, Jahrbuch für germanistische Sprachgeschichte 1, Berlin / New York 2010, 254-268, hier 256-259.



wird demnach also rd. 60.000 Wortartikel enthalten. Die Gründe für die gegenüber den Vorgängerwörterbüchern geringere Artikelzahl sind ein engerer Quellenzeitraum und abweichende Lemmatisierungsprinzipien.

Die zweite Hauptkomponente der elektronischen Materialbasis ist die digitalisierte Quellensammlung. Sie besteht zurzeit aus rd. 210 Texten mit zusammen rd. 7 Mio. laufenden Wortformen. Aus diesen Texten ist vor allem durch Verfahren halbautomatischer Lemmatisierung, ergänzend auch durch Lemmatisierung von Einzelstellen, das elektronische Belegarchiv erhoben worden. Es umfasst (nach Ausarbeitung der dritten Doppellieferung) rd. 1.425.000 Textbelege, die sich auf rund 27.000 Artikelkandidaten verteilen. Diese Zahlen besagen, dass trotz umfassender Digitalisierungs- und Lemmatisierungsanstrengungen durch das Projekt selbst immer noch nur rund die Hälfte – nämlich 27.000 von voraussichtlich knapp 60.000 artikelwertigen Wortschatzeinheiten – im digitalen Belegarchiv vertreten sind. Entsprechendes gilt dann wiederum für die Ebene der einzelnen Wortschatzeinheit und dem Belegmaterial, das für die Beschreibung ihres Gebrauchs – Bedeutungen, Wendungen, syntaktisches Verhalten – nötig ist. Die dazu für die ersten drei Doppellieferungen (wie Anm. 2, S. 257) erhobenen Zahlen ergeben, dass sich vom Gesamt der in einer gedruckten Doppellieferung zitierten oder mit einer Stellenangabe vertretenen Belege nur rd. die Hälfte im elektronischen Belegarchiv der Wörterbuchdatenbank fanden, während die anderen nachexzerpiert werden mussten. Zwar ist der Anteil des elektronischen Belegarchivs kontinuierlich gesteigert worden von 40 % in der ersten Doppellieferung bis auf 54 % in der dritten Doppellieferung, und in der gerade im Druck erschienenen vierten Doppellieferung ist er abermals leicht erhöht worden, aber mindestens ein Drittel der Belege stammt auch in ihr immer noch aus Nachexzerption, die zu einem großen Teil auf den Nachweisen in den Vorgängerwörterbüchern beruht.

Die<sup>5</sup> nachexzerpierten Belege haben gegenüber den Datenbankbelegen auch ganz abgesehen von dem Aufwand, der für ihre Erhebung bei laufender Artikellarbeit getrieben werden muss, verschiedene Nachteile, von denen hier nur der gravierendste für Lexikographen wie für Benutzer des Wörterbuchs hervorgehoben werden soll: Die Exzerpte haben keinen Volltext hinter sich, den man als Artikelbearbeiter im Redaktionssystem oder als Benutzer der Online-Fassung

---

<sup>5</sup> Dieser Absatz ist unverändert wiederholt aus PLATE (wie Anm. 4), 258f.

des Wörterbuchs aufrufen könnte, um den Textzusammenhang über den exzerpierten Ausschnitt hinaus zu prüfen. – Hierzu sei eine Seitenbemerkung gestattet, die von allgemeinerem methodischen Interesse für die historische Linguistik ist. Als wir in den 1990er Jahren damit begannen, das elektronische Text- und Belegarchiv für das neue Wörterbuch einzurichten, wurde die Diskussion über den nötigen Umfang stark von der Befürchtung einiger Beteiligter bestimmt, dass schnell Belegmengen entstehen könnten, die nicht mehr bearbeitbar wären und geeignet, das eigentliche Vorhaben, die Ausarbeitung des Wörterbuchs, lahmzulegen. Dies betrifft tatsächlich jedoch nur eine kleine Gruppe von höchstfrequenten Lexemen; für sie müssen bei der halbautomatischen Lemmatisierung geeignete Filter oder Sperren eingebaut werden. Der größte Teil des Wortschatzes dagegen bedarf einer sehr großen Textmenge, um angemessen vertreten zu sein in der Belegsammlung. Der Aufbau einer umfassenden elektronischen Quellensammlung für die deutsche Sprachgeschichte ist aber leider immer noch ein dringendes Desiderat; es sollte heute nicht mehr Aufgabe einzelner Projekte sein, sich die elektronischen Corpora für ihre Untersuchungszwecke selbst zu erarbeiten.

Die Leitfrage nach den Nutzungsperspektiven elektronischer Corpora ist also soeben für die Wörterbuchmacher wie folgt beantwortet worden: Der Nutzen ist enorm, man sollte keine Kosten und Mühen scheuen, solche Corpora für die historische Linguistik bereitzustellen, und man sollte sich dabei vor allem nicht von zu engen Gesichtspunkten leiten lassen. Bereits gestreift worden ist eben auch schon die Perspektive der Wörterbuchnutzer, also von uns allen. Zu dieser Perspektive jetzt abschließend einige Anmerkungen am Beispiel unseres Internetangebots MWB Online.

Auf MWB Online wird nicht nur das Wörterbuch selbst für elektronische Nutzung bereitgestellt, sondern zusätzlich auch die elektronischen Materialien, die für die Ausarbeitung genutzt worden sind bzw. für die weitere Ausarbeitung bereitstehen. Zentraler Einstiegsort in das Online-Angebot ist die Komponente Lemmaliste / Belegarchiv (vgl. Abb. 2). Sie enthält die gesamte Stichwortliste des Vorhabens, also auch die Stichwörter, die nur in den alten Wörterbüchern mit einem Artikel bzw. mit einem eigenen Artikel vertreten sind. In der Lemmaliste wird auf ökonomische Weise angezeigt, zu welchen Stichwörtern es Belegmaterial gibt und wie viele Belege, außerdem, ob es dazu Artikel im neuen und / oder in den alten mittelhochdeutschen Wörterbüchern gibt. Am Beispiel des Stichworts *abbet* ‚Abt, Vorsteher eines Klosters‘: Dazu findet sich in Klammern eine

Häufigkeitsangabe (26), die die Zahl der Stellen im elektronischen Belegarchiv anzeigt, anschließend folgen die Siglen MWB und MWV für das neue Wörterbuch und den Verbund der alten Wörterbücher im Internet. Die Häufigkeitsangabe und die beiden Siglen sind als Hyperlinks gestaltet, über die man in die entsprechenden Komponenten gelangt.

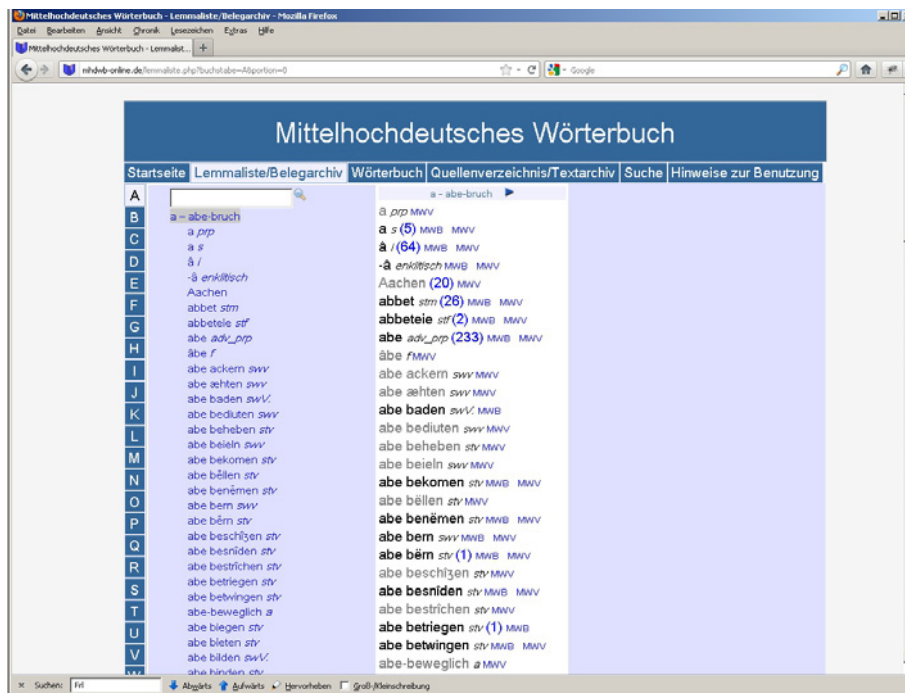


Abb. 2: MWB Online: Komponente Lemmaliste/Belegarchiv

Abb. 3 zeigt die Ausgabe des elektronischen Belegarchivs zum Stichwort *abbet*, die bereitgestellt wird, wenn man auf die eingeklammerte Frequenzangabe klickt. Wenn man dagegen dem Link in den Artikel selbst des neuen Wörterbuchs folgt, erhält man die Ausgabe, die in Abb. 4b abgebildet ist; dem elektronischen Artikel gegenübergestellt ist dort in Abb. 4a die Satzausgabe des Artikels im gedruckten Wörterbuch.

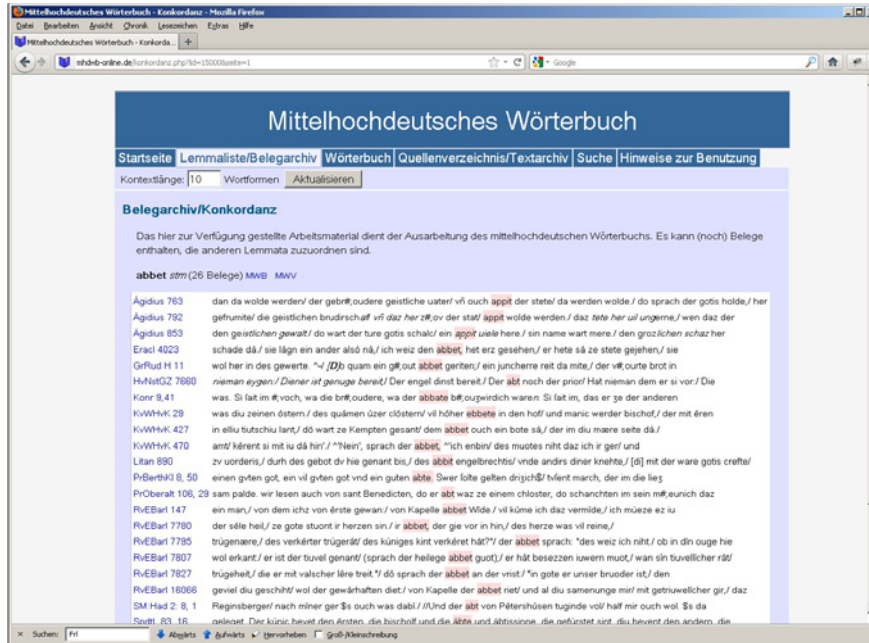


Abb. 3: MWB Online: Komponente Belegarchiv, Konkordanz der Belege für abbet

**abbet** *stfM*. Pl. eb(βe)te; auch abbāt [ːslāt] UvZ-LANZ 3864, abbate KONR 9,41, abte [Akk.] PR-BERTHKL 8,50; *verschoben* -p(p)- (zu den Schreibformen s. WMU 1,13); lat. abbas, -atis, *ahd.* abbat (vgl. *Etymol. Wb. d. Ahd.* 1,19f.). 'Abt, Vorsteher einer Mönchsgemeinschaft' der gebirg.ouere geistliche uater / vnd ouch appit der stete ÄGIDIUS 763; wir lesen auch von sant Benedicten, do er abt waz ze einem chloster PROBERALT 106,29; der meister über die münche was, / daz was ein wizzic abbas UVZLANZ 3832; ich han getan gehorsam ir, / reht als ein munch eim apte tut MINNEB 1887; *sprichwörtl.* swā der apt die wüfel treit, / spilnt dā die münche daz ist niht ein wunder MARNER S. 160 (vgl. TPMA 1,17f.). – als hoher Würden- und Herrschaftsträger: die äbte und äbtissinne, die gefürstet [in den Fürstenstand erhoben] sint SPÖTL 83,16; eines tages keiser Ott reit / mit den fürsten [...] kurzwillen an daz velt, / [...] der abt von Vult neben im reit ENIKWCHR 27811; des quāmen über clōstern / vil höher ebbete in den hof / und manic werder bischof KVWHVK 29; nur die man infel tragen sach, / bischof, ebt und cardinal OTTOK 13525. – Gebrauch als Titel in Verbindung mit Namens- und/oder Ortsangabe: von Kapelle abbet Wide RVEBARI 147; gedenke [...] zv vorderis [...] des abbit Engelbrechtis LITAN 890; der appet sante Brandan RVEWCHR 3060; in der Anrede mit her(re): der herzog Albreht [...] sprach: 'nū verjeht, / her abt, waz iu werre.' OTTOK 36064

Abb. 4a: Artikel abbet (Druckfassung)

*Etymol. Wb. d. Ahd.* 1,1f.; *umfangr. mhd. Belegammlung von I. V. Zingerle, Germania 7 (1862), S. 257-267 (älteste Belege dort „gegen das Ende des 12. Jh.s“; S. 266)*

**abbet** *stfM*. Pl. eb(βe)te; auch abbāt [ːslāt] UvZLANZ 3864, abbate KONR 9,41, abte [Akk.] PR-BERTHKL 8,50; *verschoben* -p(p)- (zu den Schreibformen s. WMU 1,13); lat. abbas, -atis, *ahd.* abbat (vgl. *Etymol. Wb. d. Ahd.* 1,19f.). 'Abt, Vorsteher einer Mönchsgemeinschaft' der gebirg.ouere geistliche uater / vnd ouch appit der stete ÄGIDIUS 763; wir lesen auch von sant Benedicten, do er abt waz ze einem chloster PROBERALT 106,29; der meister über die münche was, / daz was ein wizzic abbas UVZLANZ 3832; ich han getan gehorsam ir, / reht als ein munch eim apte tut MINNEB 1887; *sprichwörtl.* swā der apt die wüfel treit, / spilnt dā die münche daz ist niht ein wunder MARNER S. 160 (vgl. TPMA 1,17f.). – als hoher Würden- und Herrschaftsträger: die äbte und äbtissinne, die gefürstet [in den Fürstenstand erhoben] sint SPÖTL 83,16; eines tages keiser Ott reit / mit den fürsten [...] kurzwillen an daz velt, / [...] der abt von Vult neben im reit ENIKWCHR 27811; des quāmen über clōstern / vil höher ebbete in den hof / und manic werder bischof KVWHVK 29; nur die man infel tragen sach, / bischof, ebt und cardinal OTTOK 13525. – Gebrauch als Titel in Verbindung mit Namens- und/oder Ortsangabe: von Kapelle abbet Wide RVEBARI 147; gedenke [...] zv vorderis [...] des abbit Engelbrechtis LITAN 890; der appet sante Brandan RVEWCHR 3060; in der Anrede mit her(re): der herzog Albreht [...] sprach: 'nū verjeht, / her abt, waz iu werre.' OTTOK 36064

abbeteie, abbette, abdie, abtei *stf*; auch abbeteie, abbate, abbacie, appiteie, aptie; *vereinzelt sw.*, mit Uml. epigen (Dat. Sg.) PLGERF 9822. Neben Formen mit dem Suffix -eie, die zuerst in Hss. des 12./13. Jh.s im SÜMMHERR belegt sind, stehen solche mit dem

Abb. 4b: MWB Online: Artikel abbet (Online-Version)

Hier soll weder inhaltlich auf den Artikel eingegangen werden, noch sollen die Vorzüge und Nachteile der beiden Darstellungen diskutiert werden. Worauf es ankommt, sind die weiterführenden Benutzungsmöglichkeiten des Bildschirmartikels. Sie sind angedeutet in den farblich hervorgehobenen Links, die sich bei Quellensiglen und bei Stellenangaben finden. Die Quellensiglen sind durchgehend mit Links versehen, die zu den bibliographischen Angaben im Quellenverzeichnis führen. Die Stellenangaben sind nur dann als Hyperlinks gestaltet, wenn es sich um Belege aus den Quellen des elektronischen Textarchivs handelt. Durch Anklicken der Stellenangabe kann dann eine elektronische Version der Textausgabe an der betreffenden Textstelle aufgeschlagen werden.

Dies ist in Abb. 5 für den Beleg aus Konrads von Würzburg Vers-Erzählung ‚Heinrich von Kempten‘ vorgeführt, der sich im gedruckten Artikel in Zeile 20-22 findet. Es handelt sich um eine Volltextanzeige, in der man beliebig weit nach vorne oder hinten zurückblättern bzw. -rollen kann, um so die Interpretation eines Textbelegs, wie sie der Wörterbuchartikel ausdrücklich vornimmt oder durch den Artikelzusammenhang zu verstehen gibt, im ursprünglichen Textzusammenhang zu prüfen oder vertiefend nachzuvollziehen.

Zusammenfassend lässt sich zur Leitfrage der Tagung festhalten: Der Nutzen elektronischer Textcorpora für die Ausarbeitung von historisch-philologischen Wörterbüchern ist evident, auch wenn diese Einsicht – wenigstens für die historische Lexikographie des Deutschen – noch nicht hinreichend in die Tat umgesetzt worden ist. Mindestens ebenso sehr müsste die breite elektronische Fundierung von Wörterbüchern, aber auch von anderen Hilfsmitteln der historischen Linguistik, eine Forderung der Nutzer dieser Hilfsmittel sein, wie der Kurzdurchgang durch MWB Online zeigen sollte.

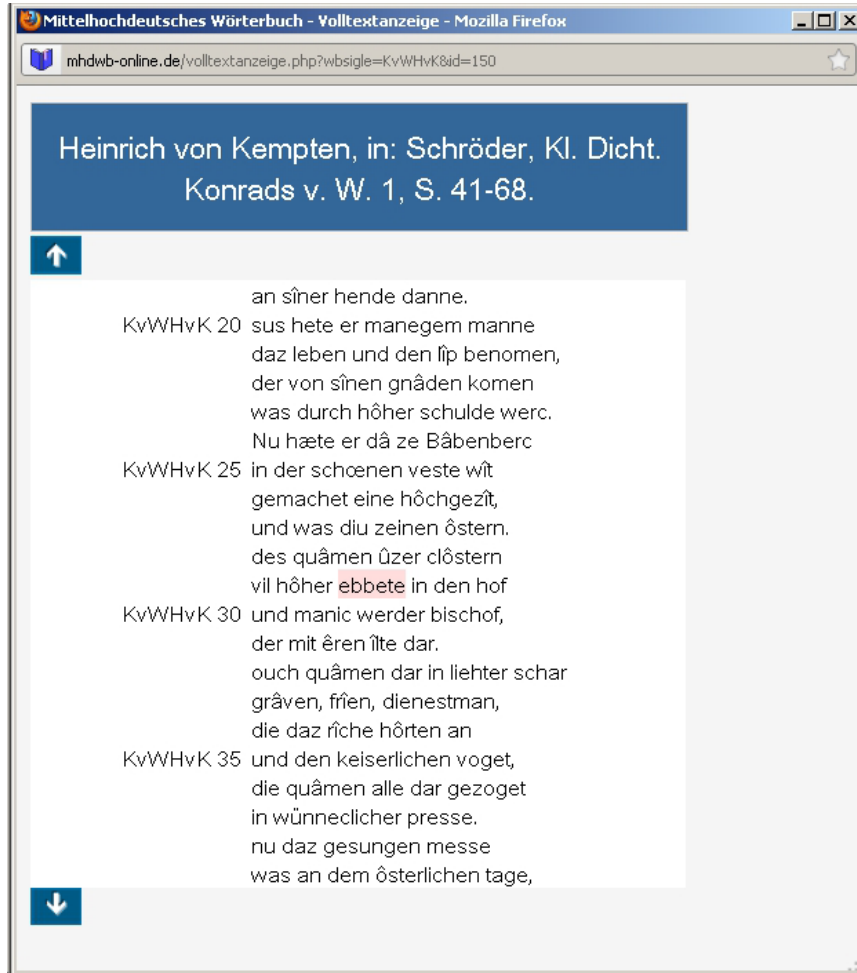


Abb. 5: MWB Online: Volltextanzeige eines Wörterbuchbelegs (Konrad von Würzburg ‚Heinrich von Kempten‘ v. 29)

Netzadressen der erwähnten Hilfsmittel (dort auch die bibliographischen Angaben zu den Druckwerken) und von Seiten mit weiterführenden Informationen:

<http://www.MWV.uni-trier.de/> (Die Vorgängerwörterbücher im Mhd. Wörterbuchverbund [MWV] online).

<http://www.mhdwb-online.de/> (Online-Angebot des MWB).

<http://www.mwb.uni-trier.de/> (Homepage der Trierer Arbeitsstelle des MWB).

<http://www.zfda.de/beitrag.php?id=782&mode=maphilinet> (Überblicksartikel zur Online-Lexikographie des mittelalterlichen Deutsch, aus ZfdA 138,1 [2009]).

[http://www.uni-trier.de/fileadmin/forschung/maw/MWB/Plate/Digitale\\_Editionen.pdf](http://www.uni-trier.de/fileadmin/forschung/maw/MWB/Plate/Digitale_Editionen.pdf)  
(Referat über die E-Texte des MWB; Abstract:  
[http://www.telota.de/telota/nachrichten/abstracts\\_edworkshop/plate](http://www.telota.de/telota/nachrichten/abstracts_edworkshop/plate)).