

*Thomas Burch/Jost Gippert/Sarah Ihden/Sarah Kwekkeboom/Ralf Plate/
Lena Schnee/Ingrid Schröder/Roland Schuhmann/Johanna Wittmack*

Neue Wege in der Analyse der historischen Wortbildung des Deutschen

Aufbau und Nutzungsmöglichkeiten der digitalen Forschungsumgebung ‚Wortfamilien diachron‘ (WoDia)

Abstract This paper outlines the long-term project ‘Wortfamilien diachron’ (= Word Families in Diachrony; WoDia), funded by Deutsche Forschungsgemeinschaft (DFG) and executed at the Universities of Frankfurt a. M., Hamburg, Rostock, and Trier. Its primary goal is the development of an online database-driven research environment for the historical word formation of German, and it is intended for implementation in both research and teaching. The dataset comprises lemmas from the Old High German, Middle High German, Old Saxon and Middle Low German dictionaries. WoDia integrates these lemmas into an overarching word family structure by putting each word in its position within its word family. The individual word formation elements and the word formation hierarchy are mapped in a structural formula. The four vocabularies are interconnected and linked to the online dictionary resources and to the reference corpora of historical German. This approach used in the research environment WoDia enables the user to analyse processes of change within the historical vocabularies of German, particularly transformations in word family structures and the usage of word formation elements. The methodical approach employed in constructing the database is presented in this paper, and concrete possibilities of application are illustrated by means of three exemplary case studies.

Keywords word formation, language change, word families, morphology, database

1. Einleitung

In den letzten Jahren hat die digitale Aufbereitung von Sprachdaten erheblich zugenommen, was sowohl der Forschung als auch der Lehre neue Möglichkeiten eröffnet. Für die historische Sprachwissenschaft sind in diesem Zusammenhang unter anderem die grammatisch annotierten Referenzkorpora zur deutschen Sprachgeschichte im Verbund ‚Deutsch Diachron Digital‘ mit den Korpora Altdeutsch, Mittelhochdeutsch, Frühneuhochdeutsch, Mittelniederdeutsch/Niederrheinisch und Deutsche Inschriften (vgl. Ihden et al. 2023)¹ von besonderer Bedeutung. Im Bereich der Lexikographie ist vor allem die Digitalisierung zahlreicher historischer Wörterbücher hervorzuheben.² Für die Forschung zur historischen Wortbildung steht jedoch bislang noch kein geeignetes Instrument zur Verfügung, denn die Annotationen in den Korpora konzentrieren sich auf Wortart, Flexion und ansatzweise

1 Vgl. auch die Webseite des Verbundes ‚Deutsch Diachron Digital‘ unter www.deutschdiachrondigital.de/ (Stand: 1.7.2025).

2 Die Wörterbücher sind im Trierer Wörterbuchnetz online publiziert und digital verfügbar unter <https://woerterbuchnetz.de> (Stand: 1.7.2025).

auf die Satzstruktur, während Wortbildungsstrukturen nicht ausgezeichnet sind. Das WoDia-Projekt schließt diese Lücke, indem eine digitale Forschungsumgebung geschaffen wird, die den historischen Wortschatz des Deutschen varietäten- und sprachstufenübergreifend auf der Wortbildungsebene analysiert, ihn in einer Wortfamilienstruktur abbildet und miteinander verbindet sowie eine Verknüpfung mit den Lemmata der Online-Wörterbücher und der Referenzkorpora vornimmt.

2. Ziele des Projekts

Im WoDia-Projekt wird eine umfassende empirische Grundlage für die Forschung zur historischen Wortbildung des Deutschen bereitgestellt. Die Basis bilden die Referenzwörterbücher dieser Wortschätze:

- Für Althochdeutsch das Wörterbuch von Jochen Splett (1993) (AWB^{Spl}) und das noch in Ausarbeitung befindliche Leipziger ‚Althochdeutsche Wörterbuch‘ (AWB).
- Für Mittelhochdeutsch Lexers Handwörterbuch (1872–1878) (Lex) und das in Ausarbeitung befindliche neue Mittelhochdeutsche Wörterbuch (MWB).
- Für Altsächsisch das Handwörterbuch von Heinrich Tiefenbach (2010) (ASWB) und wiederum das AWB, das auch den gesamten altsächsischen Wortschatz der kleineren Texte und Glossen mitbearbeitet.
- Für Mittelniederdeutsch das in Bearbeitung befindliche Mittelniederdeutsche Handwörterbuch (MNWB) sowie ergänzend das Wörterbuch von Lübbers/Walther (1888) (LW).

Mit diesen Wörterbüchern ist die gesamte ältere deutsche Sprachgeschichte bis ins 15. Jahrhundert (Lex für das Mittelhochdeutsche) bzw. bis zum Anfang des 17. Jahrhunderts (MNWB für das Mittelniederdeutsche) abgedeckt. Für das Mittelhochdeutsche wird dabei jeweils notiert, ob ein Lemma bereits vor 1350 (rd. 50.000 Wörter) oder erst nach 1350 (rd. 31.000 Wörter) bezeugt ist. So ist der Anteil des Wortschatzes aus der älteren frühneuhochdeutschen Überlieferung unterschieden von dem mittelhochdeutschen Wortschatz im engeren Sinne der heutigen Epochengliederung.

Die Referenzwörterbücher werden als Materialbasis für WoDia gewählt, da der Umfang ihrer Wortschätze den der entsprechenden Referenzkorpora ReA (Referenzkorpus Althochdeutsch = Althochdeutsch + Altsächsisch), ReM (Referenzkorpus Mittelhochdeutsch) und ReN (Referenzkorpus Mittelniederdeutsch/Niederrheinisch) erheblich übersteigt. Die Referenzkorpora zielen nicht primär auf eine lexikalische, sondern stärker auf eine grammatische, vor allem flexionsmorphologische Erschließung der jeweiligen Sprachstufe. Tabelle 1 zeigt die unterschiedlichen Umfänge.

Sprache	Wörterbuch	Referenzkorpus
Althochdeutsch	28.500 (AWB ^{SpI})	10.900 (ReA)
Altsächsisch	5.700 (ASWB)	4.100 (ReA)
Mittelhochdeutsch	81.000 (Lexer, bis 15. Jh.) davon 50.000* bis 1350 (MWB)	22.300 (ReM, bis 1350)
Mittelniederdeutsch	80.000* (MNWB)	17.800 (ReN)

Tab. 1: Anzahl der Lemmata in Wörterbüchern und Referenzkorpora; * = hochgerechnete Endumfänge in noch nicht abgeschlossenen Wörterbuch-Projekten

Die vier Wortschätze werden im Projekt miteinander verknüpft, wortbildungsmorphologisch analysiert, in eine übergreifende Wortfamilienstruktur überführt und in einer Forschungsumgebung mit benutzerfreundlichem Frontend für eine Vielzahl von Fragestellungen der historischen Wortbildungslehre bereitgestellt. Die Lemmata der Datenbank sind über ihre IDs mit den Online-Angeboten der Referenzwörterbücher verlinkt, sodass zu jedem Wort seine lexikographische Darstellung verglichen werden kann, im Falle der großen Belegzitatzwörterbücher (AWB, MWB) auch ihr Gebrauch in den Belegzitaten. Die Online-Angebote für das AWB^{SpI}, das ASWB, das MNWB und für LW werden vom Projekt selbst eingerichtet und im Trierer Wörterbuchnetz publiziert, in dem bereits die übrigen Referenzwörterbücher zugänglich sind. Über die Lemma-IDs erfolgt auch eine Verlinkung in die Referenzkorpora zur deutschen Sprachgeschichte, sodass ihre flexionsmorphologische Annotation um die wortbildungsmorphologische Analyse in WoDia ergänzt werden kann. Das ReM ist bereits mit der Mittelhochdeutsch-Lemmaliste (Lexer + MWB) verknüpft, für das ReA und das ReN ist die Verlinkung in die Referenzwörterbücher noch einzubringen.

Das auf neun Jahre (2024–2033) angelegte Projekt wird in drei Arbeitsgruppen an vier Standorten durchgeführt, die eng miteinander kooperieren. Mittelniederdeutsch wird an den Universitäten in Hamburg und Rostock bearbeitet, die Arbeiten zum Althochdeutschen, Altsächsischen und Mittelhochdeutschen werden an der Universität Frankfurt a. M. durchgeführt. Die Einrichtung der zentralen Datenbank, die Online-Publikation der Referenzwörterbücher und die Organisation des digitalen Workflows liegt in den Händen der Arbeitsgruppe am Trierer Kompetenzzentrum (Trier Center for Digital Humanities).³

3 In Frankfurt a. M. setzt sich das Projektteam aus Jost Gippert, Ralf Plate (Projektleitung), Sarah Kwekkeboom, Roland Schuhmann (wissenschaftliche/r Mitarbeiter/in), Alexandre Jouan und Janina Damberg (Hilfskräfte) zusammen (vgl. die Webseite am Frankfurter Standort unter www.uni-frankfurt.de/153903162/Wortfamilien_diachron, Stand: 1.7.2025), in Hamburg aus Ingrid Schröder (Projektleitung), Lena Schnee (wissenschaftliche Mitarbeiterin), Ronia Seiri und Sarah Möller (Hilfskräfte) (vgl. die Webseite am Hamburger Standort unter www.slm.uni-hamburg.de/niederdeutsch/forschung/projekte/wortfamilien-diachron.html, Stand: 1.7.2025), in Rostock aus Sarah Ihden (Projektleitung), Johanna Wittmack (wissenschaftliche Mitarbeiterin) und Lara Stapel (Hilfskraft) (vgl. die Webseite am Rostocker Standort unter <https://www.germanistik.uni-rostock.de/forschung/sprachwissenschaft/wodia>, Stand: 16.2.2026) und in Trier aus Thomas Burch (Projektleitung), Julia Fischer, Felix Thielen (wissenschaftliche/r Mitarbeiter/in) sowie Marina Spielberg und Bianca Ottenberg (Hilfskräfte) (vgl. die Webseite am Trierer Standort unter <https://tdh.uni-trier.de/de/projekt/wortfamilien-diachron-wodia>, Stand: 1.7.2025).

3. Arbeitspakete und Methoden

Die Bearbeitung der digitalen Daten in WoDia basiert auf einem mehrstufigen Verfahren zur Analyse und Strukturierung der historischen Wortschätze. Diese Vorgehensweise wurde in einem Pilotprojekt an der Universität Frankfurt a.M. in den Jahren 2016–2019 erprobt (vgl. Plate 2020, 2022).⁴ Das Ziel bestand darin, die althochdeutsche Wortfamiliengliederung nach dem Modell von Splett (AWB^{Spl}) auf das Mittelhochdeutsche zu übertragen. Besonders hervorzuheben sind die Vorarbeiten von Thomas Klein (2018), auf denen die WoDia-Systemarchitektur aufbaut.

Die Analysen in WoDia folgen einem systematisch modularisierten Ablauf, der es ermöglicht, große Wortschatzmengen effizient zu bearbeiten. Für jede Sprachvarietät (Althochdeutsch, Mittelhochdeutsch, Altsächsisch, Mittelniederdeutsch) werden bestimmte Arbeitsschritte durchlaufen, wobei aufgrund der oben erwähnten Vorarbeiten ein unterschiedlicher Bearbeitungsstand in den Sprachstufen vorliegt. Aus diesem Grund sind die Arbeitspakete nicht zwingend nacheinander angeordnet, sondern können je nach Sprachstufe teilweise auch parallel liegen. Dieser Abschnitt bietet einen Überblick über die Arbeitspakete in WoDia und stellt die technischen Verfahren vor, mit denen die teil-automatisierte Analyse der historischen Wortschätze ermöglicht werden soll.

3.1 Erstellung der Lemmalisten

Die Datenbasis bilden die auf den jeweiligen Referenzwörterbüchern basierenden Lemmalisten. Für Althochdeutsch und Mittelhochdeutsch liegen diese bereits vor und sind miteinander verknüpft, für Altsächsisch und Mittelniederdeutsch müssen sie zunächst aus den Digitalisaten der Referenzwörterbücher zusammengestellt werden. Insbesondere für den umfangreichen mittelniederdeutschen Wortschatz stellt dies keine triviale Aufgabe dar. Aufgrund der im MNWB teilweise uneinheitlichen formalen Auszeichnung beispielsweise der Verweislemmata oder Nestartikel⁵ lässt sich nicht für alle Lemmata eindeutig formal festlegen, was ein Lemma ist und wie/ob es in die Liste aufgenommen werden soll, sodass dieser Schritt nicht voll automatisierbar ist. Für jedes Lemma werden aus den mittelniederdeutschen Wörterbüchern (MNWB + LW) zusätzlich die grammatischen Angaben sowie – soweit verfügbar – die Bedeutungsangaben extrahiert (vgl. exemplarisch Tab. 2 für das Mittelniederdeutsche).

Lemma	Grammatische Angabe	Bedeutungsangabe
begräven	stv.	begraben
begrävinge	f.	Bestattung, Begräbnis
begrëvenisse	f.	Bestattung, Begräbnis

Tab. 2: Auszug aus der mittelniederdeutschen Lemmaliste

4 Das Projekt ‚Wortfamilien. Computergestützte Etablierung epochenübergreifender Wortfamilienstrukturen‘ war Teil der mit Mitteln des BMBF geförderten Initiative für ein eHumanities-Zentrum für historische Lexikographie (ZHISTLEX) und wurde von Jost Gippert und Ralf Plate geleitet.

5 Nestartikel im MNWB sind in den Artikel eines durch Fettdruck gekennzeichneten Hauptlemmas (z.B. *veftich*) eingebundene Wörterbuchartikel weiterer, aus dem Hauptlemma gebildeter Lemmata (z.B. *veftichmäker*); letztere werden typographisch durch Kursivierung abgehoben.

3.2 Segmentierung der Lemmata und Ermittlung der Wortfamilienstämme

Ein erster Schritt auf dem Weg zur Zuordnung der Wortfamilien besteht in der Segmentierung der Lemmata in ihre Konstituenten. Diese wird mithilfe von zuvor erstellten Affixlisten durchgeführt. Auf diese Weise können die Affixe angesteuert und abgetrennt und so die Stämme der jeweiligen Lemmata freigelegt werden. Zusätzlich werden mögliche Kompositionsglieder anhand von Simplexlisten identifiziert, die zuvor durch ein Skript erstellt wurden (vgl. Abschn. 3.4). Dabei können auch Fugenelemente erkannt und abgetrennt werden.⁶ Durch die Segmentierung, d. h. das Abtrennen der Affixe (wie *be-*, *-inge* und *-nisse* in Tab. 3) bzw. das Trennen von Simplizia, wird der Stamm zunächst in der Form ermittelt, in der er im entsprechenden Lemma realisiert ist. So zeigt z. B. in mnd. *begrēvenisse* (vgl. Tab. 3) der freigelegte Stamm *-grēv-* den durch das Suffix *-nisse* ausgelösten Umlaut. Dieser Stamm muss von Umlaut, Ablaut und anderen Lautwechselphänomenen bereinigt werden, um den Wortfamilienstamm, die Grundform des Stammparadigmas eines Lexems (vgl. Duden-Grammatik 2022, § 1050; Duden-Grammatik 2016, S. 664 f.; Eisenberg 2020, S. 30, 231), zu ermitteln. So erhält man hier im Ergebnis die Grundform *grāv*, den verbalen Stamm der Wortfamilie *GRĀVEN*, der die Bildungen damit zugeordnet sind. Alle Lemmata, die derselben Wortfamilie angehören, hier *GRĀVEN*, werden zu einem späteren Zeitpunkt nach dem Muster im AWB^{SpI} in ihrer Wortfamilie gruppiert (vgl. Abb. 1 in Abschn. 4.2.1).

Lemma	Segmentierung	Stammform	Wortfamilienstamm	Wortfamilie
begrāven	be-grāv-en	grāv	grāv	GRĀVEN
begrāvinge	be-grāv-ingē	grāv	grāv	GRĀVEN
begrēvenisse	be-grēv-e-nisse	grēv	grāv	GRĀVEN

Tab. 3: Segmentierung und Stammermittlung anhand mittelniederdeutscher Beispiele

In WoDia erfolgt eine Verknüpfung der vier historischen Varietäten auf zwei Ebenen: Einerseits zwischen den einzelnen Lemmata (z. B. ahd. *bigraban* – mhd. *begraben* – as. *bigravan* – mnd. *begrāven*, vgl. Tab. 4), andererseits über die Wortfamilien (z. B. ahd. *GRABAN* – mhd. *GRABEN* – as. *GRAVAN* – mnd. *GRĀVEN*),⁷ wodurch auch miteinander verwandte Lemmata derselben Familie, die keine etymologische Entsprechung in einer der anderen Varietäten haben, einander zugeordnet sind (z. B. mnd. *grefte*, f. ‚Graben‘ und mnd. *grāvengenger*, m. ‚Wächter für Wälle und Gräben‘).

6 Die Untersuchung und Diskussion der Frage, ab welchem Zeitpunkt es sich um ein Fugenelement und nicht mehr um ein Flexionsmorphem handelt, kann im Rahmen von WoDia nicht geleistet werden. Auf Flexion basierende potenzielle Fugenelemente werden daher grundsätzlich als Fugenelement annotiert.

7 In der Forschungsumgebung kann bei der Suche zwischen Lemma, Wortfamilie und Strukturformel differenziert werden. Eine Suche nach z. B. mhd. *graben* auf der Ebene der Wortfamilie gibt auch die Wortfamilien der anderen Varietäten aus.

Sprachstufe	Lemma	Segmentierung	Wortfamilie
Althochdeutsch	bigraban	bi-grab-an	GRABAN
Mittelhochdeutsch	begraben	be-grab-en	GRABEN
Altsächsisch	bigravan	bi-grav-an	GRAVAN
Mittelniederdeutsch	begräven	be-gräv-en	GRÄVEN

Tab. 4: Verknüpfung zwischen den Varietäten über das Einzelemma und die Wortfamilie

Die Tatsache, dass für das Althochdeutsche die Segmentierung, die Zuordnung zu Wortfamilien und die wortbildungsmorphologische Analyse bereits im AWB^{Spl} vorliegen, kann in WoDia auch für die Arbeiten an den anderen Sprachstufen genutzt werden. Die Verknüpfung mit den etymologisch entsprechenden Einträgen einer bereits analysierten Lemmaliste (Mittelhochdeutsch mit Althochdeutsch, Altsächsisch mit Althochdeutsch, Mittelniederdeutsch mit Mittelhochdeutsch) bietet den Vorteil, dass deren strukturelle Analyse (vgl. Abschn. 3.3) übernommen werden kann.

3.3 Erstellung von Strukturformeln

Die wortbildungsmorphologische Analyse der Lemmata in WoDia wird in einer Strukturformel wiedergegeben. Das Vorbild hierfür bieten die Wortfamilienwörterbücher zum Althochdeutschen (1993) und zur deutschen Gegenwartssprache (2009) von Jochen Splett. Diese bieten nicht nur eine Gliederung des Wortschatzes in Wortfamilien, sondern liefern für jedes Lemma eine hierarchische, geklammerte Strukturformel mit Informationen zur Konstituentenstruktur, zur Wortart der Stämme und zur Wortbildungsart (Komposition, Derivation, Konversion). Einige Beispiele aus dem Wortfamilienwörterbuch zur Gegenwartssprache (Splett 2009, Bd. 1, S. 416–417) verdeutlichen dieses Prinzip:

- *Bau-er*: (wV)sS – zu lesen als: substantivische Ableitung aus dem Verbstamm BAU mit dem Suffix *-er* (w = Wurzel/Stamm; V = Verb; s = Suffix; S = Substantiv).
- *Acker-bau-er*: (wS)((wV)sS) – Kompositum mit den Konstituenten ACKER und BAU-ER.
- *Boots-bau-er*: ((wS)((wV)S))sS – substantivische Ableitung aus dem zusammengesetzten Substantiv BOOTS-BAU mit dem Suffix *-er*; *-bau* wiederum ist eine substantivische Konversion des Verbstamms BAU.
- *Er-bau-er*: (p(wV))sS – Substantivableitung aus dem präfigierten Verbstamm ER-BAU mit dem Suffix *-er* (p = Präfix).

Aus Doppel- oder Mehrfachmotivationen ergeben sich Mehrdeutigkeiten, denen durch die Angabe alternativer Strukturformeln Rechnung getragen wird, z.B. im Fall des Adjektivs *baukünstlerisch* (Splett 2009, Bd. 1, S. 414) mit drei Formeln für die drei möglichen Analysen: Mit *baukünstler-isch* vs. *bau-künstlerisch* als unmittelbaren Konstituenten, und wiederum im ersten Fall mit *bau-künstler* vs. *baukunst-ler* als unmittelbaren Konstituenten des ersten Elements.

Die Strukturformeln bieten mit ihrer formalisierten Beschreibung der zugrunde liegenden Wortbildungsschritte und der beteiligten Morpheme eine einzigartige Möglichkeit zur Untersuchung historischer Wortbildungsregularitäten, die in den Printpublikationen der Splettschen Wortfamilienwörterbücher bislang kaum genutzt werden konnte. Mit der Online-Publikation des AWB^{Spl} wird in einem ersten Schritt die Suche über die Strukturformeln für das Althochdeutsche ermöglicht. Später wird die WoDia-Forschungsumgebung für alle erwähnten historischen Wortschätze die Strukturformeln sämtlicher Lemmata durchsuchbar und exportierbar machen.

3.4 Teil-Automatisierung der Arbeitsschritte

Die oben beschriebenen Arbeitsschritte können durch den Einsatz regelbasierter Programme unterstützt werden. Als methodische Leitlinie für alle Arbeitsschritte gilt ein iteratives teil-automatisiertes Verfahren. Dies bedeutet, dass mithilfe von Skripten automatische Vorschläge, z. B. zur Segmentierung der Lemmata, zum Wortfamilienstamm, zur Verknüpfung der Lemmata und zu den Strukturformeln, erzeugt werden, die in einem nächsten Schritt händisch zu prüfen sind. Dabei können Fehlerquellen entdeckt und die Skripte daraufhin angepasst und erneut eingesetzt werden, wodurch die automatischen Vorschläge sukzessive verbessert werden. Die Skripte zur Bearbeitung des mittelhochdeutschen Wortschatzes wurden bereits im Rahmen der erwähnten Vorarbeiten erstellt (zur technischen Umsetzung siehe auch Abschn. 3.5). Für das Mittelniederdeutsche müssen sie entsprechend angepasst werden. Am Beispiel der Arbeiten am mittelniederdeutschen Wortschatz soll im Folgenden die Teil-Automatisierung der Arbeitsschritte genauer beschrieben werden.

Für die Segmentierung der mittelniederdeutschen Lemmata wird ein Skript verwendet,⁸ das Präfixe und Suffixe formal identifiziert und abtrennt (vgl. Tab. 3 in Abschn. 3.2).⁹ In das Skript werden dabei zusätzlich für einzelne Suffixe zu vermeidende Fehltreffer eingegeben, die formal wie das Suffix enden, bei denen das jeweilige Suffix aber nicht vorliegt und deshalb keine Segmentierung durchgeführt werden soll (zum Beispiel ist *dēgedinge* keine Bildung mit *-inge* und sind Komposita mit *-stēn* keine Bildungen mit *-in* usw.). In einem iterativen Verfahren werden im Zuge der Durchsicht der Vorschläge weitere solcher Fehltreffer identifiziert und von der Segmentierung ausgeschlossen. Auf der Basis der formal identifizierten Affixe und der übrigbleibenden Stämme wird zudem automatisch ein Vorschlag für eine nicht-hierarchische Strukturformel erstellt,¹⁰ d. h., die Reihenfolge, in der sich die jeweiligen Konstituenten zusammenfügen, wird als Formel dargestellt, z. B. für das oben erwähnte mnd. *be-grāv-en* die Formel *p/wV*. Für alle Lemmata, bei denen im automatischen Prozess keine Segmentierung erfolgt (d. h., die als reine Stämme aufgefasst werden), wird zunächst manuell geprüft, ob es sich tatsächlich um Simplizia handelt. Die daraus resultierende Simplizialiste wird wiederum verwendet, um – analog zum Vorgehen bei den Affixen – für diejenigen Lemmata, bei denen keine Segmentierung der Komposition aus den Vorarbeiten vorliegt, Segmentierungsvorschläge zu produzieren. Auch fließt die Wort-

8 Alle im Rahmen von WoDia entstandenen Skripte sollen zu einem späteren Zeitpunkt veröffentlicht werden.

9 Die Affixlisten basieren auf den Vorarbeiten von Dieter Möhn und Ingrid Schröder zur Wortbildung des Mittelniederdeutschen (vgl. z. B. zur Adjektivbildung Möhn/Schröder 2009), die auch in eine geplante korpusbasierte mittelniederdeutsche Grammatik einfließen sollen (vgl. Ihden/Schröder 2021). Für weitere Vorarbeiten zur mittelniederdeutschen Wortbildung, die Wörterbuch- und/oder Korpusdaten einbeziehen, vgl. Ihden/Schröder (2024) sowie Schröder (2022).

10 Für bestimmte Lemmata, die Kompositionsprodukte sind, kann zusätzlich auf Vorarbeiten aus dem ReN-Projekt zurückgegriffen werden, da diese dort bereits manuell segmentiert wurden.

artangabe der Simplizia direkt in den Strukturformel-Vorschlag ein. So werden schrittweise neue Teilmengen des Wortschatzes segmentiert und analysiert, wodurch die Skripte jeweils erweitert und verbessert werden können.

Ein weiterer Schritt, bei dem regelbasierte Skripte zum Einsatz kommen, ist die Verknüpfung des Wortschatzes; beispielhaft wird hier das Vorgehen für die Verknüpfung von Mittelhochdeutsch und Mittelniederdeutsch skizziert.¹¹ Ein Skript produziert auch hier Vorschläge, die anschließend zu überprüfen sind. Die Verknüpfungsrichtung verläuft dabei vom Hochdeutschen zum Niederdeutschen, da die lautlichen Korrespondenzen in diese Richtung überwiegend eindeutig sind.¹² Auf der Grundlage von knapp 70 Ersetzungsregeln, die für hochdeutsche Laute (bzw. Grapheme) die niederdeutschen Entsprechungen aufführen, erstellt das Skript für jedes mittelhochdeutsche Lemma eine ‚niederdeutsche‘ Form. Beispielsweise wird mhd. <zz> grundsätzlich zu mnd. <t> verändert, wodurch für mhd. *wazzer* mnd. *water* generiert wird. Diese Ersetzungsregeln sind je nach Laut (bzw. Graphem) unterschiedlich komplex und beziehen in einigen Fällen die Position im Wort sowie die folgenden Laute mit ein. Ausgenommen von diesen allgemeinen Lautkorrespondenzen werden Affixe, die systematisch durch ihre mittelniederdeutsche Entsprechung ersetzt werden, da die Lautkorrespondenzen hier nicht greifen. Schließlich werden einige Sonderregeln angewandt, wie z. B., dass mhd. *alt* mnd. *ōlt* entspricht oder dass die Korrespondenz von mhd. <ou> zu mnd. <ô> im Falle des Lexems *vrouwe* (mhd. und mnd.) nicht besteht.

Das Skript gleicht die künstlich hergestellten niederdeutschen Formen mit der mittelniederdeutschen Lemmaliste ab. Existiert eine Form in der mittelniederdeutschen Lemmaliste, wird diese als Verknüpfungsvorschlag zum mittelhochdeutschen Lemma ausgegeben.¹³ Um weiter einzugrenzen, ob es sich um eine Entsprechung handelt, wird außerdem auf beiden Seiten die Wortart der Lemmata mit einbezogen. Werden für ein Lemma mehrere mögliche Verknüpfungen gefunden – beispielsweise bei Homonymen –, werden alle Möglichkeiten vorgeschlagen und bei der anschließenden Durchsicht wird eine Entscheidung getroffen.

Derzeit liegen bereits für knapp 15% der mittelniederdeutschen Lemmata Verknüpfungsvorschläge vor, wobei zu berücksichtigen ist, dass auf mittelniederdeutscher Seite eine noch unvollständige Lemmaliste zum Einsatz kommt, der zudem die Angaben zur Wortart fehlen. Daher ist mit einer deutlichen Verbesserung dieser Bilanz zu rechnen, sobald die mittelniederdeutsche Lemmaliste komplett vorliegt und auf beiden Seiten die Wortart mit einbezogen werden kann. Da es zudem nur eine Teilüberschneidung beider Wortschätze gibt, stellt dieser Anteil automatischer Verknüpfungsvorschläge bereits einen enormen Mehrwert für das Projekt dar. Neben ihrem übergeordneten Zweck, varietätenübergreifende Studien zu ermöglichen, erleichtert die Verknüpfung der verschiedenen Wortschätze die Arbeit im Projekt, da, wie erwähnt, bei der Bearbeitung des Mittelniederdeutschen für

11 Hilfreich waren hierfür insbesondere die Vorüberlegungen von Thomas Klein (2013, Folie 25–36) zur Verknüpfung der mittelhochdeutschen und mittelniederdeutschen Lemmalisten.

12 Vgl. die Übersicht über die lautlichen Korrespondenzen, zur Verfügung gestellt von Lourens Visser unter <https://osf.io/yvfjh> (Stand: 1.7.2025).

13 Weitere Angleichungen sind nötig, um unterschiedlich gesetzten Klammerungen in beiden Wörterbüchern zu begegnen. Enthält ein Lemma, für das so keine Entsprechung im Mittelniederdeutschen gefunden werden kann, Klammern, wird entweder eine Form ohne die Klammern generiert (so wird für mhd. *acker(e)n* > *ackeren* mnd. *ackeren* gefunden), oder es wird eine Form ohne Klammern und ohne Klammerinhalt erstellt (mhd. *ē(we)brēchunge* > *ēbrēchunge* führt zu mnd. *ēbrēkinge*). Ebenso werden auf mittelniederdeutscher Seite Formen ohne Klammern bzw. ohne Klammerinhalt generiert (so kann mnd. *ping(e)stāvent* zu mhd. *pfin gestāvend* und mnd. *bū(w)knecht* zu mhd. *būkneht* gefunden werden).

das Alt- oder Mittelhochdeutsche bereits angefertigte Segmentierungen und Strukturformeln übernommen werden können.

3.5 Technische Umsetzung der digitalen Forschungsumgebung

Parallel zu den skizzierten Arbeitspaketen läuft die Konzeption, Entwicklung und Erweiterung der digitalen Forschungsumgebung. Dazu zählen unter anderem die Modellierung der Datenbank sowie die dafür notwendige Datenintegration aus den unterschiedlichen vorliegenden Quellen. Für das Projektteam wird eine Bearbeitungsschnittstelle bereitgestellt. Daneben wird das Frontend der Forschungsumgebung gestaltet und es werden Anwendungskonzepte entwickelt, die verschiedene Ausgabe-, Export- und Visualisierungstypen auf der vereinheitlichten Datengrundlage ermöglichen. So sollen in der Forschungsumgebung komplexe Suchabfragen anhand intuitiv bedienbarer Steuerelemente zusammengestellt sowie dem individuellen Forschungsinteresse entsprechend die Ergebnisse ausgegeben werden können. Die Ergebnislisten lassen sich z. B. alphabetisch nach Lemma oder Wortfamilie sowie nach der Strukturformel sortieren.

4. Anwendungsbeispiele

Die beschriebenen Strukturformeln eröffnen neue Zugänge für die historische Wortbildungsforschung. Drei exemplarische Untersuchungsfelder sollen dies verdeutlichen und die Analysemöglichkeiten von WoDia aufzeigen. Die beiden Beispiele für den hochdeutschen Sprachraum befassen sich mit den doppelsuffigierten Adjektiven (vgl. Abschn. 4.1) und der Suffixablösung bei den Nomina agentis (vgl. Abschn. 4.2.1). Für eine Analyse im Niederdeutschen wird ein Blick auf funktionsgleiche Movierungssuffixe im Mittelniederdeutschen geworfen (vgl. Abschn. 4.2.2). Beide sind klassische Themenfelder der historischen Morphologie (vgl. Wilmanns 1899; Erben 2006) und erfahren durch die computergestützte Analyse in WoDia eine neue empirische Tiefenschärfe sowie zum ersten Mal eine vergleichbare systematische Aufbereitung.

4.1 Doppelsuffigierte Adjektive

Die Doppelsuffigierung von abgeleiteten Adjektiven wird in der historischen Wortbildungslehre des Deutschen vielfach kritisch betrachtet. Gemeint sind Bildungen auf (nhd.) *-haft-ig*, *-ig-lich* usw. (*wahrhaftig*, *gemeiniglich* usw.), die sich bereits im Althochdeutschen finden. Die älteren Wortbildungslehren sprechen hier von Bildungen, in denen das zweite Suffix zum Teil als „blosse Wucherung“ ohne Bedeutungsunterschied erscheine (Wilmanns 1899, S. 490), und ähnlich wird das aus doppelt suffigierten (ahd.) *-ig-lîh-* (mhd.) *-ec-lich*-Bildungen als eigenes Suffix entwickelte mhd. *-eclich* (bei Bildungen ohne entsprechendes Adjektiv auf einfaches *-ec*) als „Wuchersuffix“ bezeichnet (Henzen 1965, S. 203, 274; KSW III, S. 313, § A 105).

Speltt (1995) hat gezeigt, wie man die Behandlung dieser Erscheinung fürs Althochdeutsche auf eine breite und systematisch erhobene Materialgrundlage stellen kann, nämlich durch Nutzung der Strukturformeln seines Wörterbuchs. In diesem Fall können mit der Suche nach der Formel ((...)sA)sA (Stamm + Adjektivsuffix + Adjektivsuffix) 140 doppelsuffigierte althochdeutsche Adjektive erhoben werden, in denen von den 41 althochdeutschen Adjektivsuffixen 28 in erster oder zweiter Position oder in beiden vorkommen können.

Zu 89 von 140 dieser doppelsuffigierten Bildungen ist ein mit dem an Erstposition stehenden Suffix einfach suffigiertes Adjektiv bezeugt (z. B. *lib-haft* neben *lib-haft-ig*). Diese 89 Wortpaare liefern ein reichhaltiges Material für die genauere Untersuchung der Funktionen der Doppelpräfigierung am althochdeutschen Adjektiv. Entsprechende Materialzusammenstellungen werden in WoDia dann auch für Altsächsisch, Mittelhochdeutsch und Mittelniederdeutsch und den diachronen Vergleich möglich sein.

4.2 Persönliche Agentiva

4.2.1 Suffixablösung bei den Nomina agentis

Wie die Doppelauffigierung von Adjektiven hat auch die Suffixablösung bei den persönlichen Nomina agentis im Althochdeutschen und Mittelhochdeutschen von jeher große Aufmerksamkeit der historischen Wortbildungslehre gefunden (vgl. Erben 2006, S. 150–155). Bei diesem Wortbildungswandel wird der ältere germanische Typus der Ableitung von Nomina agentis durch das *-an/-jan*-Suffix der schwachen Maskulina (z. B. ahd. *gebo* swM. ‚Geber, Spender, Wohltäter‘) durch den jüngeren mit dem (bereits voralthochdeutsch gebrauchten) Lehnsuffix aus lat. *-ārius*, ahd. *-āri*, mhd. *-ære/-er*, nhd. *-er* abgelöst (z. B. *gebâri*, mhd. *geber*, nhd. *Geber* stM.). Die entsprechenden femininen Nomina agentis wurden ursprünglich ebenfalls als schwache (*-ōn-*, *-jōn-*) Stämme gebildet (z. B. ahd. *becka* swF. ‚Bäckerin‘). Die Nebensilbenabschwächung zum Mittelhochdeutschen hin hat dann der deutlicheren neuen Bildungsweise mit dem *-āri/-ære*-Suffix zum Durchbruch verholfen, wobei die Feminina (im Hochdeutschen) jetzt durch Anfügung des Movierungssuffixes ahd. *-in*, mhd. *-inne*, nhd. *-in* an das Lehnsuffix gebildet wurden (nhd. *Bäcker, Bäckerin*).

Charakteristisch für den alten Bildungstyp ist, dass er (schon germanisch) vor allem in Reaktionskomposita (vgl. Krahe/Meid 1969, S. 26) bzw. Zusammenbildungen (vgl. Henzen 1965, S. 61) des Typs ahd. *brôt-becko*, *gast-geba* usw. vorkommt, wobei das Zweitglied häufig gar nicht selbstständig belegt ist (wie ahd. im Falle von **becko* swM. und **geba* swF.) und vielleicht primär gar nicht frei gebraucht wurde. Als Erstglied kommen vor allem Substantive vor, daneben auch Adjektive wie in *wîs-sago* swM. ‚Wahrsager‘, *blint-slich* swM. ‚Blindschleiche‘ (Henzen 1965, S. 69). Die Nomina agentis im Zweitglied können nicht nur von Verben abgeleitet sein, sondern auch von Substantiven wie z. B. in den *jan*-stämmigen Bildungen ahd. *scerio* swM. ‚Scharmeister‘ (nhd. *Scherge*) und *sidilo* swM. ‚Einwohner‘ zu ahd. *scara* ‚Schar‘, *sidil* ‚Siedel, Sitz‘ (Henzen 1965, S. 133).

Dieser Wortbildungswandel verdiente eine erneute Untersuchung auf der umfassenden Materialgrundlage des heutigen Stands der Lexikographie des älteren Deutsch, insbesondere des Althochdeutschen. Die Materialerhebung zu den Bildungen älteren Typs wird dabei durch Spletts Strukturformeln stark erleichtert. Die Suche in ihnen zu den deverbabilen Bildungen ergibt

- für die Formel (wV)*San* der maskulinen einfachen Bildungen 87 Treffer,
- für die Formel (wV)*Sôn* der femininen einfachen Bildungen 45 Treffer,
- für die Formeln (wS)((wV)*San*) bzw. ((wS)+(wV))*San* der maskulinen Komposita bzw. Zusammenbildungen mit substantivischem Erstglied 39 bzw. 78 Treffer,
- für die Formeln (wS)((wV)*Sôn*) bzw. ((wS)+(wV))*Sôn* der femininen Komposita bzw. Zusammenbildungen mit substantivischem Erstglied 8 bzw. 17 Treffer.

Abbildung 1 zeigt exemplarisch einen Ausschnitt des Suchergebnisses auf AWB^{Sp1} Online für die Formel (wV)San.

The screenshot shows the AWB Online search interface. The search bar at the top right contains the text 'Stichwort suchen'. The main content area displays search results for the formula (wV)San. The results are organized into sections: 'DRÄEN (1) vgl. DRÄTI' and 'DRAGA (1)'. Each section lists words and their corresponding forms, along with their meanings and grammatical information. For example, under 'DRÄEN', there are entries for 'dräen', 'gi-dräen', 'drä-faz', 'drä-isarn', 'dräo', 'dorn-dräo', 'dorn-drä-il', 'drä-ling', 'drä-t', 'wefal-drä-t', 'gi-drä-ti', 'drä-unga', and 'gi-drä-unga'. Each entry includes a list of related forms and their meanings, such as 'dräen' meaning 'drehen, dreheln' and 'dräo' meaning 'Drechsler'.

Abb. 1 : Screenshot AWB^{Sp1} (2009) Online im Wörterbuchnetz (interner Prototyp)

Entsprechend kann nach desubstantivischen Nomina agentis und nach Bildungen mit deadjektivischen Erstgliedern gesucht werden. Die Trefferlisten sind natürlich einer genauen Kontrolle zu unterziehen und ggf. zu bereinigen, z. B. um die nicht-persönlichen Bildungen (Nomina instrumenti, Tier- und Pflanzennamen usw.). Außerdem sind unter Umständen Formeln für besondere weitere Bildungsweisen einzubeziehen. Für eine nähere Untersuchung könnten in einer Pilotstudie die Strukturformeln um die Unterscheidung der persönlichen und nicht-persönlichen Bildungen angereichert werden. Solche Zusatzinformationen können später in der Forschungsumgebung als optionales Angebot zur Verfügung gestellt werden.

Für den Vergleich mit dem Mittelhochdeutschen und Mittelniederdeutschen ergibt sich durch die Nebensilbenabschwächung die Schwierigkeit des Zusammenfalls verschiedener älterer Suffixe. Im Falle der hier interessierenden agentiven schwachen Maskulina und Feminina des besprochenen alten Bildungstyps werden die ahd. Ausgänge *-o* und *-a* zu mhd. *-e* abgeschwächt. Das mhd. Suffix *-e* vereinigt nach Ring (2008, S. 40–118) und KSW III (2009, S. 61–72) mindestens acht verschiedene Funktionsklassen, von denen die deverbale Agentiva nur eine ausmachen (Ring *-e*³, KSW III *-e*⁴), die zudem wiederum auch nicht-persönliche Agentiva enthält. Hierzu könnte ähnlich wie bei den personalen und nicht-personalen Nomina agentis des Althochdeutschen in einer Detailstudie die genauere Unterscheidung der Funktionsklassen in die Formeln eingebracht und in der Forschungsumgebung zusätzlich zur Verfügung gestellt werden.

4.2.2 Funktionsgleiche Movierungssuffixe

Das Mittelniederdeutsche weist in mehreren Bereichen Affixkonvergenz auf, d. h., es gibt mehrere formal unterschiedliche Affixe, die (in etwa) die gleiche Funktion erfüllen (vgl. Barz 2021). Am Beispiel der frequentesten Movierungssuffixe soll gezeigt werden, wie WoDia für Untersuchungen solcher Konkurrenz Kandidaten eingesetzt werden kann.

Im Neuhochdeutschen wird die Femininmovierung durch das Derivationssuffix *-in* dominiert, das z. B. aus dem *Koch* eine *Köchin* macht. Weitere Suffixe treten nur sehr eingeschränkt auf (vgl. Fleischer/Barz 2012, S. 239). In den älteren Sprachstufen des Deutschen finden sich hingegen mehrere Derivationssuffixe, die nebeneinander stehen.

Im Altsächsischen sind die Suffixe *-in* und *-inne* belegt, die an deverbale Nomina agentis gehängt werden, z. B. *mak-en* mit der Formel $wV > mak-āri$ mit der Formel $(wV)sS > mak-ār-in$ mit der Formel $((wV)sS)sS$ (vgl. Lasch 1914, § 213, 372; Hucko 1904, S. 72 f.; Wilmanns 1899, S. 310–312). Die Distinktion zwischen der kürzeren Form *-in* und der längeren Form *-inne* ist hierbei auf eine morphophonologische Differenzierung zwischen dem Nominativ und den obliquen Kasus zurückzuführen. Ein weiteres Movierungssuffix ist *-sche*, dessen Ursprung auf lat. *-issa* bzw. frz. *-esse* zurückgeht.

Im Mittelniederdeutschen sind die beiden frequentesten Movierungssuffixe *-inne* und *-sche* (vgl. Seelmann 1921).¹⁴ Werth (2022, S. 3, 2015, S. 61) stellt in norddeutschen Hexenverhörprotokollen des 16. und 17. Jahrhunderts eine funktionale Differenzierung dieser beiden fest, wobei *-inne* das Lemma semantisch um das Merkmal [+weiblich] erweitert und *-sche* eine onymische Zugehörigkeitsrelation („Ehefrau von XName“) etabliert.

Für eine umfassende Untersuchung lässt sich mit Hilfe von WoDia eine Trefferliste aller Movierungen anzeigen.¹⁵ Das Suchergebnis in WoDia könnte wie in Abbildung 2 aussehen.



The screenshot shows the WoDia search interface. At the top, there's a search bar with the query '(-sche|-inne) & .sS'. Below the search bar, there's a table of results. The table has columns: Sprachstufe, Lemma, Gramm. Angabe, Bedeutungsangabe, Segmentierung, Strukturformel, and Wortfamilie. The results are for various lemmas in Mittelniederdeutsch, including kökinne, köksche, mörderische, mörderinne, vorköpersche, vorköperinne, and bindersche. Each row shows the lemma, its grammatical gender, its meaning, its segmentization, its structural formula, and its word family.

Sprachstufe	Lemma	Gramm. Angabe	Bedeutungsangabe	Segmentierung	Strukturformel	Wortfamilie
Mittelniederdeutsch	kökinne	f.	Köchin	kök-inne	((wV)sS)sS	² KÖKEN
Mittelniederdeutsch	köksche	f.	Köchin	kök-sche	((wV)sS)sS	² KÖKEN
Mittelniederdeutsch	mörderische	f.	Mörderin	mörd-er-sche	((wV)sS)sS	MÖRDEN
Mittelniederdeutsch	mörderinne	f.	Mörderin	mörd-er-inne	((wV)sS)sS	MÖRDEN
Mittelniederdeutsch	vorköpersche	f.	Verkäuferin	vor-köp-er-sche	((p(wV))sS)sS	KÖPEN
Mittelniederdeutsch	vorköperinne	f.	Verkäuferin	vor-köp-er-inne	((p(wV))sS)sS	KÖPEN
Mittelniederdeutsch	bindersche	f.	Binderin	bind-er-sche	((wV)sS)sS	BINDEN

Abb. 2: Suche in der WoDia-Forschungsumgebung am Beispiel der Movierung¹⁶

Durch die Auszeichnung der Wortbildungsstruktur kann direkt nach den beiden Suffixen gesucht werden, formgleich endende Simplizia müssen nicht mehr aussortiert werden. Die dargestellte Abfrage liefert 126 Lemmata mit *-inne* und 165 Lemmata mit *-sche*. Bei jeweils dreien davon liegt die Kombination *-ster-inne* bzw. *-ster-sche* vor. Nomina agentis auf *-er* liegen bei 50 der *-inne*-Bildungen und bei 121 *-sche*-Bildungen zugrunde.¹⁷ Insgesamt 30

14 Außerdem finden sich die weniger frequente, kurze Form *-in* sowie, ebenfalls weit weniger verbreitet, das aus dem Mittelniederländischen stammende *-(e)ster* (auch in Kombination mit *-inne* oder mit *-sche*) (vgl. Seelmann 1921). Nur bei Familiennamen dienen überdies *-s* und *-sen* zur weiblichen Sexusmarkierung (vgl. Werth 2015, S. 57–58).

15 So wird die Suche nach Lemmata, die auf *-sche* oder *-inne* enden und deren Strukturformel auf *sS* (= Substantivsuffix) ausgeht, möglich sein.

16 Es handelt sich um einen ersten Entwurf für das Design der Forschungsumgebung, das im weiteren Verlauf der Projektarbeiten noch bearbeitet wird.

17 Die Zahlen basieren auf einem präfinalen Stand der Lemmaliste, bei der neben der im MNWB noch nicht bearbeiteten Strecke einzelne weitere Lemmata fehlen können, da Nestartikel sowie bestimmte Verweise in der Auszeichnung aktuell noch nicht korrekt erfasst sind. In der finalen Fassung der Datenbank können die Trefferzahlen daher abweichen.

verschiedene Stämme liegen sowohl mit *-inne* als auch mit *-sche* vor (vgl. *kōkinne/kōksche* ‚Köchin‘ in Abb. 2). Die in Abbildung 2 ausgewählten Beispiele gehören (ihrer Bedeutungsangabe im MNWB nach) alle zum ersten Typ, bei dem die Basis um das Merkmal [+weiblich] ergänzt wird. Da im Wörterbuch Eigennamen nur in Ausnahmefällen aufgenommen sind, lässt sich die von Werth als onymischer Typ bezeichnete Movierungsform vor allem in Kombinationen mit Nomina agentis wie Berufsbezeichnungen, z. B. *vinstermēkersche* ‚Frau des Fenstermakers‘, finden. Eine eingehende Untersuchung der Movierung unter Nutzung der Wörterbuchdaten einerseits (Wortbildungsstruktur, Basen, Verbindung mit *-er*) und des über die Lemmaliste verknüpften Referenzkorpus Mittelniederdeutsch/Nieder-rheinisch (räumliche und zeitliche Verbreitung, kontextuelle Einflüsse) andererseits ist derzeit als Fallstudie in Vorbereitung.

5. Fazit

Im Projekt ‚Wortfamilien diachron‘ (WoDia) entsteht eine digitale Forschungsumgebung, die für Untersuchungen zur historischen Wortbildung des Deutschen eine umfassende Materialbasis zur Verfügung stellt. Sie bietet für alle Lemmata der Referenzwörterbücher des Althochdeutschen, Altsächsischen, Mittelhochdeutschen und Mittelniederdeutschen eine wortbildungsmorphologische Analyse und Wortfamilienzuordnung, die auf der Segmentierung der Lemmata und der Rückführung der dabei freigelegten Stämme auf ihren Wortfamilienstamm beruht. Das Ergebnis der Analyse wird in einer hierarchischen Strukturformel angegeben, in der die Konstituentenstruktur, die Wortart der Stämme und die Wortbildungsart notiert sind. Um varietätenübergreifende Analysen zu ermöglichen, sind die vier Wortschätze sowohl auf der Ebene der Einzellemmata als auch über die Wortfamilien miteinander verbunden. Alle Lemmata sind mit den entsprechenden Artikeln in ihren Referenzwörterbüchern verknüpft, darüber hinaus werden sie mit den Referenzkorpora zur deutschen Sprachgeschichte verlinkt.

Das Potenzial der in WoDia aufbereiteten Daten wurde am Beispiel dreier wortbildungsmorphologischer Phänomene in verschiedenen Sprachstufen verdeutlicht, der Doppelsuffigierung von Adjektiven, der Suffixablösung bei den Nomina agentis und der funktionsgleichen Movierungssuffixe. Über die Suche nach Strukturformeln eines bestimmten Typs in der Forschungsumgebung können die betreffenden Lemmata zusammengestellt werden. Je nach Phänomen und konkreter Forschungsfrage sind weitere Annotationen anzuschließen, z. B. im Fall der Nomina agentis zur Differenzierung zwischen persönlichen und nicht-persönlichen Bildungen oder bei den Movierungssuffixen zur Unterscheidung verschiedener Funktionstypen. WoDia ermöglicht Detailstudien nicht nur zu Wortbildungsstrukturen, Wortbildungsmitteln und deren Wandel im älteren Deutschen, sondern auch zu Veränderungen im Bereich der Wortfamilien, beispielsweise durch einen diachronen Ausbau bestimmter Wortfamilien. Die Verknüpfung zwischen den vier Wortschätzen erlaubt zudem varietätenübergreifende Untersuchungen, sodass im Ergebnis ein umfassenderes Bild der Entwicklungen im Bereich des Wortschatzes und der Wortbildung im historischen Deutschen gezeichnet werden kann.

Literatur

Wörterbücher

ASWB = Tiefenbach, Heinrich (2010): Altsächsisches Handwörterbuch. Berlin/New York: De Gruyter.

AWB = Sächsische Akademie der Wissenschaften zu Leipzig (2021): Althochdeutsches Wörterbuch. Auf Grund der von Elias von Steinmeyer hinterlassenen Sammlungen im Auftrag der Sächsischen Akademie der Wissenschaften zu Leipzig. Bd. 8,1: S – Sn-. Berlin/Boston: De Gruyter.

AWB^{Spl} = Splett, Jochen (1993): Althochdeutsches Wörterbuch. Analyse der Wortfamilienstrukturen des Althochdeutschen. Zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes. Bd. 1–3. Berlin/New York: De Gruyter.

Lexer = Lexer, Matthias (1992): Mittelhochdeutsches Handwörterbuch. Zugleich als Supplement und alphabetischer Index zum Mittelhochdeutschen Wörterbuche von Benecke-Müller-Zarncke. Bd. 1–3. Stuttgart: Hirzel. [Nachdruck der Ausgabe Leipzig 1872–1878].

LW = Lübben, August/Walther, Christoph (1888): Mittelniederdeutsches Handwörterbuch. (= Wörterbücher 2). Leipzig: Soltan.

MNWB = Lasch, Agathe/Borchling, Conrad (1956–): Mittelniederdeutsches Handwörterbuch. Fortgef. von Gerhard Cordes und Dieter Möhn. Hg. von Ingrid Schröder. Bd. 1 ff. Zuletzt Bd. 3, Lfg. 44., 2023. Kiel/Hamburg: Wachholtz.

MWB = Akademie der Wissenschaften und der Literatur | Mainz und der Niedersächsischen Akademie der Wissenschaften zu Göttingen (2006–): Mittelhochdeutsches Wörterbuch. Zuletzt Bd. 3, Lfg. 3, 2024. Stuttgart: Hirzel.

Splett, Jochen (2009): Deutsches Wortfamilienwörterbuch. Analyse der Wortfamilienstrukturen der deutschen Gegenwartssprache. Zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes. Bd. 1–18. Berlin/New York: De Gruyter.

Forschungsliteratur

Barz, Irmhild (2021): Affixkonvergenz. www.degruyterbrill.com/database/WSK/entry/wsk_id_wsk_artikel_artikel_17524/html (Stand: 1.7.2025).

Duden (2016): Der Duden in zwölf Bänden. Bd. 4: Die Grammatik. Unentbehrlich für richtiges Deutsch. 9., vollst. überarb. u. aktual. Aufl. Berlin: Dudenverlag.

Duden (2022): Der Duden in zwölf Bänden. Bd. 4: Die Grammatik. Unentbehrlich für richtiges Deutsch. 10., völlig neu verf. Aufl. Berlin: Dudenverlag.

Eisenberg, Peter (2020): Grundriss der deutschen Grammatik. Das Wort. 5., aktual. und überarb. Aufl. Stuttgart: Metzler.

Erben, Johannes (2006): Einführung in die deutsche Wortbildungslehre. (= Grundlagen der Germanistik 17). 5., durchges. u. erg. Aufl. Berlin: ESV.

Fleischer, Wolfgang/Barz, Irmhild (2012): Wortbildung der deutschen Gegenwartssprache. 4., völlig neu bearb. Aufl. Berlin/Boston: De Gruyter.

Henzen, Walter (1965): Deutsche Wortbildung. 3., durchges. u. erg. Aufl. (= Sammlung kurzer Grammatiken germanischer Dialekte B 5). Tübingen: Niemeyer.

Hucko, Matthias (1904): Bildung der Substantiva durch Ableitung und Zusammensetzung im Altsächsischen. Straßburg: DuMont-Schauberg.

Ihden, Sarah/Schröder, Ingrid (2021): Mittelniederdeutsche Grammatik: Konzeption und erste Analysen. In: Jahrbuch des Vereins für niederdeutsche Sprachforschung 144, S. 79–104.

Ihden, Sarah/Schröder, Ingrid (2024): Das Referenzkorpus Mittelniederdeutsch/Niederrheinisch als Quelle für wortbildungsmorphologische Studien. Zur Varianz des Suffixes *-schop*. In: Coniglio, Marco/Recker, Anabel/Sahm, Heike (Hg.): Mittelniederdeutsch zwischen Korpuslinguistik und Literaturwissenschaft. Göttingen: Universitätsverlag, S. 15–43.

Ihden, Sarah/Schnelle, Gohar/Schröder, Ingrid/Zeige, Lars E. (2023): Der Verbund ‚Deutsch Diachron Digital – Referenzkorpora zur deutschen Sprachgeschichte‘. Strategien der Erschließung, Analyse und nachhaltigen Nutzung historischer Sprachdaten. In: Kupietz, Marc/Schmidt, Thomas (Hg.): Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022. (= Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache/Corpus Linguistics and Interdisciplinary Perspectives on Language 11). Tübingen: Narr, S. 11–31.

Klein, Thomas (2013): Verknüpfung digitaler Lemmalisten historischer Sprachstufen des Deutschen – wie und wozu? www.uni-trier.de/fileadmin/forschung/maw/MWB/Arbeitsgesprach2013/Vortrag_Bullay_Klein.pdf (Stand: 1.7.2025). [Präsentationsfolien des Vortrags im Rahmen des Arbeitsgesprächs zur historischen Lexikographie vom 12. bis 14. April 2013 in Bullay, 13.4.2013].

Klein, Thomas (2018): Mittelhochdeutsche Wortfamilien: Ermittlung und Perspektiven. In: Zeitschrift für Wortbildung 2, 1 (Sonderheft: Wortbildung – historisch, mehrsprachig, kontrastiv), S. 11–31. <https://journals.linguistik.de/zwjw/article/view/18/103> (Stand: 8.2.2026).

Krahe, Hans (1969): Germanische Sprachwissenschaft. Bd. 3: Wortbildungslehre. 7. Aufl. (= Sammlung Götschen 2234). Berlin: De Gruyter.

KSW III = Klein, Thomas/Solms, Hans-Joachim/Wegera, Klaus-Peter (2009): Mittelhochdeutsche Grammatik. Bd. 3: Wortbildung. Tübingen: Niemeyer.

Möhn, Dieter/Schröder, Ingrid (2009): Lexembildung im Aufriss einer Grammatik des Mittelniederdeutschen. Das Adjektiv als Exempel. In: Lenz, Alexandra N./Gooskens, Charlotte/Reker, Simon (Hg.): Low Saxon Dialects across Borders – Niedersächsische Dialekte über Grenzen hinweg. (= Zeitschrift für Dialektologie und Linguistik Beiheft 138). Stuttgart: Steiner, S. 38–59.

Plate, Ralf (2020): Computergestützte Etablierung epochenübergreifender Wortfamilienstrukturen. Abschlussbericht. <https://zhistlex.de/ziele/wortfamilien/> (Stand: 1.7.2025).

Plate, Ralf (2022): Word families in diachrony. An epoch-spanning structure for the word families of older German. In: Klosa-Kückelhaus, Anette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (Hg.): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. 12–16 July 2022, Mannheim, Germany. Mannheim: Leibniz-Institut für Deutsche Sprache, S. 605–613. <https://doi.org/10.14618/ids-pub-11253> (Stand: 1.7.2025).

Ring, Uli (2008): Substantivderivation in der Urkundensprache des 13. Jahrhunderts. Eine historisch-synchrone Untersuchung anhand der ältesten deutschsprachigen Originalurkunden. (= Studia Linguistica Germanica 96). Berlin: De Gruyter.

Schröder, Ingrid (2022): Stand und Perspektiven mittelniederdeutscher Grammatikographie am Beispiel der expliziten Derivation. In: Denkler, Markus/Mähl, Stefan (Hg.): Beiträge zur historischen Wortbildung des Niederdeutschen. (= Niederdeutsche Studien 61). Wien/Köln: Böhlau, S. 13–36.

Seelmann, Wilhelm (1921): Mittelniederländische Wörter in der Mark Brandenburg. In: Jahrbuch des Vereins für niederdeutsche Sprachforschung 47, S. 40–45.

Splett, Jochen (1995): Der Adjektivtyp ((...)sA)sA im Althochdeutschen. In: Cajot, José/Kremer, Ludger/Niebaum, Hermann (Hg.): Lingua theodisca. Bd. 1: Beiträge zur Sprach- und Literaturwissenschaft. Jan Goossens zum 65. Geburtstag. (= Niederlande-Studien 16.1). Münster u. a.: Lit, S. 89–101.

Wilmanns, Wilhelm (1899): Deutsche Grammatik. Gotisch, Alt-, Mittel- und Neuhochdeutsch. 2. Aufl. Bd. 2: Wortbildung. Straßburg: Trübner.

Werth, Alexander (2015): *Gretie Dwengers*, genannt *die Dwengersche*. Formale und funktionale Aspekte morphologischer Sexusmarkierung (Movierung) in norddeutschen Hexenverhörprotokollen der Frühen Neuzeit. In: Jahrbuch des Vereins für niederdeutsche Sprachforschung 138, S. 53–75.

Werth, Alexander (2022): Der Abbau der onymischen Movierung im 18. Jahrhundert: Eine Auswertung des Deutschen Textarchivs. In: Havinga, Anna D./Lindner-Bornemann, Bettina (Hg.): Deutscher Sprachgebrauch im 18. Jahrhundert: Sprachmentalität, Sprachwirklichkeit, Sprachreichtum. (= Germanistische Bibliothek 71). Heidelberg: Winter, S. 93–113.

Online-Ressourcen

Deutsch Diachron Digital (o.D.): Referenzkorpora zur deutschen Sprachgeschichte. www.deutschdiachrondigital.de/ (Stand: 1.7.2025).

MWB Online (o.D.): Mittelhochdeutsches Wörterbuch. www.mhdwb-online.de/ (Stand: 1.7.2025).

Sächsische Akademie der Wissenschaften zu Leipzig (o.D.): Althochdeutsches Wörterbuch. <https://awb.saw-leipzig.de/AWB> (Stand: 1.7.2025).

Trier Center for Digital Humanities (o.D.): Wörterbuchnetz. Kompetenzzentrum. www.woerterbuchnetz.de (Stand: 1.7.2025).

Visser, Lourens (o.D.): Lautkorrespondenzen MHD-MND. <https://osf.io/yvfjh> (Stand: 1.7.2025).

Kontaktinformation

Thomas Burch
Kompetenzzentrum – Trier Center for Digital Humanities
Universität Trier
54286 Trier
E-Mail: burch@uni-trier.de

Jost Gippert
Centre for the Study of Manuscript Cultures
University of Hamburg
Warburgstr. 26
20354 Hamburg
E-Mail: jost.gippert@uni-hamburg.de

Sarah Ihden
Institut für Germanistik
Universität Rostock
Universitätsplatz 3
18055 Rostock
E-Mail: sarah.ihden@uni-rostock.de

Sarah Kwekkeboom
DFG-Projekt ‚Wortfamilien diachron‘
Institut für Empirische Sprachwissenschaft
Goethe-Universität
Rostocker Str. 2
60323 Frankfurt a. M.
E-Mail: sarah.kwekkeboom@adwmainz.de

Ralf Plate
Mittelhochdeutsches Wörterbuch,
Arbeitsstelle der Akademie der Wissenschaften und der Literatur | Mainz
Universität Trier
Universitätsring 15
54286 Trier
E-Mail: plate@uni-trier.de

Lena Schnee
DFG-Projekt ‚Wortfamilien diachron‘
Institut für Germanistik
Universität Hamburg
Von-Melle-Park 6
20146 Hamburg
E-Mail: lena.schnee@uni-hamburg.de

Ingrid Schröder
Institut für Germanistik
Universität Hamburg
Von-Melle-Park 6
20146 Hamburg
E-Mail: ingrid.schroeder@uni-hamburg.de

Roland Schuhmann
DFG-Projekt ‚Wortfamilien diachron‘
Institut für Empirische Sprachwissenschaft
Goethe-Universität
Rostocker Str. 2
60323 Frankfurt a. M.
E-Mail: schuhmann@em.uni-frankfurt.de

Johanna Wittmack
DFG-Projekt ‚Wortfamilien diachron‘
Institut für Germanistik
Universität Rostock
Universitätsplatz 3
18055 Rostock
E-Mail: johanna.wittmack@uni-rostock.de

Bibliografische Angaben

Dieser Text ist Teil der Publikation: Brunner, Annelen/Hansen, Sandra/Lang, Christian/Tu, Ngoc Duyen Tanja/Wolfer, Sascha (Hg.) (2026): Deutsch im Wandel – Tools, Methoden, Ressourcen. Beiträge zur Methodenmesse der IDS-Jahrestagung 1. (= *IDSopen* 16). Mannheim: IDS-Verlag. <https://doi.org/10.21248/idsopen.16.2026.63>.