

Brom, Vlastimil

Die Mittelhochdeutsche Begriffsdatenbank als ein vielseitiges Arbeitsinstrument zur Analyse älterer deutschsprachiger Texte

Brünner Beiträge zur Germanistik und Nordistik. 2019, vol. 33, iss. Supplementum, pp. 173-184

ISSN 1803-7380 (print); ISSN 2336-4408 (online)

Stable URL (DOI): <https://doi.org/10.5817/BBGN2019-S-13>

Stable URL (handle): <https://hdl.handle.net/11222.digilib/142280>

License: [CC BY-SA 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/)

Access Date: 28. 11. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

Die Mittelhochdeutsche Begriffsdatenbank als ein vielseitiges Arbeitsinstrument zur Analyse älterer deutschsprachiger Texte

Middle High German Conceptual Database. A flexible tool for analysing older German Texts

Vlastimil Brom

Abstract

The paper presents the Middle High German Conceptual Database as a tool and research infrastructure for analysing older German texts; the available functionality and typical use cases are discussed, including the conceptual framework and the representation of semantic features of the lexical units. The present options for retrieving this kind of data are outlined and some currently running projects are briefly characterised, which are aiming at enhancing the current functionality and enabling other kinds usage in multiple areas.

Keywords

Middle High German; philology; text analysis; Middle High German Conceptual Database; Middle Ages

Geschichte, Anliegen und Konzeption der MhdBDB

Eines der Pionier-Projekte des EDV-Einsatzes und im weiteren Sinne der „digital humanities“ im Kontext der Altgermanistik stellt die Mittelhochdeutsche Begriffsdatenbank (MhdBDB) dar.¹ Im vorliegenden Beitrag soll diese begrifflich-lexikographisch sowie textanalytisch orientierte Infrastruktur zusammenfassend vorgestellt werden. Eine Grundlage stellt hier eine angestrebte weitergehende (v.a. lexikalisch-semantic orientierte) Untersuchung von Texten, die vorher philologisch erschlossen wurden – der deutschen Reimübersetzung der sog. Dalimil-Chronik – *Tutsch kronik von Behem lant* (TKR), des zusammenhängenden annalistischen Abrisses (TKA),² der deutschen Übersetzung der Autobiographie Karls IV. Vita Caroli (VTC) und der deutschen Übersetzung der Pulkava-Chronik.³ Die folgenden Überlegungen gehen im Wesentlichen auf die unternommenen Arbeiten im Zusammenhang mit der Vorbereitung dieser Texte für die MhdBDB, die anschließenden Schritte der Lemmatisierung und Disambiguierung sowie die Recherchen anhand der bereitgestellten textuellen Daten zurück. Ausgewählte Aspekte wurden bereits in einer Art explorativen Teiluntersuchungen diskutiert und mittels Stichproben demonstriert;⁴ im Folgenden sollen nun auch die konzeptuellen und technischen sowie institutionellen und forschungsgeschichtlichen Grundlagen der MhdBDB eingehender angesprochen werden.

Dieses Langzeitprojekt wurde bereits 1992 durch einen Zusammenschluss von zwei spezialisierten Vorhaben ins Leben gerufen, nämlich *Namen in deutschen literarischen Texten des Mittelalters* und *Begriffswörterbuch der mittelhochdeutschen Literatur*, die ihrerseits noch jeweils eigene Vorgeschichte hatten;⁵ die Arbeitsstelle ist gegenwärtig ein Teil des Interdisziplinären Zentrums für Mittelalter und Frühneuzeit (IZMF) der Universität Salz-

1 Den Mitarbeitern der Mittelhochdeutschen Begriffsdatenbank, namentlich insbesondere Klaus M. Schmidt, Margarete Springeth, Katharina Zepezauer-Wachauer, Nikolaus Morocutti, Daniel Schlager und Peter Hinkelmanns, gehört mein großer Dank für ihr Entgegenkommen und Hilfsbereitschaft und ihre vielseitige Unterstützung sowohl bei der Neuaufnahme der Texte als auch bei anschließenden Analysen.

2 Beide Werke ediert in Brom, Vlastimil (hrsg.) (2009): *Di tutsch kronik von Behem lant. Die gereimte deutsche Übersetzung der alttschechischen Dalimil-Chronik. Rýmovaný německý překlad staročeské Dalimilovy kroniky*, Brno: MU.

3 Beide Texte in einer Online-Fassung bzw. Transkription bereitgestellt: Brom, Vlastimil (hrsg.) (2019): *Spätmittelalterliche deutsche historiographische Texte böhmischer Provenienz – philologische Analyse, elektronische Edition – Texte*. <https://www.phil.muni.cz/german/projekty/hmb/e-text-hmb-de.htm> (3. 9. 2019).

4 Vgl.: Brom, Vlastimil (2015): *Möglichkeiten der semantischen Auswertung von spätmittelalterlichen historiographischen Texten. Stichproben zum Themenbereich ‚Herrschaft‘*. In: Kusová, Jana – Malechová, Magdalena – Vodrážková, Lenka (hrsg.): *Deutsch ohne Grenzen. Linguistik*. Brno: Tribun EU, S. 51–67 (hier S. 53–54). <http://publikationen.uni-frankfurt.de/frontdoor/index/index/docId/39643> (12. 3. 2016). – Brom, Vlastimil (2017): *Zur Begrifflichkeit und sprachlichen Erfassung von ‚Herrschaft‘ als einer der zentralen konzeptuellen Domänen der historiographischen Werke des Mittelalters*. In: Kotůlková, Veronika – Rykalová, Gabriela (hrsg.): *Zentrum und Peripherie. Aus sprachwissenschaftlicher Sicht*. Opava: Slezská univerzita v Opavě, S. 323–343. <http://publikationen.uni-frankfurt.de/frontdoor/index/index/docId/46981> (18. 9. 2019).

5 Vgl. Hinkelmanns, Peter – Zepezauer-Wachauer, Katharina: *ez ist ein wârheit, niht ein spel, daz netze was sinewel. Die MHDDB im Semantic Web (im Druck)*. Tagung: *Digitale Methoden und Objekte in Forschung und Vermittlung der mediävistischen Disziplinen*. Tagung der Bamberger Graduiertenschule für Mittelalterstudien. Bamberg, 8.–10. November 2018.

burg.⁶ Das primäre Anliegen stellt die Erfassung des mittelhochdeutschen Wortschatzes dar, vornehmlich auf der semantisch-begrifflichen Ebene, natürlich mit Einbeziehung des Namenschatzes. Den Kern der Textbasis bildeten zunächst klassische literarische Werke der mittelhochdeutschen Epoche; mit der fortschreitenden Erweiterung der Textgrundlage wurden auch spätmittelhochdeutsche und frühneuhochdeutsche Werke einbezogen, einschließlich Gebrauchstexte u.a. Nach dem Datenbestand zum Zeitpunkt der herangezogenen Abfragen handelt es sich um 666 Texte mit 10656175 laufenden Wörtern (bei einem unterschiedlichen Bearbeitungsstand im Hinblick auf Lemmatisierung und Disambiguierung); diese gehören zu 279998 Wortformen (einschließlich rein graphische Varianten) und werden zu 42083 Lemmata zugeordnet, bei denen 57982 Bedeutungen unterschieden werden;⁷ es sind aktuell 556 Begriffskategorien im Gebrauch.⁸ (Die während der vorhergehenden Arbeiten primär hochgeladenen und annotierten vier historiographischen Werke umfassen in der Datenbank zusammen etwa 131449 laufende Wortformen: Pulkava-Chronik (Sigle PUC) 73076, Tutsch kronik (TKR) 41667, Abriß (TKA) 2771, Vita Caroli (VTC) 13935, d.h. insgesamt ungefähr 1,23 % des Umfangs.)

Explizit formulierte Anliegen der MhdBDB sind in den erreichbaren eigenen Stellungnahmen eher knapp gehalten, die Vorstellung der Datenbank bietet jedoch eine treffende Darlegung ihrer Kernfunktionalität: „Sie ermöglicht den Zugriff auf die wichtigsten Werke der mittelhochdeutschen Dichtung von den verschiedensten Blickwinkeln aus über eine äußerst vielseitige Suchfunktionalität.“⁹ Die Bestimmung der Vorläufer- und Frühformen der MhdBDB lässt sich im bestimmten Ausmaß der einschlägigen Fachbibliographie der beteiligten Mitarbeiter, insbesondere des Urhebers Klaus Maximilian Schmidt entnehmen:¹⁰ Zu den wichtigen mediävistischen Themen aus diesem Kontext gehört die Erstellung und Auswertung der Indizes und Begriffsglossare zu mehreren mittelhochdeutschen literarischen Werken (wobei spätestens seit den 1980er Jahren von der gerade erst zugänglich gewordenen Computertechnik intensiv Gebrauch gemacht wurde). Seit 1995 besteht eine Internet-Präsenz die durch allmähliche inhaltliche sowie technologische Fortentwicklung die heutigen Datenbestände und Funktionalität erreichte. Im aktuellen Projekt des sog. Relaunch der MhdBDB seit 2016 steht eine vielseitige Modernisierung, Standardisierung der Datenformate und eine Funktionserweiterung im Vordergrund.¹¹

6 Vgl. MhdBDB: Mittelhochdeutsche Literatur auf Knopfdruck. http://mhdadb.sbg.ac.at:8300/wordpress/mittelhochdeutsche_literatur_auf_knopfdruck_de/ (11. 9. 2019).

7 MhdBDB – Statistik. <http://mhdadb.sbg.ac.at/LastStatistics.de.html> (20. 9. 2019), Stand zum 9. 9. 2019; Text List. <http://mhdadb.sbg.ac.at/mhdadb/App?action=TextList> (20. 9. 2019).

8 Die hierarchisch gegliederte Übersicht der Kategorien ist auf der einschlägigen Seite der Datenbank zu finden: MhdBDB – Begriffssystem. <http://mhdadb.sbg.ac.at/mhdadb/App?action=BrowseCategory> (20. 9. 2019).

9 Vgl. Die mittelhochdeutsche Begriffsdatenbank (MHDDB). Mittelhochdeutsche Literatur auf Knopfdruck, Abschnitt: Über die MHDDB. http://mhdadb.sbg.ac.at:8300/wordpress/mittelhochdeutsche_literatur_auf_knopfdruck_de/ (11. 9. 2019).

10 Vgl. em. Univ.-Prof. Dr. Klaus M. Schmidt – MhdBDB. <http://mhdadb.sbg.ac.at:8300/wordpress/klaus-m-schmidt/> (11. 9. 2019). – Klaus Maximilian Schmidt – ResearchGate. https://www.researchgate.net/profile/Klaus_Maximilian_Schmidt (11. 9. 2019).

11 Vgl. Die mittelhochdeutsche Begriffsdatenbank (MHDDB). Mittelhochdeutsche Literatur auf Knopf-

Zugriffsmöglichkeiten und Abfragetypen in der MhdBDB

Die Benutzerschnittstelle bietet zwei wichtigste Perspektiven bzw. Herangehensweisen an, denen zwei Abfrage-Modi der Datenbank entsprechen: die Textstellensuche¹² und die verallgemeinerte Lemma-Anzeige¹³. Bei der letztgenannten Ansicht stehen bei jedem Lemma-Eintrag die Angaben über zugeordnete Flexionsformen und vorhandene lautlich-graphische Varianten (einschließlich einfache Wortklassenbestimmung¹⁴ und – optional – die Belegzahlen), bei Wortbildungskomplexen, sind auch die Komponenten verzeichnet. Ferner steht hier die semantische Annotation – die Zuordnung von Begriffskategorien (u.U. als eine Art Seme aufzufassen), eventuell differenziert in mehrere Bedeutungen (vergleichbar mit Sememen); auf die Beschaffenheit des Begriffssystems wird hier weiter noch näher eingegangen. Falls das jeweilige Lemma in den online verfügbaren Wörterbüchern des Mittelhochdeutschen erfasst ist, wird die entsprechende Seite verlinkt.¹⁵ Zusätzlich sind bei den einschlägigen Lemmata noch weitere externe Datenbanken verlinkt (Pflanzen des Mittelalters,¹⁶ Bildserver des IMAREAL¹⁷). Die aktuellen Bestrebungen, für diese inhaltlich benachbarten Daten umfassendere einheitliche Auswertungsmöglichkeiten zu eröffnen, stehen im Zentrum des Kooperationsprojektes ONAMA (Ontology of the Narratives of the Middle Ages),¹⁸ von dem ein gemeinsamer Beschreibungsrahmen (in Form einer Semantic-Web-Ontologie und zusammenhängender Technologien) entwickelt und bereitgestellt werden soll.

Bei der Textsuche nach Belegstellen stehen in der MhdBDB detaillierte Suchoptionen zur Verfügung: Auswahl der Einzeltexte oder Textgruppen, Suchkontext, Behandlung der Lemmata und Wortvarianten u.a. Die Anzeige der Suchergebnisse enthält auch die jeweilige Textumgebung in einstellbarer Breite (v.a. aus urheberrechtlichen Gründen werden die Volltexte nicht direkt angezeigt).

druck, Abschnitt: Relaunch der MHDDB. http://mhdadb.sbg.ac.at:8300/wordpress/mittelhochdeutsche_literatur_auf_knopfdruck_de/ (11. 9. 2019).

12 MhdBDB – Textsuche nach Belegstellen. <http://mhdadb.sbg.ac.at/mhdadb/App?action=TextQueryModule> (20. 9. 2019).

13 MhdBDB – Wortindex. <http://mhdadb.sbg.ac.at/mhdadb/App?action=DicSelect> (20. 9. 2019).

14 Die Wortklassen sind in der aktuellen Form nicht kontextsensitiv sondern nur type-bezogen, ggf. mit mehrfachen Zuordnungen ohne Differenzierung; einen Ansatz zu einer morphologischen Annotation mit Token-Differenzierung stellt der eigens für diese Daten entworfene POS Tagger dar; vgl. Echelmeyer, Nora – Reiter, Nils – Schulz, Sarah (2017): Ein PoS-Tagger für „das“ Mittelhochdeutsche. In: Digital Humanities im deutschsprachigen Raum (Dhd) 2017. Konferenzabstracts, S. 141–147. https://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/MHD_Tagger/paper.pdf (24. 9. 2019); eigenständige Implementation online: POS Tagger für mittelhochdeutsche Texte: <http://clarin05.ims.uni-stuttgart.de/mhdtt/index.html> (24. 9. 2019).

15 Wörterbuchnetz – Mittelhochdeutsches Handwörterbuch von Matthias Lexer. <http://woerterbuchnetz.de/Lexer/> (20. 9. 2019) – im Verbund mit weiteren mittelhochdeutschen Wörterbüchern.

16 Sprache: Mittelhochdeutsch – MPS – Repository. <http://medieval-plants.org/mps-daten/languages/gmh/> (20. 9. 2019).

17 IMAREAL, Institut für Realienkunde des Mittelalters und der frühen Neuzeit <http://www.imareal.sbg.ac.at/home/> (3. 9. 2019).

18 ONAMA – Ontologie der Narrative des Mittelalters. <http://onama.sbg.ac.at/> (3. 9. 2019).

Die angeführten Informationen der Wortdatenbank (Lemma mit Formvarianten einschließlich Wortklassen; Semem mit Begriffskategorien/Semen) sind mit den einzelnen laufenden Wörtern der Textdatenbank assoziiert. Beim Neueinlesen der Texte werden die einzelnen Wortformen mit den gespeicherten Einträgen verglichen, und die eindeutig identifizierten Textwörter werden den entsprechenden Lemmata zugeordnet; bei nicht erkannten oder homographen Formen ist eine manuelle Lemmatisierung erforderlich, genauso wird bei den polysemen Lemmata eine Disambiguierung vorgenommen. Bei der Neuerfassung der Texte (oder auch nachträglich beim Editieren) können diese Daten geändert werden, etwa wenn neue Homographie oder Bedeutungsvarianten identifiziert werden, oder eine Verfeinerung der begrifflichen Bestimmungen angestrebt wird u.a.

Die Textdatenbank enthält neben Wörterbuch-Verknüpfungen auch textstellenbezogene Daten, z.B. die Textgliederung (Bücher, Kapitel, Verseinheiten, handschriftliche Folien, Satzgrenzen u.a.). Zusätzlich werden Informationen über die Einzeltexte erfasst, d.h. Daten zu den Autoren, Gattungen, zur Gliederung und Nummerierung, ferner die zu Grunde liegenden Editionen bzw. Handschriften, die etwaigen Spezifika der Erfassung u.a.¹⁹

Für die Abfragen in beiden Modi stehen vergleichbare wortbezogene Informationen zur Verfügung – neben Wortgestalt (Lemma, Wortform/graphische Variante, Wortklasse) können auch die (im Weiteren näher beschriebenen) Begriffskategorien abgefragt werden sowie eine Kombination von diesen beiden Ebenen.

Dank dem Verständnis und Entgegenkommen der Mitarbeiter der MhdBDB konnte bei der Arbeit zusätzlich auch noch eine weitere, intern verfügbare Funktionalität genutzt werden, nämlich direkte SQL-Abfragen von „rohen“ Daten, die einige formalisierte Vergleiche innerhalb der gesamten Textbasis möglich machten.

Von größtem Nutzen sind für die sprachlichen Analysen die Funktionalitäten der Mittelhochdeutschen Begriffsdatenbank im Bereich der Lemmatisierung und Disambiguierung der Texte und selbstverständlich das zu Grunde liegende Begriffssystem, das bereits beim Entwurf die Besonderheiten der historischen Sprachstufen und der mittelalterlichen „Realien“ im weitesten Sinne zu berücksichtigen suchte.

Das Begriffssystem der MhdBDB im Kontext der onomasiologischen Wörterbücher

Durch diesen Ansatz stellt sich das Vorhaben in die Tradition der onomasiologischen lexikographischen Ansätze, die im gewissen Sinne eigentlich bis in die frühesten Epochen der Schriftlichkeit bzw. der Lexikographie zurückreicht.²⁰ Man kann die onomasiologische

19 Vgl. MhdBDB – Text List. <http://mhdadb.sbg.ac.at/mhdadb/App?action=TextList> (11. 3. 2017) – in den einzelnen Teilseiten, verlinkt von den Textsiglen und Autorennamen.

20 Vgl. Wiegand, Herbert Ernst (2004): Lexikographisch-historische Einführung. In: Dornseiff, Franz (Begr.) (2004): Der deutsche Wortschatz nach Sachgruppen. 8., völlig neu bearbeitete und mit einem vollständigen alphabetischen Zugriffsregister versehene Auflage von Uwe Quasthoff. Mit einer lexikographisch-historischen

Perspektive als besonders prominent in der MhdBDB betrachten, gerade dadurch hebt sie sich am deutlichsten von anderen auf ältere Sprachstufen konzentrierten lexikographischen Projekten ab.

Es ist jedoch festzuhalten, dass die im Wortindex präsentierten Daten auch eine – zumindest grobe – semasiologisch verwertbare Annotation bereitstellen, nämlich die einem Lemma jeweils zugeordneten Begriffskategorien; eine klassische, wiederum primär semasiologische, lexikographische Erschließung des Wortschatzes steht in den bereits erwähnten verlinkten externen Wörterbüchern zur Verfügung; die Vernetzung mit den Realiendatenbanken verdeutlicht hingegen wieder die onomasiologischen Anliegen und Auswertungsmöglichkeiten der MhdBDB. Ferner spielen die korpusorientierten Anforderungen eine wichtige Rolle, insbesondere das Bestreben nach einer (im Idealfall) vollständigen Lemmatisierung und Disambiguierung, die neben ausgeprägten Autosemantika auch die Synsemantika bzw. die Funktionswörter einschließen, die in den erfassten Texten vertreten sind.

Einen üblichen Fall stellt eine Mehrfachzuordnung von Begriffskategorien zu einzelnen Lemmata bzw. Bedeutungsvarianten dar; bei diesem Ansatz sind somit auch verhältnismäßig detaillierte und differenzierte semantische Annotationen durch Kombination von Begriffskategorien möglich.

Die Konzeption des Begriffssystems lässt eine Inspiration durch klassische nach Sachgruppen strukturierte Thesauri erkennen, in allgemeiner Nachfolge von Rogets Thesaurus²¹, ein konkreteres Vorbild liegt in dem von Hallig und Wartburg entworfenen Begriffssystem vor.²² Hervorzuheben sind die praktischen Anliegen der vorgenommenen begrifflichen Erfassung des Wortschatzes – dem Begriffssystem wird Werkzeugcharakter zugeschrieben²³, wobei die Erprobung der Tragfähigkeit dieser Arbeitsgrundlage [in dem wohl vorgesehenen lexikographischen Einsatz – Anm. VB] als notwendig angesehen wird.²⁴ Dabei sind aber auch Spuren der lange andauernden kontroversen

Einführung und einer ausgewählten Bibliographie zur Lexikographie und Onomasiologie von Herbert Ernst Wiegand. Berlin – New York: Walter de Gruyter. E-Book (2010). <https://www.degruyter.com/viewbooktoc/product/61548> 2010 (31. 7. 2019), S. 13*– 91* (hier S. 68*, 81*–91*).

21 Roget, Peter Mark (1852): *Thesaurus of English Words and Phrases*. Classified and arranged so as to facilitate the Expression of Ideas and assist in Literary Composition (1. Aufl.). London 1852 (ferner in zahlreichen Neuauflagen und verschiedenen Bearbeitungen). – Zit. nach Wiegand [Anm. 20], S. 31; vgl. Springeth, Margarete (2009): *Der analytische Weg ist das Ziel: Die Mittelhochdeutsche Begriffsdatenbank als Online-Textarchiv*. In: Hofmeister, Wernfried – Hofmeister-Winter, Andrea (Hrsg.): *Wege zum Text. Über die Verfügbarkeit mediävistischer Editionen im 21. Jahrhundert*. Grazer Kolloquium 17.–19. September 2008. Tübingen: Niemeyer (Beihefte zu *Editio*, Bd. 30). S. 185–202 (hier S. 194).

22 Hallig, Rudolf – Wartburg, Walther von (1963): *Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas*. 2. Aufl. Berlin: Akademie-Verlag. Vgl. Schmidt, Klaus M. (1988): *Der Beitrag der begriffsorientierten Lexikographie zur systematischen Erfassung von Sprachwandel und das Begriffswörterbuch zur mittelhochdeutschen Epik*. In: Bachofer, Wolfgang (Hrsg.): *Mittelhochdeutsches Wörterbuch in der Diskussion (Germanistische Linguistik 84)*. Tübingen: Niemeyer, S. 35–49 (hier S. 40). – Schmidt, Klaus M. (1993): *Begriffsglossar zu Ulrichs von Zatzikhoven *Lanzelet**. Tübingen: Niemeyer. (Indices zur deutschen Literatur 25), S. VIII.

23 Hallig, Rudolf – Wartburg, Walther von (1963): *Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas*. 2. Aufl. Berlin: Akademie-Verlag, S. 22.

24 Hallig – von Wartburg [Anm. 23], S. 22.

Diskussionen über die „wissenschaftlichere“ Betrachtung und Darstellung des Wortschatzes (im Kontrast zu inhärent „unwissenschaftlichen“, äußerlichen alphabetischen Wörterbuchanordnung) bemerkbar. So wird bereits in der programmatischen Einführung von Wartburg die Notwendigkeit angesprochen „bei der Abfassung von wissenschaftlichen deskriptiven Wörterbüchern die alphabetische Anordnung aufzugeben...“.²⁵ Für das als Ersatz der alphabetischen Anordnung entworfene Begriffssystem wird (auch bei gelegentlichen Relativierungen) fast eine Art universalistischer Geltungsanspruch erhofft, vgl.:

„Wir haben versucht, alle Seinsbereiche zu berücksichtigen und allen kulturellen und zivilisatorischen Verhältnissen der Gegenwart Rechnung zu tragen. Diese Gegenwartsbedingtheit schließt aber, wie wir glauben, die Möglichkeit nicht aus, dieses Begriffssystem auch für die Darstellung des Wortschatzes früherer Epochen zu verwenden. [...] Wir glauben, mit dem Begriffssystem, das wir hier vorlegen, für die Lexikographie eine Grundlage geschaffen zu haben, welche die Darstellung des ‚Wortschatzes als Gesamtgefüge‘ ermöglichen könnte, unabhängig davon, welcher Sprache, welcher Mundart oder welcher Epoche dieses Wortgut angehört.“²⁶

In der metalexikographischen Diskussion gilt die angesprochene Kontroverse um bzw. gegen den „unwissenschaftlichen“ Charakter der alphabetischen Anordnung der Wörterbücher inzwischen als irrelevant bzw. historisch, wie Wiegand in seinem umfassenden Überblick darlegt.²⁷ Als Kriterium für Wertungen des jeweiligen Ordnungsansatzes gilt demnach die Effizienz im Hinblick auf die Erfüllung der vorgesehen konkreten Wörterbuchfunktionen.

In der Umsetzung des Begriffssystems durch die MhdBDB treten die forschungspraktischen Anliegen verständlicherweise deutlicher in den Vordergrund als in dem oben erwähnten konzeptuellen Entwurf von Hallig und Wartburg (bei dem jedoch auch eine Art Probeanwendung am französischen Wortschatz geliefert wird)²⁸. Man kann sagen, dass die vorliegende Form der MhdBDB als eines flexiblen Informationssystems auch einige als kontrovers angesehene Aspekte weniger relevant erscheinen lässt: So spielt das vorgegebene Ordnungsschema keine so große Rolle wie bei Printwörterbüchern, deren äußere Gestaltung und deren (nächstliegenden) Zugriffswege davon deutlich geprägt sind. Die Datenbank mit einer Web-Schnittstelle stellt hingegen mehrere Zugriffsmöglichkeiten bereit, die als gleichwertig gelten können und je nach Fragestellung frei gewählt werden können (neben der Suche anhand von Begriffskategorien v.a. die Textstellsuche, Wortform- bzw. Lemmasuche im Wortindex u.a.). Durch die erwähnte übliche Mehrfachzuordnung von Begriffskategorien zu Lemmata wird zugleich die Bindung an die konkret vorgegebene hierarchische Strukturierung abgeschwächt – im Hinblick auf Abfragen oder in der semantischen Annotation der Lemmata können die Kategorien

25 Hallig – von Wartburg [Anm. 23], S. 52.

26 Hallig – von Wartburg [Anm. 23], S. 75.

27 Wiegand [Anm. 20], S. 68*.

28 Vgl. die „Durchführung des Begriffssystems“ in: Hallig – von Wartburg [Anm. 23], S. 113–229, der noch ein Wortregister folgt, S. 230–315.

nebeneinanderstehen; die Hierarchie bleibt so primär bei der Arbeit mit dem Begriffssystem selbst von Bedeutung.

Die oberen Gliederungseinheiten des Begriffssystems entsprechen weitgehend dem Ordnungsschema von Hallig und Wartburg;²⁹ eine Erweiterung³⁰ stellt lediglich die Aufnahme der synsemantischen Ausdrücke, die verallgemeinernd als „Funktionswörter“ (9000000–...) zusammengefasst bzw. markiert sind:

A	Universum/Welt		10000000
	I	Himmel/Atmosphäre/Himmelskörper	11000000
	II	Planet Erde	12000000
	III	Vegetation/Pflanzen	13000000
	IV	Tierwelt	14000000
B	Mensch		20000000
	I	Mensch/Körperliches Wesen	21000000
	II	Geist/Seele	22000000
	III	Mensch als soziales Wesen	23000000
	IV	Sozialer Aufbau	24000000
	V	Kunst/Kunstgewerbe	25000000
C	Mensch und Welt		30000000
	I	A Priori	31000000
	II	Wissenschaft und Technik	32000000
X	Funktionswörter		90000000
	I	Nicht bestimmbare Formen	91000000
	II	Nhd. Kommentare	92000000
	III	Abkürzungen	93000000

Begriffssystem der MhdBDB – obere Gliederungseinheiten³¹

Praktische Vorteile der Arbeit mit der MhdBDB ergeben sich aus jahrelanger stetiger Erweiterung der Textbasis um neue Texte, wobei neue lexikalische Einheiten bzw. Wortvarianten einbezogen werden; zugleich werden stufenweise auch neue Sachgebiete erschlossen, bzw. werden bei vorhandenen Kategorien weitergehende Präzisierungen vorgenommen, was in die Datenbank fließend eingeht. Dadurch kann eine flexible Repräsentation auch bei Sachbereichen gewährleistet werden, die erst mit neu aufgenommenen Texten erschlossen werden.³²

Die hierarchisch strukturierten Begriffskategorien sind mit numerischen Werten assoziiert, die bei den Datenbank-Abfragen (sowohl im Wortindex als auch bei der Ab-

29 Vgl. Hallig – von Wartburg [Anm. 23], S. 101–112.

30 In Vorbereitung ist ferner die Einführung der eigenständigen Kategorie „Namen“ (40000000–...), die im aktuell vorliegenden Schema in dieser Form nicht enthalten ist.

31 MhdBDB – Begriffssystem. <http://mhdadb.sbg.ac.at/mhdadb/App?action=BrowseCategory> [Komplette Übersicht der Einzelkategorien vgl.: <http://mhdadb.sbg.ac.at/mhdadb/App?action=BrowseCategory&cat=0&depth=10&all=%3E%3E%3E>] (19. 9. 2019).

32 Vgl. Schmidt, Klaus M. (1980): Begriffsglossare zu Ulrich von Lichtenstein (2 Bände). Tübingen: Niemeyer (Indices zur deutschen Literatur 14/15), S. VII.

frage von Textstellen) benutzt werden können. Die semantische Beschreibungsebene mittels Begriffskategorien wird den Lemma-Einträgen in der Datenbank zugeordnet, häufiger werden noch mehrere Bedeutungsvarianten (als Sememe aufzufassen) innerhalb von Einzelemmata unterschieden. Dabei können die Begriffskategorien für einzelne Sachgebiete im Allgemeinen auch als eine Art Seme fungieren. Diese Nutzung von Begriffskategorien als kombinierbaren Elementen der semantischen Annotation ist wahrscheinlich als eine Innovation der MhdBDB anzusehen; in dem zu Grunde gelegten Begriffssystem von Hallig und Wartburg ist eine solche Anwendung (anhand der verfügbaren Darlegungen) nicht vorgesehen und eigentlich sind da die Auffassung und Perspektive unterschiedlich: Hier werden die Lexeme als Einheiten zu (einem oder ggf. mehreren) „Begriffen“³³ nach der wahrgenommenen inhaltlichen Zugehörigkeit zugeordnet; in der MhdBDB wird in vielen Fällen de facto die lexikalische Bedeutung einzelner Lemmata (oder deren einzelnen Sememe) durch eine Kombination von zugeordneten begrifflichen Kategorien annotiert (je nach Einzelfall mehr oder weniger exakt oder eher approximativ). Aus nahe liegenden Gründen ist die Beschreibung primär auf die ausgeprägten Autosemantika konzentriert, die Erfassung von Funktionswörtern und semantisch vagen Wörtern ist in der Regel eingeschränkt und wenig spezifiziert; immerhin werden jedoch auch solche Einheiten erfasst und klassifiziert (z.B. als die bereits erwähnten Funktionswörter u.a.). Einige Lemmata bzw. Sememe sind allerdings auch mit genau einer Begriffskategorie verbunden; hierbei lassen sich mehrere spezifische Gruppen unterscheiden: Besonders häufig sind da Eigennamen vieler Typen, die i.d.R. gerade nur als solche klassifiziert werden, wozu die jeweilige Einzelkategorie in konkreten Fällen ausreicht, vgl. z.B. *Aachen* (Stadt/Namen – 24212000).³⁴ Eine sehr ähnliche Situation ist bei detaillierten Gliederungen wie bei Pflanzen- oder Tierbezeichnungen, z.B. *buoche* (Waldbäume/Flurbäume – 13022000) festzustellen. Es gibt aber auch weitere Appellativa, zu deren plausiblen semantischen Annotation eine Begriffskategorie im Wesentlichen genügt – z.B. dank einer geringeren Spezifiziertheit oder der semantischen Nähe zum betreffenden Begriff, vgl. eines der üblichen Sememe von *vluz* (Binnengewässer 12021000), oder die Zahlwörter, z.B. *drîe* (Kardinalzahlen – 31422000). Interessanterweise gibt es dabei relativ wenige Fälle, wo in den betreffenden Bezeichnungen auch identische Morpheme bzw. Etyma vorliegen, z.B. *dorf* (Gemeinde/Ortschaft/Dorf – 24220000). Zu belegen sind ferner bedeutungsarme Ausdrücke – Partikel, Konjunktionen u.a. für die eine Art Restkategorie vorgesehen ist, z.B.: *doch*, *nein*, *und*, *obe* u.a. (Funktionswörter – 90000000); dabei sollte aber nochmals betont werden, dass auch diese Ausdrücke durch zusätzliche semantische Kategorien näher spezifiziert werden, wie es auch in den angeführten Beispielen der Fall ist.

33 Diese Begriffe als Schlüsseleinheiten des als Ordnungsschema entworfenen Begriffssystems sind freilich durch abstrahierend, metasprachlich gebrauchte sprachliche Ausdrücke bezeichnet; vgl. ferner die informelle Wendung „... der ‚geistige Gegenstand‘, der Begriff, der durch diese Bezeichnung in meinem Bewusstsein festgehalten wird, ...“ [Anführungszeichen im Original], Hallig – von Wartburg [Anm. 23], S. 11.

34 In den hier und weiter angeführten Beispielen handelt es sich um Einträge des Wortindexes, abrufbar nach Lemma-Formen, vgl.: MhdBDB – Wortindex – Suchformular. <http://mhdadb.sbg.ac.at/mhdadb/App?action=DicSelect> (12. 9. 2019).

Die Begriffskategorien, die in größerer Anzahl bei den Sememen auftreten, werden als einfache Mengen aufgefasst, d.h. ohne formale Hierarchie oder explizit markierte Strukturbeziehungen (dabei spielt gewiss die praktische Handhabbarkeit dieser Merkmale bei den Abfragen eine wichtige Rolle). Während die fehlende Hierarchie in einigen Fällen eine weniger übersichtliche semantische Annotation zur Folge hat (z.B. bei den an sich durchsichtigen Wortbildungskonstrukten), kann der so entstehende erweiterte Spielraum in der Verarbeitung der semantischen Annotation u.U. vorteilhaft werden – je nach Zielsetzung und Herangehensweise.

Zu beachten ist ferner der Umstand, dass die einzelnen Begriffskategorien oft nicht als gleich „gewichtig“ anzusehen sind für die Kodierung der Lemma- bzw. Semembeutung; eine solche Gewichtung kann aber bei der Aufzählung der Kategorien nicht explizit vermittelt werden. So ist z.B. die Begriffskategorie Autorität/Anordnung/Befehl – 22826100 bei dem Lemma *künig*³⁵ als ein maßgebliches Element im begrifflichen Gehalt dieser Herrscherbezeichnung aufzufassen, bei *unterscriben*³⁶ erscheint es vielmehr als eine zusätzliche Spezifizierung, nämlich der gesellschaftlichen Sanktionierung der persönlichen Identifizierung des betreffenden (unterschreibenden) Subjekts. Diese potentielle Asymmetrie ist weniger problematisch für die Lemma-Angaben im Wortindex, die differenzierende oder ggf. präzisierende Informationen im unmittelbaren Zusammenhang bieten können. Diese sind aber zunächst nicht direkt vorhanden bei den Ergebnissen von Suchabfragen nach Textbelegen oder in manchen verallgemeinernden intern zugänglichen direkten SQL-Abfragen. In solchen Fällen gilt es, die besagten Aspekte der Bedeutungsbeschreibung und der Abfragemodi in der Auswertung und Interpretation der Ergebnisse zu respektieren und zu berücksichtigen.

Es lässt sich abschließend festhalten, dass die im jahrelangen Einsatz bewährten Kernfunktionalitäten der MhdBDB im Bereich der Textsuche und der semantischen Annotation mittels eines Begriffssystems durch aktuell verlaufende Relaunch-Arbeiten um weitere Facetten und Anwendungsmöglichkeiten bereichert werden sollen. Besonders positiv hervorzuheben ist dabei die Akzentuierung der Kompatibilität und der potenziellen Interoperabilität mit weiteren Informationssystemen und Datenquellen unterschiedlicher Typen – sei es auf der Ebene der Speicherung von Textdaten oder der Erschließung und Vernetzung der wortsemantischen Informationen. Dazu sollen vor allem die Ausrichtung nach TEI und eine Integration mit dem Semantic Web wesentlich beitragen. Einige zusammenhängende Projekte lassen die Aussicht realistisch erscheinen, dass damit hilfreiche textanalytische Mittel bereitgestellt werden, durch die die Einsatzmöglichkeiten dieser inzwischen traditionsreichen Datenbank maßgeblich erweitert werden.

35 MhdBDB – Wortindex: künig. <http://mhdadb.sbg.ac.at:8000/mhdadb/App?lid=3507&frequencys=0&action=Dic&mode=00> (23. 9. 2019); sämtliche zugeordnete Begriffskategorien: Männlich – 21012000, Autorität/Anordnung/Befehl – 22826100, Ämter – 23310200, Gesellschaftliche Ordnung/Allg. Ordnungskategorien – 24300000, Geburtsadel – 24321000, Rechtswesen – 24330000, Ordnungsmacht – 24340000, Wert/Unwert – 31330000.

36 MhdBDB – Wortindex: unterschriben. <http://mhdadb.sbg.ac.at:8000/mhdadb/App?lid=19475&frequencys=0&action=Dic&mode=00> (23. 9. 2019); sämtliche zugeordnete Begriffskategorien: Objektbezogene Aktivität/Tätigkeit – 21072000, Autorität/Anordnung/Befehl – 22826100, Schriftliche Kommunikation – 23122000, Raum – 31500000.

Quellen und Literatur

Primärliteratur

- Brom, Vlastimil (hrsg.) (2009): Di tutsch kronik von Behem lant. Die gereimte deutsche Übersetzung der alttschechischen Dalimil-Chronik. Rýmovaný německý překlad staročeské Dalimilovy kroniky. Brno: MU.
- Brom, Vlastimil (hrsg.) (2019): Spätmittelalterliche deutsche historiographische Texte böhmischer Provenienz – philologische Analyse, elektronische Edition – Texte. <https://www.phil.muni.cz/german/projekty/hmb/e-text-hmb-de.htm> (3. 9. 2019).

Sekundärliteratur

- Brom, Vlastimil (2017): Möglichkeiten der semantischen Auswertung von spätmittelalterlichen historiographischen Texten. Stichproben zum Themenbereich ‚Herrschaft‘. In: Kusová, Jana – Malechová, Magdalena – Vodrážková, Lenka (hrsg.): Deutsch ohne Grenzen. Linguistik. Brno: Tribun EU 2015, S. 51–67. <http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/39643> (12. 3. 2016).
- Brom, Vlastimil (2017): Zur Begrifflichkeit und sprachlichen Erfassung von ‚Herrschaft‘ als einer der zentralen konzeptuellen Domänen der historiographischen Werke des Mittelalters. In: Kotůlková, Veronika – Rykalová, Gabriela (hrsg.): Zentrum und Peripherie. Aus sprachwissenschaftlicher Sicht. Opava: Slezská univerzita v Opavě, S. 323–343. <http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/46981> (18. 9. 2019).
- Echelmeyer, Nora – Reiter, Nils – Schulz, Sarah (2017): Ein PoS-Tagger für „das“ Mittelhochdeutsche. In: Digital Humanities im deutschsprachigen Raum (Dhd) 2017. Konferenzabstracts, S. 141–147. https://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/MHD_Tagger/paper.pdf (24. 9. 2019).
- Em. Univ.-Prof. Dr. Klaus M. Schmidt – MhdBDB. <http://mhdadb.sbg.ac.at:8300/wordpress/klaus-m-schmidt/> (11. 9. 2019).
- Hallig, Rudolf – Wartburg, Walther von (1963): Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas. 2. Aufl. Berlin: Akademie-Verlag.
- Hinkelmanns, Peter – Zeppezauer-Wachauer, Katharina (im Druck): ez ist ein wârheit, niht ein spel, daz netze was sinewel. Die MHDDB im Semantic Web (im Druck). Tagung: Digitale Methoden und Objekte in Forschung und Vermittlung der mediävistischen Disziplinen. Tagung der Bamberger Graduiertenschule für Mittelalterstudien. Bamberg, 8.–10. November 2018.
- IMAREAL, Institut für Realienkunde des Mittelalters und der frühen Neuzeit <http://www.imareal.sbg.ac.at/home/> (3. 9. 2019).
- Klaus Maximilian Schmidt – ResearchGate. https://www.researchgate.net/profile/Klaus_Maximilian_Schmidt (11. 9. 2019).
- MhdBDB – Statistik. <http://mhdadb.sbg.ac.at/LastStatistics.de.html> (20. 9. 2019; Stand zum 9. 9. 2019).
- MhdBDB – Text List. <http://mhdadb.sbg.ac.at/mhdadb/App?action=TextList> (20. 9. 2019).
- MhdBDB – Textsuche nach Belegstellen. <http://mhdadb.sbg.ac.at/mhdadb/App?action=TextQueryModule> (20. 9. 2019).
- MhdBDB – Wortindex – Suchformular. <http://mhdadb.sbg.ac.at/mhdadb/App?action=DicSelect> (12. 9. 2019).

- MhdBDB – Wortindex: künic. <http://mhdbdb.sbg.ac.at:8000/mhdbdb/App?lid=3507&frequency=0&action=Dic&mode=00> (23. 9. 2019).
- MhdBDB – Wortindex: unterschriben. <http://mhdbdb.sbg.ac.at:8000/mhdbdb/App?lid=19475&frequency=0&action=Dic&mode=00> (23. 9. 2019).
- MhdBDB: Mittelhochdeutsche Literatur auf Knopfdruck. http://mhdbdb.sbg.ac.at:8300/wordpress/mittelhochdeutsche_literatur_auf_knopfdruck_de/ (11. 9. 2019).
- ONAMA – Ontologie der Narrative des Mittelalters. <http://onama.sbg.ac.at/> (3. 9. 2019).
- POS Tagger für mittelhochdeutsche Texte. <http://clarin05.ims.uni-stuttgart.de/mhdt/index.html> (24. 9. 2019).
- Roget, Peter Mark (1852): Thesaurus of English Words and Phrases. Classified and arranged so as to facilitate the Expression of Ideas and assist in Literary Composition (1. Aufl.). London 1852 (ferner in zahlreichen Neuauflagen und verschiedenen Bearbeitungen). – Zit. nach Wiegand, Herbert Ernst (2004).
- Schmidt, Klaus M. (1993): Begriffsglossar zu Ulrichs von Zatzikhoven Lanzelet. Tübingen: Niemeyer. (Indices zur deutschen Literatur 25).
- Schmidt, Klaus M. (1988): Der Beitrag der begriffsorientierten Lexikographie zur systematischen Erfassung von Sprachwandel und das Begriffswörterbuch zur mittelhochdeutschen Epik. In: Bachofer, Wolfgang (hrsg.): Mittelhochdeutsches Wörterbuch in der Diskussion (Germanistische Linguistik 84). Tübingen: Niemeyer, S. 35–49.
- Schmidt, Klaus M. (1980): Begriffsglossare zu Ulrich von Lichtenstein. (2 Bände). Tübingen: Niemeyer. (Indices zur deutschen Literatur 14/15).
- Sprache: Mittelhochdeutsch – MPS – Repository. <http://medieval-plants.org/mps-daten/languages/gmh/> (20. 9. 2019).
- Springeth, Margarete (2009): Der analytische Weg ist das Ziel: Die Mittelhochdeutsche Begriffsdatenbank als Online-Textarchiv. In: Hofmeister, Wernfried – Hofmeister-Winter, Andrea (hrsg.): Wege zum Text. Über die Verfügbarkeit mediävistischer Editionen im 21. Jahrhundert. Grazer Kolloquium 17.–19. September 2008. Tübingen: Niemeyer (Beihefte zu Editio, Bd. 30), S. 185–202.
- Wiegand, Herbert Ernst (2004): Lexikographisch-historische Einführung. In: Dornseiff, Franz (Begr.) (2004): Der deutsche Wortschatz nach Sachgruppen. 8., völlig neu bearbeitete und mit einem vollständigen alphabetischen Zugriffsregister versehene Auflage von Uwe Quasthoff. Mit einer lexikographisch-historischen Einführung und einer ausgewählten Bibliographie zur Lexikographie und Onomasiologie von Herbert Ernst Wiegand. Berlin – New York: Walter de Gruyter. E-Book (2010). <https://www.degruyter.com/viewbooktoc/product/61548> 2010 (31. 7. 2019), S. 13*– 91*.
- Wörterbuchnetz – Mittelhochdeutsches Handwörterbuch von Matthias Lexer. <http://woerterbuchnetz.de/Lexer/> (20. 9. 2019).

Mgr. Vlastimil Brom, Ph.D. / brom@phil.muni.cz

Masarykova univerzita, Filozofická fakulta, Ústav germanistiky, nordistiky a nederlandistiky
Arna Nováka 1, 602 00 Brno, CZ



This work can be used in accordance with the Creative Commons BY-SA 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-sa/4.0/legalcode>). This does not apply to works or elements (such as image or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights