

Schlussbericht des Verbund-Projekts ZHistLex

(„eHumanities-Zentrum für Historische Lexikographie“)

Förderkennzeichen: 01UG1603A-F

Datum: 28.3.2020

VerfasserInnen: Burch, Thomas; Gloning, Thomas; Harm, Volker; Herold, Axel; Hoenen, Armin; Plate, Ralf; Recker-Hamm, Ute

I. Kurze Darstellung zu Aufgabenstellung, Voraussetzungen, Planung und Ablauf, Anknüpfungspunkte, Zusammenarbeit

I.1 Aufgabenstellung

Historische Wörterbücher und historisch-lexikologische Darstellungen zur Entwicklung des Wortschatzes und des Wortgebrauchs in historischer Perspektive sind zentrale wissenschaftliche Darstellungs- und Dokumentationsformen im Schnittbereich zwischen Philologie, historischer Sprach-, Literatur- und Kulturwissenschaft sowie je nach Gegenstandsbereich auch den Forschungsdisziplinen, die auf die sprachliche Verfasstheit einzelner historischer Gegenstandsbereiche, ihrer geschichtlichen Entwicklung und ihrer Erforschung bezogen sind.

Beim Übergang ins digitale Zeitalter waren im Bereich der historischen Lexikographie und Lexikologie vor allem zwei zeitlich versetzte Strategien zu beobachten. Zunächst wurden die Informationsbestände der vorhandenen Wörterbücher retrodigitalisiert (entweder als Volltext oder als Bilder), die digitalen Versionen waren zum Teil direkte Gegenstücke zu den gedruckten Fassungen. Eine zweite Strategie bestand und besteht darin, vorhandene und hinzukommende Daten mit neuen Darstellungs-, Interaktions- und Vernetzungsmöglichkeiten anzureichern, die durch digitale Techniken ermöglicht werden. Hierzu gehören unter anderem erweiterte Suchoptionen, Ein- und Ausblend-Techniken bei der Artikeldarstellung, die Verbindung von Wortartikeln zu den Korpus-texten oder auch die Anreicherung mit neuen Komponenten, die dynamisch aus dem verfügbaren Datenbestand erzeugt werden, z.B. visuelle Zeitleisten.

Ein wesentlicher Unterschied zur gegenwartssprachlichen Lexikographie in digitalen Umgebungen ist der historisch-dynamische Charakter des Gegenstandes, also des histo-

rischen Wortgebrauchs in seiner Entwicklung in der Zeit. Daraus ergeben sich auch besondere Anforderungen und Aufgaben. In Bezug auf digitale Umgebungen stellen sich Leitfragen wie die folgenden:

Welche (neuen) Handlungsoptionen ergeben sich in digitalen Umgebungen

- für die Beschreibung und Dokumentation der Bedeutungsentwicklung;
- für Formen der Belegdokumentation;
- für die Charakterisierung der Gebrauchsdynamik von Wörtern;
- für die Erfassung von Zusammenhängen im Wortschatz und ihrer Entwicklung;
- die die Charakterisierung der Variation im Wortgebrauch und ihrer Entwicklung;
- usw.

Neben der Spezifik des Gegenstandsbereichs ist auch die wissenschaftliche Organisation der historischen Lexikographie des Deutschen ein wichtiger Bedingungsfaktor. Die großen historischen Wörterbücher, z.B. die Sprachstadienwörterbücher und historische Spezialwörterbücher wie das Goethe-Wörterbuch oder das Deutsche Rechtswörterbuch, sind an Akademien verortet. Historische Wort(schatz)forschung findet darüber hinaus auch an Universitäten und an eigenen Zentren, wie dem Trierer Kompetenzzentrum (TCDH), statt.

Die **Ziele** des Projekts lassen sich in drei Dimensionen charakterisieren (vgl. Antrag, Abschnitt 6): (i) Beiträge zu einer integrierten Dateninfrastruktur im Bereich der digitalen historischen Lexikographie; (ii) Konzeptionelle Systematisierung und Erweiterung von digital unterstützten Untersuchungs-, Erschließungs- und Präsentationsverfahren; (iii) Kollaborationsstruktur, in der Akteure aus den Arbeitsstellen der Akademien, der Universitäten und des Trierer Kompetenzzentrums gemeinsam an der Weiterentwicklung historisch-lexikographischer Darstellungs- und Dokumentationsformen arbeiten.

Diese Weiterentwicklungsziele umfassten im Wesentlichen folgende **Aufgaben**, denen spezifische Arbeitspakete in der Projektplanung zugewiesen waren.

AP1 Projektkoordination

AP2 Standards und Webservices

AP3 Exemplarische Aufbereitung von Wörterbuchressourcen

AP4 Metadatenschema für ein übergreifendes Quellenverzeichnis

AP5 Exemplarische bidirektionale Verknüpfung von Wörterbuch-Belegen mit digitalen Volltexten

AP6 Epochenübergreifende Wortfamilienstrukturen

AP7 Innovative Zugriffs- und Darstellungsformen

Diese Aufgaben und die darauf bezogenen Arbeitspakete sind nicht isoliert, sondern weisen enge Zusammenhänge auf. Sie werden ggf. im Rahmen der Ergebnisdarstellung im Abschnitt II.1 erläutert.

I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Literaturlage und metalexikographischer Forschungsstand

Zur gegenwartssprachlichen Lexikographie in digitalen Umgebungen gibt es inzwischen einen breiten Forschungsstand und eine gute Literaturlage, vgl. u.a. Klein 2004; Klein/Geyken 2010; Granger/Paquot 2012; Klosa 2011; Schnörch 2011; vgl. weiter die Informationen zum Netzwerk „Internetlexikografie“ unter www.internetlexikografie.de. Im Jahr 2013 wurden diese Entwicklungen in einem Zusatzband zum bisher dreibändigen Handbuch „Dictionaries“ der HSK-Reihe dokumentiert, der zwar den Untertitel „Recent Developments with Focus on Electronic and Computational Lexicography“ aufweist, im Hinblick auf digitale Entwicklungen in der historischen Lexikografie aber keinen eigenen Überblicksbeitrag enthält. Der Beitrag 48 zu den historischen und etymologischen Wörterbüchern des Deutschen streift einzelne digitale Entwicklungen in diesem Bereich nur ganz kurz. Älter sind Traditionen, in denen elektronische Unterstützungsmöglichkeiten bei der Bearbeitung historischer Wörterbücher behandelt werden, z.B. in mehreren Bänden der Reihe „Maschinelle Verarbeitung altdeutscher Texte“, vgl. weiterhin auch Formen der Begleitreflexion zur Arbeit in den einzelnen Wörterbuch-Arbeitsstellen, z.B. Plate 2007.

Der Stand der technisch-informatischen Entwicklung

Der Stand der technisch-informatischen Entwicklung ist bestimmt durch eine beträchtliche Heterogenität der Ansätze und der technischen Lösungen. Während sich eine XML-Kodierung im Rahmen des TEI-P5-Standards durchgesetzt hat, erlaubt dieser Standard doch eine Fülle von „Dialekten“, die eine interoperable Ansprechbarkeit von Wörterbuchdaten erschwert. Auch die technischen Grundlagen variieren je nach Standort. Hier sollte im Projekt vor allem die Entwicklung von Webservices die Interoperabilität von digitalen Wörterbuchangeboten ermöglichen bzw. verbessern.

I.3 Planung und Ablauf des Vorhabens

Die Planungen des Vorhabens, seine Arbeitspakete und die zeitlichen Abläufe, sind im Projektantrag dokumentiert. Die Planungen wurde inhaltlich und auch in zeitlicher Hinsicht im Wesentlichen erfüllt. Wenige Abweichungen werden im Abschnitt II. bei den Resultaten der einzelnen Arbeitspakete erwähnt und erläutert.

I.4 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Die wesentlichen Bezugspunkte für unsere Arbeit waren:

- Fachliteratur zu den Bereichen: Lexikographie, historische Lexikographie, digitale/elektronische Lexikographie, (historische) Semantik, (historische) Lexikologie und zu spezifischen Bereichen wie z.B. Informationsvisualisierung;
- der technisch-informatische Stand im Bereich der XML-Kodierung, der Datenhaltung, der Gestaltung von Webservices, der Organisation von Abfrage-Szenarien;
- verfügbare Standards, die bereits verfügbar waren (v.a. TEI-P5), und Standards, die im Projektzeitraum weiterentwickelt wurden, besonders relevant war der neue gemeinsame Standard für die Basiskodierung von Wörterbüchern, TEI Lex0;
- der Diskussionsstand zu spezifischen Teilaspekten der historischen Sprachwissenschaft, die sich auf den Wortgebrauch, den Wortschatz sowie seine Organisation und seine Geschichte beziehen (z.B. Arbeiten zur Modellierung der Wortschatzstruktur, zu Wortfamilien und ihrer historischen Entwicklung);

Wichtig waren darüber hinaus die Erfahrung und die laufende Praxis des Wörterbuchmachens und -digitalisierens, die an den beteiligten Wörterbuchstandorten (Mainz, Trier, Göttingen, Berlin) vorlagen.

I.5 Zusammenarbeit mit anderen Stellen

Die wesentliche Form der Zusammenarbeit war projektintern, also zwischen den Verbundpartnern. Darüber hinaus sind drei weitere Einrichtungen zu nennen:

- das Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim: Hier fanden ein regelmäßiger fachlicher Austausch und zwei gemeinsame Veranstaltungen statt;
- das „Arbeitsgespräch zur historischen Lexikographie“ in Bullay, das von ZHistLex-Mitglied Ralf Plate jährlich in Bullay organisiert wird: Hier wurden bei jedem Jahrestreffen auch ZHistLex-Ergebnisse vorgestellt und diskutiert;
- das „Zentrum für digitale Lexikographie“ (ZDL), an dem mit der Berlin-Brandenburgischen und der Göttinger Akademie der Wissenschaften sowie Thomas Gloning (Uni Gießen) drei ZHistLex-Projektpartner beteiligt sind. Die Ergebnisse des ZHistLex-Projekts werden direkt relevant sein für die Ansprache historischer Datenbestände im ZDL-Portal.

Teil II: Eingehende Darstellung zu den Arbeitspaketen

Im Folgenden berichten wir über die in den einzelnen Arbeitspaketen erzielten Ergebnisse, dabei gehen wir jeweils auch auf die durchgeführten Arbeitsschritte und den Nutzen ein. Die erfolgten und noch geplanten Veröffentlichungen sowie die Vorträge und Präsentationen dokumentieren wir im Abschnitt II.8 und ordnen sie den Arbeitspaketen und ggf. auch übergeordneten Gesichtspunkten zu.

II.1 AP1 Projekt-Koordination (Leitung: Uni Gießen)

Im Rahmen der Projekt-Koordination wurden folgende Aufgaben erfüllt: Organisation von Videokonferenzen und von persönlichen Treffen; Organisation des Aufbaus einer Projekt-Webseite (zhistlex.de) und einer redmine-Umgebung für die interne Dokumentation (an den Arbeiten selbst waren dann alle Projektpartner beteiligt); Kommunikation mit dem Projektträger; Organisation der projektinternen Kommunikation (u.a. Mailingliste); Organisation des Berichtswesens.

II.2 AP2 Standards und Webservices

Leitung: BBAW Berlin

Beteiligte Projektpartner: AdW Göttingen, AdWL Mainz, Uni Frankfurt, Uni Trier

II.2.1 Standards

Im Bereich Standards führte die BBAW eine Sichtung und Bewertung der für die Implementation der technischen Infrastruktur für ZHistLex zu berücksichtigenden Standards in den Bereichen lexikalische Daten, Wörterbuchmetadaten, bibliografische Daten und Webservices durch.

Die lexikalischen Daten aller Projektpartner lagen bereits in unterschiedlichen Repräsentationen gemäß TEI (Text Encoding Initiative) vor, einem Standard-Framework innerhalb der Digital Humanities. Dieses Framework zeichnet sich – vor allem wegen seines universellen Anspruchs – durch eine hohe Variabilität hinsichtlich Textperspektiven, Annotationsebenen und Annotationstiefen aus. Dies spiegelte sich auch in der Heterogenität der Datenauszeichnung bei den Projektpartnern wider. Aus diesem Grund wurde bereits im ersten Projektjahr beschlossen, in Zusammenarbeit und Abstimmung mit der internationalen Initiative TEI Lex-0 (im Rahmen des DARIAH-Projektes, <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>) an der Standardisierung der Repräsentation lexikalischer Daten innerhalb des TEI-Frameworks

mitzuwirken. Die BBAW hat in der Folge durch einen Mitarbeiter für ZHistLex an zahlreichen TEI-Lex-0-Arbeitstreffen teilgenommen und diese Entwicklung aktiv begleitet und unterstützt. Ein Whitepaper zur Nutzung von TEI Lex-0 zur Auszeichnung von ZHistLex-konformen Wörterbuchdaten befindet sich auf der Projekthomepage unter <https://zhistlex.de/ziele/ap2/tei-lex-0/>. Auch über das Projektende hinaus wird die BBAW die Entwicklung von TEI Lex-0 im Hinblick auf ZHistLex-konforme Daten weiter unterstützen.

Nach Abstimmung mit AP4 werden in ZHistLex bibliografische Datensätze nach DublinCore zur Metadatenbeschreibung verwendet. Die Identifikation von Werken erfolgt über ihre PPN, wie sie im Göttinger Verbundkatalog verzeichnet sind. Eine eingehendere Darstellung hierzu findet sich im Abschnitt zu AP4.

II.2.2 Webservices

Die Spezifikation von ZHistLex-konformen Webservices erfolgte durch die BBAW in enger Abstimmung mit den Projektpartnern. Sie liegt in Form einer YAML-Spezifikation vor und ist in dieser Form auf der Projektwebseite interaktiv eingebunden:

- für Wörterbücher:
<https://zhistlex.de/ziele/ap2/webservices-woerterbuecher/interaktiv/>
- für Quellentexte:
<https://zhistlex.de/ziele/ap2/webservices-quellentexte/interaktiv/>
- für Literaturverzeichnisse:
<https://zhistlex.de/ziele/ap2/webservices-bibliographien/interaktiv/>

Die Spezifikation folgt für alle Ressourcentypen dem RESTful-Paradigma.

Wie vorgesehen erfolgt die Anpassung der lexikalischen Ressourcen an den Standorten der Projektpartner AdWL Mainz, Uni Trier und Uni Frankfurt zur Auslieferung über ZHistLex-konforme Webservices dynamisch. Dadurch wird sichergestellt, dass auch zukünftige Weiterentwicklungen der Webservice-Spezifikation von diesen Projektpartnern mit geringem Aufwand umgesetzt werden können.

Der Projektantrag sah die Spezifikation von Webservices in drei Bereichen vor: Wörterbuchartikel, bibliografische Daten und Korpustexte.

Für das Frankfurter Textarchiv Titus implementierte die Uni Frankfurt einen Webservice, der sowohl die Quellenbibliographie als auch erste Texte gemäß der ZHistLex-Spezifikation umsetzt. Auch nach Projektende werden noch weitere Texte über den Webservice zugänglich gemacht. Der Quellenservice wurde umfassend mit Bezug auf die im Lesekorpus Altdeutsch (LeA) verwendeten bibliografischen Ressourcen umgesetzt. Fragen der individuellen Umsetzung betrafen die Links zu den digitalen Textquellen wie dem Titusarchiv und LeA. Eine Dokumentation, die die Funktionsweise der Ti-

tusservices ist unter <http://titus.fkidg1.uni-frankfurt.de/database/titusinx/zhl/DocumentationZHLTitus.pdf> verfügbar. Darin werden auch die standortspezifischen Umsetzungen beschrieben und Interpretationsmöglichkeiten bestimmter nicht näher spezifizierbarer Details diskutiert.

Für das MWB wurden von der AdWL Mainz Webservices nach der ZHistLex-Spezifikation für das Wörterbuch und das Quellenverzeichnis implementiert. Ein erweiterter Service für Volltexte der Quelleneditionen ist vorbereitet und wird aus Eigenmitteln fertiggestellt. Der Endpoint der Service-Implementationen ist <http://tares.uni-trier.de/ZHistLex/API/>. Die implementierten Services werden bereits produktiv in der Wörterbucharbeit genutzt, nämlich im Wesentlichen durch das Redaktionssystem, indem es Informationen verschiedener Ressourcen dynamisch bündelt, und für die Vernetzung mit dem Wörterbuchnetz.

Die Implementierung der Webservices seitens der Uni Trier nach der ZHistLex-Spezifikation erfolgte durch das TCDH in Eigenleistung. Über den Endpoint unter der Basisadresse <http://urts55.uni-trier.de:8111/api> kann auf die Daten von AWB, BMZ, Lexer und ¹DWB zugegriffen werden. Im Rahmen der Beteiligung des TCDH am europäischen Verbundvorhaben ELEXIS (<http://www.elex.is/>) wird die Schnittstelle weiterentwickelt und gepflegt. Sie wird dort insbesondere zum Aufbau eines Wörterbuchviewers (vgl. Hierzu http://www.nytud.hu/kozerdeku/eu_palyazatok/grant_agreement_elexis.pdf, S. 46) als Verbindung zwischen Backend und Frontend genutzt.

BBAW, AdWL Mainz und Uni Trier werden auch über das Projektende hinaus eine Weiterentwicklung der ZHistLex-Spezifikationen anstreben. Die Spezifikation wird unter einer freien Lizenz unter der Adresse <https://github.com/zhistlex/> öffentlich gepflegt und weiterentwickelt.

Uni Frankfurt, AdWL Mainz und Uni Trier werden in Eigenleistung ihre ressourcenspezifischen API-Endpoints über das Projektende hinaus betreiben und weiterentwickeln.

II.2.3 Adressierung von Artikeln und Artikelteilen

Die Versionierung und Adressierung von Ressourcen wird in ZHistLex über die URLs der Webservices der Projektpartner erreicht. Detaillierte Beschreibungen für die Webservices der einzelnen Ressourcentypen und die realisierten Adressierungspfade sind auf der Projekthomepage in separaten Whitepapers durch die BBAW dokumentiert worden:

- Webservices für Wörterbücher:
<https://zhistlex.de/ziele/ap2/webservices-woerterbuecher/>
- Webservices für Quellentexte:
<https://zhistlex.de/ziele/ap2/webservices-quellentexte/>
- Webservices für Bibliographien:

<https://zhistlex.de/ziele/ap2/webservices-bibliographien/>

Insbesondere die Entwicklung der Zugriffsmöglichkeiten auf Quellentexte und Bibliographien erfolgte in enger Abstimmung mit den entsprechenden Arbeitspaketen 5 und 4.

II.2.4 Portalplanung

Auf der Projekthomepage <https://zhistlex.de> wurden verschiedene Möglichkeiten der Umsetzung einer zentralen Nutzerschnittstelle eruiert und zur Illustration ein einfacher Demonstrator von Uni Frankfurt und AdWL Mainz zur Quellenbibliographie entwickelt (<http://tares.uni-trier.de/ZhistLex/Quellenverzeichnisse>). Der Demonstrator führt eine gleichzeitige Anfrage an alle implementierten Services der Projektpartner durch, bereitet das Ergebnis auf und visualisiert die Überschneidung der Quellennutzung in den beteiligten Wörterbüchern. Der Demonstrator soll exemplarisch die Möglichkeit illustrieren, die Interaktion mit den ZHistLex-Webservices über Suchinterfaces auch menschlichen Nutzern zu erschließen. Er verwirklicht bewusst nur ein Basis-Konzept und realisiert nicht alle spezifizierten Suchoptionen. Er verdeutlicht aber, dass die gewählte Architektur geeignet ist, die Ressourcen der beteiligten Standorte interoperabel zu machen. Da alle Services ohne Zugangsbeschränkungen verfügbar sind, können von beliebiger Seite weitere Anwendungen auf ihrer Basis entwickelt werden (beispielsweise eine föderierte Suche über eine Menge von verschiedenen Wörterbüchern mit anschließenden komplexen Aggregationen, Synopsen und Darstellungen der Ergebnismengen).

In Kooperation mit AP7 (Innovative Zugriffs- und Darstellungsformen) wurden verschiedene Konzepte von allen beteiligten Partnern auf der Basis von antizipierten Forschungsfragen entwickelt und hinsichtlich ihrer technischen Realisierbarkeit bewertet.

Generell folgte ZHistLex der Maxime, die einzelnen Datengeber nicht zur strikten Umsetzung aller in der Spezifikation festgelegten Anforderungen zu verpflichten, sondern – vor allem aufgrund des heterogenen Charakters der verschiedenen Ausgangsmaterialien (in Bezug auf Art und Tiefe der vorhandenen Annotationen) – den Grad der Umsetzung ressourcenspezifisch festzulegen. Die implementierten Services werden an den jeweiligen Projektstandorten betrieben und die Pflege und Entwicklung der zugrundeliegenden Daten bleibt an diesen Standorten verortet. ZHistLex stellt eine gemeinsame Infrastruktur zur Abfrage und zum gezielten Zugriff auf die Daten der Projektpartner bereit.

Die ZHistLex-Homepage wurde bis Anfang 2019 von der Universität Gießen betrieben. Für den Langzeitbetrieb über das Projektende hinaus übernimmt die BBAW in Eigenleistung Hosting und Betreuung der Seite. In Kooperation mit Uni Frankfurt wurde

der Umzug umgesetzt. Damit fungiert die Projektseite nicht nur als Dokumentationsplattform für ZHistLex, sondern ebenso als Präsentationsplattform für die exemplarisch implementierten Infrastrukturkomponenten.

II.3 AP3 Exemplarische Aufbereitung von Wörterbuchressourcen

Leitung: Uni Trier (Trier Center for Digital Humanities)

Beteiligter Projektpartner: AdWL Mainz

II.3.1 Aufgaben und Ziele

Durch die an ZHistLex beteiligten Antragsteller konnten die im Rahmen des Akademienprogramms geförderten sprachhistorischen Großwörterbücher des Deutschen als zentrale digitale Ressourcen in das Zentrum eingebracht werden. Dies sind zum einen die Wörterbücher zu einzelnen Sprachstadien des Deutschen: das Althochdeutsche Wörterbuch (AWB), das Mittelhochdeutsche Wörterbuch (MWB), sowie zum anderen die Erst- (¹DWB) und Neubearbeitung (²DWB) des Deutschen Wörterbuchs von Jacob Grimm und Wilhelm Grimm als grundlegende Informationsquelle für die gesamte Sprachgeschichte des Deutschen vom Überlieferungsbeginn bis in die Gegenwart. Zusätzlich wurden die durch das TCHD im Rahmen eines DFG geförderten Vorhabens entstandenen digitalen Daten zu den großen mittelhochdeutschen Belegwörterbüchern (MWV) des 19. Jh.s von Benecke / Müller / Zarncke und von Lexer mit in dieses Arbeitspaket einbezogen.

Dieses Arbeitspaket diene der Aufbereitung dieser Ressourcen, so dass sie in die gemeinsame Infrastruktur als Primärdaten mit Metadaten und Identifikatoren integriert und über die im Rahmen von AP2 entwickelten Webservices adressiert, abgerufen, durchsucht oder verlinkt werden können. Im einzelnen sollten in Bezug auf die Wörterbuchdaten folgende Arbeitsschritte durchgeführt werden:

AWB: Die Daten des AWB lagen aus der Kooperation der SAW mit dem Trierer Kompetenzzentrum bereits sehr gut erschlossen im TEI-P5-Format vor. Weiterer inhaltlicher Erschließungsbedarf bestand vor allem in der Auflösung von Textverdichtungen für Suchzwecke, in der textstrukturellen Auszeichnung sowie in der Verknüpfung der Belege mit dem Quellenverzeichnis.

¹DWB: Hauptziel der inhaltlichen Kuration von ¹DWB war die Identifikation und Annotation von Prosabelegen (ca. 1.250.000) und deren Verknüpfung zum Quellenverzeichnis. Bereits annotiert waren die Lyrikbelege (ca. 385.000) und zugehörigen Siglen.

²DWB: Für die jüngeren Teile des Wörterbuchs standen die Satzdaten (PageMaker bzw. TUSTEP) zur Verfügung; dies betrifft Bd. VII, Lfg. 6-7, Bd. IX sowie die vor dem Abschluss stehenden Bände IV und V. Die in PageMaker erstellten Daten sollten über

eine Zwischenstufe im RTF-Format in TUSTEP konvertiert und anschließend die typographisch abgrenzbaren Informationseinheiten per Programmen mit TEI-P5-konformen XML-Kodierungen versehen werden. Schrittweise sollten dann weitere Informationspositionen, die aufgrund des Artikelaufbaus leicht identifizierbar sind (z. B. Genusangaben, etymologische Angaben, Bedeutungsteil, Kompositionsgruppen), markiert werden.

MWB: Durch das für das MWB verwendete Redaktionssystem, das der gedruckten wie der elektronischen Publikation zugrunde liegt, stehen weitgehend alle vorgesehenen inhaltlichen Informationspositionen eindeutig in XML-Kodierung zur Verfügung. Inhaltliche Aufbereitung war lediglich im Bereich der MWB-internen Verweise auf Artikel bzw. Artikelteile erforderlich, da sie in ihrer ursprünglichen Form nicht automatisch in Hyperlinks überführt werden können.

MWV: Die beiden im MWV enthaltenen Wörterbücher lagen als TEI-P2-codierte (im Wesentlichen noch nach SGML-basierenden Regeln) Daten vor. Die Überführung nach XML/TEI-P5 sollte insbesondere solche Stellen bearbeiten, an denen an strukturell relevanten Stellen Elemente zu erschließen sind, die nach dem früher angewandten SGML-Standard unterspezifiziert bleiben durften. Ferner sollten im Rahmen der Erstauszeichnung getroffene Entscheidungen an heutige Standards angepasst werden, indem Elemente des Wörterbuchttextes, z.B. Interpunktion zwischen Lemmavarianten oder bestimmten Artikelteilen nachgetragen werden. Elemente der Textverdichtung (z.B. Elision wiederholter Lemmakonstituenten) waren für die elektronischen Nutzungszwecke des Zentrums wiederherzustellen.

Wörterbuchverlinkung: Ein zentrales Ziel des Zentrums war es, die Voraussetzungen für eine umfassende Vernetzung der Wörterbücher untereinander zu schaffen, um so die Basis zu einem vernetzten Informationssystem zum historischen Deutsch zu legen. Dies betraf einerseits die Auszeichnung der in den Wörterbüchern zahlreich explizit angelegten Verweise und andererseits im Zusammenhang mit AP2 die Schaffung eindeutiger Adressierungsmöglichkeiten auf Artekelebene.

II.3.2 Durchgeführte Arbeitsschritte

AWB: Ausgehend von der im Rahmen der Kooperation zwischen SAW und TCDH erarbeiteten Codierung des AWB wurde dessen Auszeichnung erweitert. Im Einzelnen betraf dies die semantische Auszeichnung etymologischer und bibliographischer Verweise, den Formen-Teil und den Bedeutungs-Teil. Diese vertiefte Auszeichnung wurde im Laufe der Förderphase auf sämtliche neu erschienene Lieferungen des AWB angewandt, inklusive der damit verbundenen Anpassungsarbeiten (Einarbeitung neuer Sonderzeichen, Auflösung verkürzter Lemmaansätze, Kontrolle der grammatischen Angaben

etc.). Zum Berichtszeitpunkt ist das AWB bis einschließlich Band 7, Lieferung 12 bearbeitet und bis zum Lemma „quedan“ im Internet verfügbar (vgl. awb.saw-leipzig.de). Ergänzend wurden die verschiedenen Literaturverzeichnisse des Althochdeutschen Wörterbuchs semantisch annotiert und zu einer Gesamtdatenbank zusammengeführt. Diese ergänzenden Arbeiten konnten aus eigenen Mitteln des Trierer Kompetenzzentrums durchgeführt werden.

¹DWB: Die Kuratierung der Daten des ¹DWB betraf einerseits die Erweiterung der vorliegenden XML-Auszeichnung des Wörterbuchs selbst und andererseits die Vertiefung der Auszeichnung des Quellenverzeichnisses und anschließend die weitgehende Vernetzung zwischen beiden. Im Zuge der weiteren Auszeichnung des Quellenverzeichnisses wurden dabei ausgesparte oder verkürzt wiedergegebene Informationen zu Autoren, Werken und Referenzangaben ergänzt bzw. in maschinell auswertbarer Form hinterlegt. Außerdem wurden die fast 15.000 Einträge automatisch auf PPN-Nummern des Gemeinsamen-Verbund-Katalogs bezogen. Die Evaluierung dieses Ergebnisses erfolgt durch Hilfskräfte, die aus zentralen Mitteln des Trierer Kompetenzzentrums beschäftigt werden.

Die Vernetzung zwischen Wörterbuch und Quellenverzeichnis war eine herausfordernde Aufgabe. Aufgrund der diskursiven Struktur des ¹DWB sind im Allgemeinen zu einer Sigle mehr als ein Beleg verzeichnet, sodass man hier von einer geschätzten Zahl von 2 Millionen Belegen ausgehen kann. Die Identifikation der rechten und linken Grenzen von Prosabelegen und deren Zuordnung zu einer Sigle konnte daher nur semiautomatisch, d.h. mit manueller Nacharbeit, durchgeführt werden. Darüber hinaus mussten die Autorenteile der Siglen einem konkreten Autor im Quellenverzeichnis zugeordnet werden. Erschwerend kam hinzu, dass das Quellenverzeichnis unvollständig ist – geschätzte 10 % der Siglen haben keine Entsprechung im Quellenverzeichnis. Aufbauend auf einer von der BBAW eingebrachten Access-Datenbank wurde mit der Zuordnung der Belegautoren in ihren jeweiligen Schreibvarianten zu den Referenzformen im Quellenverzeichnis begonnen. Nach der Verknüpfung auf Autorenebene stand die Zuordnung der Werktitel im Fokus. Diese gestaltete sich aufgrund der starken Varianz und den routinemäßigen Verkürzungen innerhalb der Belegreferenzen als noch einmal deutlich aufwendiger und machte ein neues Verfahren notwendig. Es wurde deshalb jeder Nennung im Wörterbuch und jedem Titelansatz im Quellenverzeichnis ein eigenständiger Vergleichsschlüssel zugeordnet, der aus Autor (in normierter Form) und, so vorhanden, Titel und Datum besteht. Da dieser Schlüssel nur als technisches Bindeglied fungiert, konnte er im Hinblick auf den maschinellen Abgleich erstellt und optimiert werden. So wurde beispielsweise die Schreibung von Umlauten vereinheitlicht, bestimmte Zeichenfolgen, die für historisch bedingte Varianz anfällig sind, vereinfacht (z.B. Abkürzungspunkte durch Wildcards ersetzt etc.). Es folgte ein Abgleich in mehreren

Schritten: Zunächst wurde Zeichenidentität vorausgesetzt, dann wurden für die Restmenge Wildcards berücksichtigt, dann fehlende oder abweichende Datumsangaben zugelassen. Es verblieb jedoch ein relativ großer Rest von Belegreferenzen, bei denen Titel und Datum ausgespart sind. Diese wurden zunächst automatisch der jeweiligen Referenzausgabe des Autors zugeordnet, was aber nicht immer zu korrekten Ergebnissen führt. Hier bleiben sukzessive manuelle Eingriffe, insbesondere bei den hochfrequenten Werken, notwendig.

²DWB: Die Datenkuratierung der Neubearbeitung des Deutschen Wörterbuchs von Jacob Grimm und Wilhelm erfolgt im Rahmen eines laufenden DFG-geförderten Vorhabens federführend durch das TCDH in Zusammenarbeit mit der AdWG und der BBAW. Dabei wird das ²DWB volltextdigitalisiert, in XML/TEI-P5 annotiert, ¹DWB und ²DWB auf Stichwortebene miteinander vernetzt, und das ²DWB nach den etablierten Verfahren mit den weiteren Wörterbüchern im Trierer Wörterbuchnetz und im DWDS verlinkt und in die Digitale Plattform der AdWG eingebunden. Die zur Einbindung der ²DWB-Daten in das Konzept von ZHistLex erforderlichen Arbeiten wurden vom TCDH in Eigenleistung eingebracht.

MWV: Die TEI-P2 codierten Daten der im MWV enthaltenen Wörterbücher (Lexer, BMZ) sowie des „Findebuchs zum Mittelhochdeutschen Wortschatz“ wurden federführend durch die Arbeitsstelle des MWB in TEI-P5 konvertiert und gemeinsam für die elektronische Nachnutzung im Rahmen von ZHistLex vorbereitet.

MWB: Wie im Projektantrag vorgesehen wurden die MWB-internen Verweise überarbeitet und so ausgezeichnet, dass sie auch auf die Mikrostrukturebene des Wörterbuchs als Hyperlinks realisiert und für lexikographische Anfragen verwendet werden können. Nachdem in AP4 die Entscheidung gefällt worden war, PPNs als zentrales Merkmal der Quellenverzeichnisse des Zentrums anzusetzen, wurden diese für das gesamte Siglenverzeichnis des MWB nachgerüstet. Umfassende Aufgaben der Datenkuratierung betrafen die Umstellung des Volltextarchives des MWB auf TEI-P5, die als zusätzliche Aufgabe identifiziert worden war, nachdem in AP5 von einer exemplarischen Verknüpfung von Texten und Wörterbüchern abgesehen worden war zugunsten einer umfassenden Verknüpfung über Webservices (vgl. unten, AP5). Die zusätzlichen Aufgaben wurden überwiegend in Eigenleistung der AdWL Mainz in das ZHistLex eingebracht.

II.3.3 Zusammenfassung der erzielten Ergebnisse und Nutzen

Als Ergebnis von AP3 liegen die betrachteten Wörterbuchressourcen in Form von XML/TEI-P5 codierten Fassungen vor und werden über die in AP2 entwickelten Webservices zugreifbar gemacht. Die erzielten Ergebnisse werden auch im Rahmen ei-

ner Überarbeitung des Trierer Wörterbuchnetzes durch das TCDH verwendet, hier insbesondere die Vernetzung zwischen ¹DWB und Quellenverzeichnis, welches bisher nicht online verfügbar war. Die Wörterbuchdaten werden des Weiteren vom TCDH im Rahmen seiner Beteiligung am europäischen Verbundvorhaben ELEXIS (www.elex.is) als Ressourcen eingebracht.

II.4 AP4 Metadatenchema für ein übergreifendes Quellenverzeichnis

Federführung: AdW Göttingen in Eigenleistung

II.4.1. Aufgaben und Ziele

Das Arbeitspaket umfasst alle im Rahmen des ZHistLex anfallenden Aufgaben, die außer den eigentlichen Wörterbuch-Ressourcen auch die Informationen der einzelnen Quellenverzeichnisse der Ressourcengeber betreffen. Diese sollen gemeinsam absuchbar gemacht werden, so dass die Überschneidungen im Quellenbereich sichtbar gemacht werden können. Dies schließt auch die Suche in den Metadaten zu den Quellen (z.B. Sprachraum, Textsorte) ein, die eine tiefergehende Erschließung der lexikographischen Arbeitsweise sowie eine gesicherte Evaluierung der jeweiligen Befunde erlaubt. Die Herausforderung besteht daher zum einen darin, Werke unterschiedlicher Editionen und unterschiedlicher Titelaufnahmen virtuell zusammenzuführen, und zum anderen darin, die in den einzelnen Quellenverzeichnissen ggf. schon kodierten Metadaten nutzbar zu machen. Während der Projektlaufzeit zeigte sich die wichtige Bedeutung der Quellenverzeichnisse für die bidirektionale Verknüpfung von Wörterbüchern und digitalen Volltexten in AP5, sodass als zusätzliche, über den Projektantrag hinausgehende Aufgabe für AP4 die Erweiterung der Datenmodellierung und Recherchemöglichkeiten in Bezug auf Referenzschemata identifiziert wurde.

II.4.2 Durchgeführte Arbeitsschritte

Um die Voraussetzung für die Bearbeitung dieser Aufgaben zu schaffen, wurde ein zweistufiges Verfahren entwickelt: In einem ersten Schritt annotieren die Ressourcengeber die Einträge ihrer Quellenverzeichnisse mit Hilfe der sog. Pica-Produktionsnummer (PPN), die im Gemeinsamen Verbundkatalog (GVK) verzeichnet ist und die eine bibliotheksübergreifende Identifikation von Katalogeinträgen gewährleistet. Über einen automatischen Abgleich der im GVK angegebenen PPN können Katalogeinträge zu einer potentiell zusammengehörigen Gruppe zusammengefasst werden, wodurch z.B. separate Textausgaben und Ausgaben in Sammeleditionen aufeinander beziehbar zu machen sind. In einem zweiten Schritt sieht das Verfahren vor, die einschlägigen Metadaten des GVK, der DNB oder des OPAC auszuwerten, die über die PPNs automatisch abgefragt

werden können. Teilen zwei Einträge oder Eintragsgruppen die relevanten Metadatenfelder, so können die Einträge zusammengefasst werden. Auf diese Weise lassen sich etwa unterschiedliche Editionen eines Textes aufeinander beziehen und zu einem gemeinsamen Werkeintrag zusammenführen. Für die Frage, in welchen Wörterbüchern ein bestimmter Text als Quelle auftritt und wie diese Quelle dort behandelt wird, kann auf diese Weise eine wesentliche Grundlage bereitgestellt werden.

Als Alternative zur Identifikation über PPN wurde auch ein Identifikationsverfahren erwogen, das auf bereits bestehenden normierten Werkverzeichnissen wie z.B. dem Verfasserlexikon beruht. Da die Anwendungsmöglichkeiten dieses Verfahrens auf den Bereich der älteren deutschen Literatur (Alt- und Mittelhochdeutsch, Frühneuhochdeutsch) beschränkt sind, wurde dieser Weg allerdings nicht weiterverfolgt.

Von der im Projektantrag vorgesehen Erarbeitung eines zentralen Metadatenschemas für alle beteiligten Ressourcen wurde also abgesehen zugunsten eines flexiblen Verfahrens, in dem die Datengeber ihre annotierten Meta- und bibliographischen Angaben über die PPNs miteinander teilen können. Insgesamt folgt das Arbeitspaket der Gesamtarchitektur des ZHistLex, indem die Quellenverzeichnisse (wie auch die Wörterbücher und Textarchive) bei den einzelnen Datengebern verbleiben und über einen ZHistLex-Webservice (siehe AP2) abgefragt werden können. Auch ggf. annotierte Metadaten wie Sprachraum, Entstehungszeit, Textsorte usw. werden nicht einem zentralen Verzeichnis gehalten, sondern bei den einzelnen Datengebern über den Service abgefragt.

Folgende Arbeitsschritte wurden im Einzelnen durchgeführt:

MWB: Das gesamte Siglenverzeichnis wurde durch die AdWL Mainz mit PPN aus dem GVK annotiert. In einem weiteren Arbeitsgang wurden die bereits vorhandenen Metadaten (Entstehungszeit, Überlieferungszeit, Textsorte, Sprachraum) angereichert und verfeinert. Der in AP2 spezifizierte Webservice für Quellenverzeichnisse wurde für das MWB implementiert.

FWB und ²DWB: Annotation der Verzeichnisse mit PPN und übergreifenden Titelleinträgen aus dem OPAC einschließlich einer Evaluierung der PPN nach bibliothekarischen Qualitätsstandards im Rahmen der Projektförderung im Akademienprogramm bzw. in Kooperation mit der SUB Göttingen. Die beiden Webservices wurden durch die SUB Göttingen implementiert.

Titus: Annotation der PPNs im Verzeichnis der verfügbaren digitalen Volltexte durch die Arbeitsgruppe Uni Frankfurt und Implementation des Webservices.

¹DWB: Den ca. 15.000 Einträgen des Quellenverzeichnisses wurden über eine maschinelle Vorauszeichnung automatisch PPN-Nummern des Gemeinsamen-Verbund-Katalogs zugeordnet, die anschließend manuell überprüft, korrigiert u. ggf. ergänzt werden mussten, was innerhalb der Projektlaufzeit zu ca. 70% fertiggestellt werden konnte.

Demonstrator: Um die Arbeitsweise der Webservice-basierten virtuellen Zusammenführung der Quellenverzeichnisse auch für die Nutzung mit einer GUI zu illustrieren,

wurde von der AdWL Mainz ein Demonstrator implementiert, der über ein Suchformular eine einfache Anfrage entgegennimmt, alle zur Verfügung stehenden Services der Zentrumpartner aufruft und das kumulierte Ergebnis in beispielhafter Weise darstellt, z.B. durch Visualisierung der Quellenüberschneidungen (<http://tares.uni-trier.de/ZHistLex/Quellenverzeichnisse>).

II.4.3 Erzielte Ergebnisse

Es konnte gezeigt werden, dass die virtuelle Zusammenführung von ganz unterschiedlichen Quellenverzeichnissen über die PPN und Webservices eine gangbare Methode darstellt. Dies trifft zu, obwohl im GVK ein gewisser Wildwuchs an PPNs mit mehreren PPNs pro Werk zu verzeichnen ist, der seine Ursache in verschiedenen Titelaufnahmen der beteiligten Bibliotheken hat. Dies ist für das gewählte Verfahren im ZHistLex jedoch unproblematisch, solange zwei bibliographische Einheiten mindestens eine PPN miteinander teilen. Die Annotation der PPNs durch die Datengeber erwies sich (von einzelnen Sonderfällen abgesehen) ebenfalls als unproblematisch, da sie halbautomatisch mit anschließendem Korrekturgang durch Hilfskräfte erfolgen konnte.

Das zweite Ziel, nämlich die Metadaten der so virtuell zusammengeführten bibliographischen Einträge automatisch zu teilen und Editionsgruppen zu Werkeinträgen zu vereinigen, konnte mit den vorhandenen Ressourcen, die in Eigenleistung der AdWG erfolgten, jedoch nicht erbracht und von den anderen Arbeitsgruppen nicht stellvertretend übernommen werden.

Dasselbe gilt für zusätzliche Aufgaben, die AP4 in Zusammenarbeit mit AP5 zuwachsen. Hier wurde deutlich, dass die bidirektionale Verknüpfung von Wörterbüchern und digitalen Textarchiven, die nicht von vornherein aufeinander abgestimmt sind, nur über ihre Quellenverzeichnisse sinnvoll einzurichten ist. Hier hätten weitere Arbeiten sowohl für die Spezifikation der Services, ihrer Implementation und der Datenkuratierung erfolgen müssen. Letzteres betrifft vor allem die technische Abbildung und Nutzbarmachung der Referenzmuster, nach denen die Quellen bei verschiedenen Datengebern zitiert werden.

II.5 AP5 Exemplarische bidirektionale Verknüpfung von Wörterbuch-Belegen mit digitalen Volltexten

Leitung: AdWL Mainz

Beteiligte Projektpartner: BBAW Berlin, Uni Frankfurt

II.5.1 Aufgaben und Ziele

In den wissenschaftlichen Wörterbüchern des Deutschen, die die Wissensbasis des eHumanities-Zentrums für Historische Lexikographie bilden, sind Belegzitate und Stellenangaben von zentraler Bedeutung. Im digitalen Medium kann diese Vernetzung von Wörterbüchern und Texten durch Verlinkung in beide Richtungen besonders effektiv genutzt werden: Können digitale Volltexte vom Wörterbuch aus angesprochen werden, ist eine Rekontextualisierung des Zitats im ursprünglichen Textzusammenhang möglich. Sind Texte mit Wörterbucheinträgen verknüpft, sind Verständnishilfen unmittelbar verfügbar.

In Wörterbüchern, die mit traditionellen Mitteln erarbeitet wurden, kann die Vernetzung nur im Nachhinein erfolgen, während sie bei Wörterbüchern mit digitalem Ausarbeitungskonzept, denen ein digitales Textkorpus zugrunde liegt, von vornherein gegeben ist. Dies ist bei den historischen Wörterbüchern des Deutschen bislang nur beim Mittelhochdeutschen Wörterbuch (MWB) der Fall.

Ausgangsvorstellung für die bidirektionale Verlinkung von Wörterbüchern und Quelltexten war bei der Projektbeantragung daher die schon vorhandene Verknüpfung des MWB mit seinen Quelltexten, die über feste Verlinkungen realisiert sind und im Online-Angebot¹ in beide Richtungen verfolgt werden können.

Für das Zentrum für Historische Lexikographie kann es jedoch nicht um Einzelverknüpfung von bestimmten Wörterbüchern mit ihren Quelltexten gehen, sondern um die Bereitstellung einer Schnittstelle, die Texte und Wörterbücher in geeigneter Weise abfragbar und interoperabel macht. Entsprechend der ZHistLex-Architektur sollen solche Abfragen mittels Webservices erfolgen, die dezentral von den Anbietern der jeweiligen Wörterbücher und Texte realisiert werden. Für die Einrichtung der Webservices stellt das ZHistLex die nötige Spezifikation zur Verfügung.

Wörterbücher und Texte verbleiben bei diesem Verfahren bei ihren Anbietern und können von jeder beliebigen anderen Stelle abgefragt werden, Ressourcen können auf diese Weise miteinander geteilt werden. Vorzüge dieses Webservice-basierten Ansatzes sind leichte Erweiterbarkeit, Flexibilität und dezentrale Datenhaltung. Bereits vorhandene feste Verlinkungen können weitgehend ohne Mehraufwand in das Webservice-basierte Verfahren überführt werden.

¹ <http://www.mhdwb-online.de/>

Für die Verlinkung von Wörterbüchern und Texten über Webservices wird in der ZHistLex-Architektur – wie im gedruckten Wörterbuch – die Zitierreferenz der Texte benutzt. Eine andere Möglichkeit wäre die Verknüpfung über die Lemmatisierung. Dies setzt jedoch eine zuverlässige, vollständige Lemmatisierung aller Texte in Bezug auf die Wörterbücher voraus, die im Falle von historischen Texten aber sehr aufwändig ist, weil sie – anders als bei gegenwartssprachlichen Texten – nicht vollautomatisch durchgeführt werden kann. Die Verknüpfung über die Referenz funktioniert auch bei abweichenden Lemma-Ansetzungen und anderen Erschwernissen der Zuordnung.

II.5.2 Durchgeführte Arbeitsschritte

II.5.2.1 Webservice-Spezifikation für Texte

Der ursprüngliche Arbeitsplan sah zunächst nur die technische Webservice-Spezifikation für Wörterbücher vor. Damit auch elektronische Volltexte über die Schnittstelle abgefragt werden können, musste sie entsprechend erweitert werden. Ein zentraler Text-Abfragetyp ist z.B. die auf ein Belegzitat im Wörterbuch bezogene Anforderung: *„Gib die Textumgebung aus dem Text ‚Kaiserchronik‘ für die Referenz 575 aus“*. – Diese Erweiterung der Spezifikation für Texte erfolgte in enger Zusammenarbeit mit der Arbeitsgruppe „Standards“ (AP2; vgl. oben, Abschnitt AP2).

II.5.2.2 Datenkuratierung der Textarchive

Im Projektantrag war in AP 5.1 für das AWB die exemplarische Verknüpfung mit Otfrids ‚Evangelienbuch‘ im DDD-Korpus ReA und für das MWB die Verknüpfung mit der ‚Kaiserchronik‘ im Textarchiv des MWB und im DDD-Korpus ReM vorgesehen. Mit der Modifikation des Arbeitsplans zugunsten der Webservice-basierten Arbeitsweise (s.o.) wurde von der exemplarischen Aufbereitung einzelner Texte abgesehen. Stattdessen wurden die Textarchive insgesamt so eingerichtet, dass die Webservices wie gewünscht arbeiten können.

Diese Aufbereitung betraf in erster Linie das „Referenzkorpus Altdeutsch“ (ReA) beim Projektpartner Uni Frankfurt. Die gut 100 tiefannotierten Volltexte wurden in Lesetexte konvertiert und so das „Lesekorpus Altdeutsch“ (LeA)² geschaffen. Die HTML-Textdateien des LeA liegen parallel auch im XML-Format (TEI-P5) vor und bieten somit die Voraussetzungen für eine Verknüpfung mit anderen Ressourcen, zunächst vor allem mit dem Althochdeutschen Wörterbuch (AWB). Die Zitierreferenz war für den

² <http://titus.uni-frankfurt.de/lea/>

Großteil der Texte bereits identisch mit der im AWB gebrauchten, für die übrigen wurde sie an die des AWB angepasst.

Das ReA/LeA-Korpus ist vollständig lemmatisiert, allerdings nicht nach dem noch in Ausarbeitung befindlichen AWB, sondern nach dem bereits 1992 publizierten althochdeutschen Wörterbuch von Jochen Splett (AWB^{SpI}). Als Wortfamilienwörterbuch wurde AWB^{SpI} auch in AP6 benutzt, und es wurde für die Verknüpfung der ahd. Wortfamilien mit dem AWB eine Konkordanz der Lemmalisten von AWB^{SpI} und AWB hergestellt. Über diese Konkordanz kann – unabhängig von den entwickelten Webservices – zukünftig auch die Verknüpfung des ReA/LeA mit dem AWB pro Textwort realisiert werden.

Im mittelhochdeutschen Textarchiv des MWB war keine inhaltliche Datenkuratierung nötig (s.o.), allerdings wurde die Datenstruktur fast des gesamten Archivs von rund 240 Quellentexten in der Arbeitsgruppe der AdWL Mainz auf TEI/XML umgestellt (vgl. oben, Abschnitt AP3). Die Verbindung der Editionstexte des MWB-Korpus mit den Handschriftentexten des ReM, und damit zweier Textarchive derselben Periode, wurde vorbereitet.

AP 5.2 des Projektantrags sah die Verknüpfung von Texten des DTA-Korpus mit dem DWB vor. Diese Verknüpfung konnte nicht umgesetzt werden, da es nach ausführlichen Stichproben keine Überschneidungen zwischen dem Textbestand des Deutschen Textarchivs und dem Quellenbestand des ¹DWB gab, ein überraschender Befund. Die nach der Ermittlung verbliebene Arbeitszeit wurde deshalb an anderer Stelle, in AP2, eingesetzt. Der wissenschaftlich-technische Ertrag des Projekts wird dadurch nicht beeinträchtigt, denn die Verfahrensweisen und Umsetzungsprinzipien sind durch die anderen Beispiele für bidirektionale Verknüpfung hinreichend gestützt.

Eine weitere Schicht der Datenkuratierung in den beiden Textarchiven betraf die Anreicherung der Quellen mit Metadaten und insbesondere die Annotation der Pica Produktionsnummer (PPN), die das zentrale Identifikationsmerkmal der Texte in der Webservice-basierten Architektur des ZHistLex darstellt (vgl. oben, Abschnitt AP4).

II.5.2.3 Implementation entsprechender Webservices durch Anbieter elektronischer Volltexte

Beide Textarchive des Zentrums implementierten Webservices für Volltextausschnitte auf Basis der ZHistLex-Spezifikation:

- Referenzkorpus Altdeutsch (durch Uni Frankfurt):
<http://titus.fkidg1.uni-frankfurt.de/database/titusinx/zhl/DocumentationZHLTitus.pdf>
- MWB (durch AdWL Mainz):
<http://tares.uni-trier.de/ZHistLex/API/>

II.5.3 Anwendung des Verfahrens

Die zentrale technische Idee des ZHistLex ist es, die einzelnen beteiligten Ressourcen des Zentrums (Wörterbücher und Quellentexte) über ein standardisiertes System von Webservices interoperabel zu machen. Die Entwicklung einer Oberfläche für menschliche Nutzung (GUI) war dabei von zweitrangiger Ordnung und nicht im Arbeitsplan verankert. Ein eigenständiger Demonstrator wie er für AP4 als Eigenleistung implementiert worden ist (s. AP4), wurde für die bidirektionale Verknüpfung nicht entwickelt. Dennoch ist die konkrete Anwendung der Webservices durch verschiedene Online-GUIs bereits in der Umsetzung: Das LeA wird von der Lemmaangabe ins AWB-Online³ zum betreffenden Stichwort verlinken, außerdem von der Stellenangabe ebenfalls ins AWB. Dort wurden die Zitierreferenzen im Zuge der Datenkuratierung (s. AP2) so ausgezeichnet, dass sie – unter Ausnutzung der entwickelten Webservices – als Ausgangspunkte für Verweise in das LeA sein können. In einer internen Fassung des Online-AWB am TCDH sind Verlinkungen von der Belegreferenzen zum Volltext im LeA in Vorbereitung. Für das Mittelhochdeutsche ist die Nutzung der entwickelten Services für die Volltexte in Bezug auf die Mittelhochdeutsche Begriffsdatenbank⁴ in Planung.

Ein besonderes Erschwernis für die Bewältigung des AP war der Umstand, dass sich erst während der Arbeit die besondere Bedeutung der Quellenverzeichnisse für die webservice-basierte Verknüpfung von Texten und Wörterbüchern verschiedener Datengeber zeigte. Damit verbunden sind auch Wörterbuch-übergreifenden Fragestellungen, wie der Wunsch, Überschneidungen in der Behandlung von Textstellen durch unterschiedliche Wörterbücher abfragbar zu machen. Bei jedem dieser Abfragetypen ist es erforderlich, dass erstens die bibliographischen Einheiten der Texte und zweitens die Referenzschemata aufeinander bezogen werden können. Die erste Anforderung wurde in AP4 für Gruppen von Editionen durch Verwendung der PPN umgesetzt, die zweite konnte in AP4, das in Eigenleistung von der AdWG erbracht wurde, nicht geleistet werden. Die entwickelten Anwendungen für das Althochdeutsche und das Mittelhochdeutsche bedienen sich daher Hilfskonstruktionen, die in Zukunft auf Basis der ZHistLex-Architektur verbessert und weiter ausgearbeitet werden sollten.

II.5.4 Zusammenfassung der erzielten Ergebnisse und Nutzen

Statt einer exemplarischen festen Vernetzung spezieller Texte mit Wörterbüchern wurden zwei umfassende Textarchive, nämlich das Referenzkorpus Altdeutsch und das Mittelhochdeutsche Textarchiv insgesamt für die standardisierte Abfrage über Webservices vorbereitet und damit die Verlinkung mit lexikographischen Ressourcen angelegt. Die

³ <http://awb.saw-leipzig.de>

⁴ <http://mhdbdb.sbg.ac.at/>

nötige Webservice-Spezifikation wurde im Rahmen des Zentrums vorgenommen, die Services wurden durch die Datengeber implementiert. Die erzielten Arbeitsergebnisse gehen damit über den im Projektantrag vorgesehen exemplarischen Charakter hinaus und sind bereits jetzt von arbeitspraktischem Nutzen, wie die Vernetzung des AWB und des LeA/ReA zeigt. Ein weiterer Ausbau des entwickelten Verfahrens in Bezug auf weitere Datengeber und ggf. weitere Abfragemöglichkeiten ist geplant.

II.6 AP6 Epochenübergreifende Wortfamilienstrukturen

Leitung: Uni Frankfurt

Beteiligter Projektpartner: AdWL Mainz (Arbeitsstelle des MWB)

Externe Kooperation mit Prof. Dr. Thomas Klein, Universität Bonn; Prof. Dr. Jochen Splett, Universität Münster, hat das Projekt wohlwollend verfolgt und die Nutzung seines ahd. Wortfamilienwörterbuchs für Projektzwecke gestattet.

II.6.1 Ziel: Computergestützte Etablierung epochenübergreifender Wortfamilienstrukturen

Ob historische Wörterbücher den Wortschatz streng einzelwortbezogen oder in Wortfamilien bearbeiten und darbieten sollen, ist seit den Anfängen der wissenschaftlichen Lexikographie umstritten. Während in der Anfangszeit mehrfach auch Wortfamilienwörterbücher realisiert wurden (z.B. Schmellers Bayerisches Wörterbuch 1827-1837, Graffs Althochdeutscher Sprachschatz 1834-1842, das Mittelhochdeutsche Wörterbuch von Benecke/Müller/Zarncke 1854-1866), dominiert heute bei den großen Belegwörterbüchern das einzelwortbezogen alphabetisch fortschreitende Wörterbuch. Ausnahmen bilden das Schweizerische Idiotikon und das Deutsche Fremdwörterbuch, in gewisser Weise auch das Althochdeutsche Wörterbuch.

Der Grund für die Bevorzugung der einzelwortbezogen alphabetisch fortschreitenden Wörterbucharbeit ist nicht inhaltlicher Natur, sondern vor allem wörterbuchplanerischer Art. Bei Ausarbeitung von Wortfamilien lässt sich der Arbeitsfortschritt scheinbar nicht so stringent planen, kontrollieren und durch Publikation dokumentieren wie bei Abarbeitung einer alphabetischen Stichwortliste. Dies spielt vor allem eine Rolle für die Großunternehmen der Akademienforschung, die mit engen planerischen Vorgaben und Auflagen arbeiten. Dass auch in diesen Arbeit nach Wortfamilien möglich ist, zeigt das Beispiel des Althochdeutschen Wörterbuchs (AWB) der Sächsischen Akademie der Wissenschaften, dessen Begründer in einer längeren Vorbereitungsphase zunächst mit der Ausarbeitung von Wortfamilien begonnen haben (vgl. z. B. den Werkstattbericht von Karg-Gasterstädt 1942); das ist bis heute im Wesentlichen beibehalten worden, publiziert wird allerdings (gemäßigt) einzelwort-alphabetisch.

Der Erkenntniswert der Wortfamilieneinheit für die Lexikographie und darüber hinaus für sprachgeschichtliche Fragen überhaupt, vor allem für die Wortbildungslehre, ist völlig unbestritten (vgl. etwa Hundsnurscher 2002, Splett 2002, Klein 2018). Eindrucksvolle Zeugnisse dafür sind die beiden Wortfamilienwörterbücher zum Althochdeutschen und zum Gegenwartsdeutschen, die – abseits der Akademienforschung – von Jochen Splett an der Universität Münster ausgearbeitet worden sind (erschieden 1993 und 2009). Der Fokus liegt bei ihnen auf der Analyse der Wortbildungen und ihrer Ableitungsbeziehungen, die Bedeutungsangaben beruhen auf den zugrundeliegenden Belegwörterbüchern und weiteren lexikographischen Hilfsmitteln.

Wo es bereits digitale Lemmalisten gibt, können Wortfamilienstrukturen computergestützt erarbeitet und epochen- und (unter bestimmten Bedingungen auch) varietätenübergreifend verknüpft werden, wie präzise und auf umfangreiche Materialien gestützte Überlegungen und Erprobungen von Thomas Klein (2013, 2018) zeigen. Diese Wortfamilienstrukturen können auch nachträglich mit den Artikeln für einzelwortbezogen alphabetisch publizierte Wörterbücher verknüpft werden, so dass für diese dann eine Benutzung nach Wortfamilien möglich wird. Darüber hinaus ist die wortbildungsmorphologische Analyse der betreffenden Stichwortbestände, die der Wortfamiliengliederung zugrundeliegt, ein unschätzbares Hilfsmittel der historischen Wortbildungslehre.

ZHistLex hat sich im Rahmen dieses Aufgabengebiets zum Ziel gesetzt, die Möglichkeiten der computergestützten Etablierung einer epochenübergreifenden Wortfamilienstruktur exemplarisch für das Althochdeutsche und das Mittelhochdeutsche zu erkunden; für diese beiden Sprachepochen sind die Bedingungen aufgrund der exzellenten Vorarbeiten von Splett und Klein und aufgrund des Vorliegens von umfassenden lexikographischen Online-Angeboten (in Gestalt des AWB, des Verbunds der älteren Mittelhochdeutschen Wörterbücher im Trierer Wörterbuchnetz und des in Ausarbeitung befindlichen Mittelhochdeutschen Wörterbuchs) besonders günstig.

II.6.2 Erprobung am Althochdeutschen und Mittelhochdeutschen

II.6.2.1 Segmentierung der mittelhochdeutschen Lemmata

Voraussetzung für die Zuweisung von Lexemen zu Wortfamilien ist ihre wortbildungsmorphologische Segmentierung (Abtrennung der Affixe, Feststellung der Grundmorpheme) und die weitere Feststellung der Kernwort-Stämme (ggf. durch Ablaut- und Umlautbereinigung) und der Kernwörter der Wortfamilien selbst; vgl. dazu ausführlich Klein 2018, S. 17-25. So ist etwa mhd. *bētegültig* ‚steuerpflichtig‘ als *bēt-e-gült-ig* zu segmentieren, nach Um- und Ablautbereinigung ergeben sich als Kernwort-Stämme *bit* und *gält*, als Kernwörter (Wortfamilien-Köpfe) selbst *bitten* und *gälten* (vgl. Klein 2018,

S. 17). Die von Thomas Klein für diese Analyseschritte zur Verfügung gestellten Scripte wurden in mehreren Durchläufen erprobt und angepasst.

Als Hauptproblem der automatischen Segmentierung hat Klein (2018, S. 18f.) die mhd. nicht mehr produktiven „Alt-Suffixe“ erkannt. Für die beiden häufigsten von ihnen (-*el* und -*en*) wurde daher eine vollständige Einzelprüfung und ggf. Korrektur der automatischen Segmentierung der rd. 6.000 mhd. Stichwörter durchgeführt, in der diese Zeichenfolge als Suffix abgetrennt wurde. Beispiele, die den Korrekturbedarf illustrieren, sind die Rückgängigmachung der Segmentierung, im Falle von Erbwörtern etwa bei *sib-en* 'sieben' (42 Lemmata) und *sat-el* 'Sattel' (35 Lemmata), bei Lehnwörtern etwa von *zirk-el* 'Zirkel' (20 Lemmata) und *tav-el* 'Tafel' (53 Lemmata).

Bei diesem Korrekturgang zur Prüfung der automatischen Abtrennung von -*en/-el* wurde zugleich auch die automatische Wortfamilien-Zuordnung (vgl. die folgenden Punkte) geprüft und in zahlreichen Fällen korrigiert.

II.6.2.2 Ermittlung der mhd. Fortsetzungen ahd. Wortfamilien

II.6.2.2.1 Zuordnung der Wortstämme mhd. Fortsetzungen ahd. Lexeme zu ahd. Wortfamilien

Die Zuordnung des mhd. Wortschatzes zu den bereits im Ahd. bezeugten Wortfamilien konnte sich die Vereinigung der ahd. und mhd. Lemmalisten zunutze machen, die bereits von Thomas Klein halbautomatisch (d.h. sorgfältig geprüft) durchgeführt worden war (vgl. Klein 2018, S. 13-16). Dabei ist zunächst bei den rd. 12.500 der rd. 73.000 mhd. Lexeme anzusetzen, für die ahd. Vorgänger bezeugt sind (17%). Diese 12.500 mhd. Lemmata weisen (bei Mehrfachzählung der Komposita entsprechend der Zahl ihrer Stämme / Grundmorpheme) insgesamt 15.000 durch Segmentierung und Alternanzbereinigung gewonnene Kernwortstämme auf, die sich nun automatisch den Wortfamilien-Kernwörtern ihrer ahd. Entsprechungen zuordnen ließen, so etwa im Fall der ahd./mhd. Simplex-Entsprechungen *bitten* / *bitten* und *geltan* / *gëlten* die mhd. Wortstämme *bit* und *gëlt* den ahd. Wortfamilien BITTEN bzw. GELTAN.

II.6.2.2.2 Zuordnung mhd. Lexeme ohne ahd. Vorgänger zu ahd. Wortfamilien

Von rd. 73.000 mhd. Lexemen sind für rd. 60.500 (83%) keine ahd. Vorgänger bezeugt; bei den rd. 100.000 Wortstämmen (Stamm-tokens) dieser Lexeme, die für die Wortfamilienzuweisung relevant sind, sind es rund 85.000. Doch können für jenen Teil dieser Wortstämme (types), die in Schritt 2.2.1 bereits ahd. Wortfamilien zugeordnet worden waren, automatisch Zuordnungsvorschläge erzeugt werden: So hat etwa mhd. *bët-e-gült-ig* keine ahd. Entsprechung, kann aber aufgrund seiner Wortstämme *bit* und *gëlt* den ahd. Wortfamilien BITTEN und GELTAN zugewiesen werden.

Die automatische Zuordnung ist allerdings nicht immer eindeutig. So wird etwa der mhd. Wortstamm *dërb* den drei ahd. Wortfamilien -DERBAN (z.B. mhd. *ver-dërb-en* / ahd. *firderban* st.V.), DERB (mhd. *dërb* / ahd. *derb* Adj. 'ungesäuert usw.') und DURFAN (z.B. mhd. *bí-derbe* / ahd. *biderbi* 'nützlich, brauchbar' Adj. [nhd. *bieder*]) zugewiesen. In einem ersten Durchgang durch die rd. 85.000 mhd. Wortstämme ohne ahd. Vorgänger ergab die automatische Zuordnung

- für rd. 50.000 Stämme (rd. 59%) eine 1:1-Zuordnung,
- für rund 12.000 Stämme (14%) eine 1:2-Zuordnung,
- für 9.000 Stämme (11%) eine Zuordnung von 1:3 bis 1:6, in einigen Fällen auch von 1:9,
- und für rd. 14.000 Stämme (16%) keine Zuordnung zu einer Wortfamilie.

Durch verschiedene Maßnahmen konnte dieses Ergebnis noch einmal beträchtlich verbessert werden. Dazu zählt die Berücksichtigung der Korrekturen und Zuweisungen, die bei der Abarbeitung der *-en/-el*-Liste (s. oben unter 2.1) gemacht wurden, vor allem aber die Einbeziehung der Wortfamiliengliederung im Mhd. Wörterbuch von Benecke /Müller/ Zarncke (BMZ, vgl. oben unter 1.). Dazu wurden aus den Daten des Mhd. Wörterbuchverbands die Wortfamilien-Lemmata des BMZ in die mhd. Lemmaliste übernommen, so z.B. zu *becke* swM. 'Bäcker' das BMZ-Wortfamilienlemma BACHE (1.Sg. des stV. *bachen* 'backen').

Mhd. *becke* sw.M. hat als Simplex keinen ahd. Vorgänger (nur als Grundwort im Kompositum *brôt-becko*). Nach der automatischen Zuordnung ergab sein Wortstamm BACK keinen 1:1-Zuordnungsvorschlag zu einem ahd. Kernwort, sondern einen 1:3-Vorschlag zu den ahd. Kernwörtern BACKAN 'backen', BACKO 'Backe' und BECKÎN 'Becken'. Das BMZ-Wortfamilienlemma BACHE findet sich aber nur zu mhd. Entsprechungen von ahd. Wörtern der Wortfamilie BACKAN. So konnte mhd. *becke* swM. automatisch der Wortfamilie ahd. BACKAN zugewiesen werden. Knapp 5.100 Zuordnungen konnten auf diese Weise vorgenommen werden.

Im Ergebnis konnten die Fälle ohne Zuordnung auf rund 11.500 (13,5 %) und die 1:2- bis 1:9-Zuordnungen auf rund 15.500 (18 %) statt bisher 21.000 (25 %) gesenkt werden. Die Zahl der 1:1-Zuordnungen ließ sich entsprechend von rund 50.000 (59 %) auf rund 58.000 (ca. 68 %) steigern.

II.6.2.2.3 Feststellung der mhd. Fortsetzungen ahd. Wortfamilien

Ein anderer Weg, die Fortsetzung ahd. Wortfamilien im Mhd. festzustellen, besteht in der systematischen Prüfung, ob die Kernwörter der ahd. Wortfamilien im Mittelhochdeutschen bezeugt sind. Dazu wurden zu den ahd. Wortfamilien-Köpfen ihre regelhaften mhd. Entsprechungen in der mhd. Lemmaliste gesucht (im Falle der Wortfamilie GELTAN z.B. *gëlten*) und bei Erfolg als mhd. Wortfamilien-Köpfe (z.B. GËLTEN) angesetzt. Auf diese Weise ließen sich zu 2.170 von 3.623 ahd. Wortfamilien (rund 60%)

automatisiert mhd. Entsprechungen ermitteln. Diese Liste wurde sodann einzeln geprüft und korrigiert, ggf. fehlende mhd. Entsprechungen ergänzt, umgekehrt vermeintliche mhd. Entsprechungen für tatsächlich nicht fortgesetzte ahd. Wortfamilien wieder getilgt. Als Ergebnis kann festgehalten werden, dass die ahd. Wortfamilien im Mhd. zu knapp 82 % (2.957 von 3.623) erhalten sind. Der Wortfamilien-Bestand ist also viel konstanter als der Lexem-Bestand (nur 47% der ahd. Lexeme haben mhd. Fortsetzer).

II.6.2.3 Nutzung von Strukturformeln

Sowohl Spletts Wörterbuch als auch Kleins Programme halten das Ergebnis der wortbildungsmorphologischen Analyse in Strukturformeln fest. Kleins Formeln sind sehr einfach gehalten und nur für einen Programmschritt der Segmentierung der Lemmata gedacht, die Abtrennung der Affixe und Feststellung der Kernwortstämme (vgl. Klein 2018, S. 17-22). Die Formeln Spletts sind hingegen sehr viel komplexer, denn sie sollen "die hierarchische Struktur der Wörter, das darauf basierende Beziehungsgeflecht innerhalb der Wortfamilie und die Verbindung zu anderen Wortfamilien" ausdrücken (Bd. I, S. XXX), nicht zuletzt durch hierarchische Klammerung. Die Strukturformeln wurden daher aus dem Frankfurter Digitalisat des Splettschen Wörterbuchs in die Projektdaten übernommen, wie bereits bei Klein 2018 (S. 16) angeregt worden ist. Der Versuch der Nutzung dieser Formeln zur automatischen Rekonstruktion der hierarchischen Binnenstruktur der Wortfamilien, z.B. in der Visualisierung in einem sog. "Stufungsprofil" (vgl. die ZHistLex-Seite unter Materialien & Papiere⁵), stieß aber bei vielen Lexemen auf bislang ungelöste Schwierigkeiten, die darauf beruhen, dass nicht jedem der hierarchisierten Klammerelemente der Strukturformel ein Element der segmentierten Wortform entspricht. Vgl. das Beispiel der dreigliedrig segmentierten Form *ana-far-tôn* 'angreifen' mit der viergliedrigen Strukturformel $((p(wV))sS)Vôn$ (zu lesen als: *anafartôn* < *anafart* < *anafaran* < *faran*), in der das Suffix *t* der Substantivbildung *fart* nicht als eigenes segmentiertes Morphem erscheint (vgl. Splett I,1, S. XXXI und im Zusammenhang der Wortfamilie *faran* auf S. 207). Insgesamt enthält die Wortliste 12.741 Einträge mit mindestens einem coverten Element.

II.6.2.4 Desiderate

II.6.2.4.1 Prüfung der automatisch erzeugten Wortfamilienzuweisungen

Für die Etablierung der ahd./mhd. Wortfamilienstruktur dürften mit den im Projekt geleisteten Arbeiten die Möglichkeiten der Computerunterstützung methodisch umfassend

⁵ <http://zhistlex.de/buwenPrototyp.html>

ausgeschöpft sein. Für eine vollständige Durchführung der ahd./mhd. Wortfamiliengliederung sind Vorschlagslisten zu prüfen, die mit den in 2.2.2 beschriebenen Verfahren erzeugt worden sind. Dies betrifft die Fortsetzung der ahd. Wortfamilien im Mhd. Die neuen mhd. Wortfamilien (zum großen Teil im Bereich des lateinischen und vor allem französischen Lehnwortschatzes) und die verbleibenden mhd. Einzellexeme erfordern eigene weitere Arbeitsschritte.

II.6.2.4.2 Einrichtung einer ahd./mhd. Wortfamilien-Datenbank

Die Projektergebnisse werden zur Zeit in einer Excel-Tabelle aufbewahrt. Die Überführung in eine Datenbank bleibt Desiderat, sie konnte im Rahmen des Projekts nicht mehr in Angriff genommen werden, vor allem weil die Informatikstelle erst im letzten Halbjahr des Projektzeitraums besetzt werden konnte.

II.6.3 Exemplarische wortgeschichtliche Auswertung am Beispiel der Wortfamilie *bauen*

Im Zusammenhang der Mitarbeit in AP7 ergab sich die Aufgabe einer Demonstration des Werts der epochenübergreifenden Wortfamilien-Struktur für wortgeschichtliche Fragen. Dies geschah am Beispiel der Geschichte der Wortfamilie *bauen* im Alt- und Mittelhochdeutschen. Darüber ist in zwei ausführlichen Referaten auf dem Arbeitsgespräch zur historischen Lexikographie in Bullay 2018 ([Abstract](#), [Handout](#) und [Literaturexzerpte](#) online auf der [Tagungsseite](#))⁶ und auf dem 26. Deutschen Germanistentag in Saarbrücken 2019 berichtet worden.

II.6.4 Erprobung von Visualisierungsmöglichkeiten

Ebenfalls im Rahmen der Mitarbeit an AP7 wurden Möglichkeiten der Visualisierung erprobt. Dies betraf zu einen die automatisierte Darstellung der Ableitungswege in einer Wortfamilie (sog. Stufungsprofil), zum anderen die interaktive Darstellung der Entwicklung der Wortfamilie *bauen* im Ahd. und Mhd. entsprechend den Folien zum Referat in Bullay 2018; vgl. oben, Abschnitt 3 und auf der ZHistLex-Seite unter Materialien & Papiere.

II.6.5 Nutzen und Verwertbarkeit

Das Projekt verfolgte keine direkt wirtschaftlich verwertbaren Ziele. Der Nutzen besteht 1. in der weitgehend vorbereiteten Ausweitung der ahd. Wortfamilienstruktur auf das

⁶ <https://www.uni-trier.de/index.php?id=65102>

Mittelhochdeutsche, 2. in den methodischen Erkenntnissen, die dabei für die halbautomatische Wortfamilienanalyse weiterer deutscher Teil-Wortschätze gewonnen wurden. Konkret setzt sich die Deutsch Diachron Digital-Initiative, die die Zusammenführung der einzelnen Referenzkorpora zur deutschen Sprachgeschichte beabsichtigt und ein großes Verbundprojekt vorbereitet, die Einbeziehung der wortbildungsmorphologischen Analyse-Ebene zum Ziel. Das dieser Aufgabe gewidmete Teilprojekt wird auf den in AP6 erzielten materiellen und methodischen Ergebnissen aufbauen.

II.6.6. Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen

Andere Stellen, die sich im Projektzeitraum mit dem Aufgabenkreis von AP6 befasst hätten, sind nicht bekannt.

II.6.7 Zitierte Forschungsliteratur

Hundsnurscher, Franz, Das Wortfamilienproblem in der Forschungsdiskussion, in: Lexikologie / Lexicology. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen / An international handbook on the nature and structure of words and vocabularies, hg. v. D. Alan Cruse u.a., Bd. 1 (HSK 21,1), Berlin / New York 2002, Art. 86, S. 675-680.

Karg-Gasterstädt, Elisabeth, Aus den Ergebnissen unserer Arbeit am Althochdeutschen Wörterbuch (Aus der Werkstatt des Althochdeutschen Wörterbuchs 15), PBB 65 (1942), S. 241-256.

Klein, Thomas, Verknüpfung digitaler Lemmalisten historischer Sprachstufen des Deutschen – wie und wozu? [Vortrag](#) auf dem [Arbeitsgespräch zur historischen Lexikographie 2013](#).

Klein, Thomas, Mittelhochdeutsche Wortfamilien: Ermittlung und Perspektiven, in: [Zeitschrift für Wortbildung / Journal of Word Formation](#), Jg. 2 (2018), S. 11-31.

Splett, Jochen, Althochdeutsches Wörterbuch. Analyse der Wortfamilienstrukturen des Althochdeutschen, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes, 3 Bd.e, Berlin/New York 1993.

Splett, Jochen, Bedingungen des Aufbaus, Umbaus und Abbaus von Wortfamilien, in: Lexikologie [wie Hundsnurscher], Art. 88, S. 688-699.

Splett, Jochen, Deutsches Wortfamilienwörterbuch. Analyse der Wortfamilienstrukturen der deutschen Gegenwartssprache, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes, 18 Bd.e, Berlin/New York 2009.

II.7 AP7 Innovative Zugriffs- und Darstellungsformen

Leitung: Uni Gießen

Beteiligte Projektpartner: AdW Göttingen, AdWL Mainz, Uni Frankfurt

II.7.1 Voraussetzungen, Aufgaben und Ziele

Die Situation in der historischen Lexikographie ist seit langem geprägt von einer internationalen Tendenz zur Digitalisierung. Viele historische Wörterbücher wurden retrodigitalisiert, viele werden derzeit entweder nur noch digital (z.B. OED) oder parallel (digital und gedruckt, z.B. Deutsches Rechtswörterbuch) erarbeitet. Dabei sind zwei gegenläufige Tendenzen zu beobachten: Zum einen sind Darstellungsformen und Zugriffsweisen nach wie vor geprägt von traditionellen Verfahren der Artikelgestaltung und der Organisation des Zugriffs. Zum anderen finden sich in der internationalen Landschaft aber auch vielfältige Formen der Anreicherung von Wörterbuchsubstanz mit neuartigen und innovativen Darstellungselementen und Zugriffsweisen, die sich in besonderer Weise den medialen Möglichkeiten digitaler Umgebungen verdanken.

Die metalexikographische Forschung hat sich dem Thema Digitalisierung in den letzten Jahren und Jahrzehnten zwar intensiv angenommen, allerdings die spezifischen Themen und Fragestellungen der historischen Lexikographie nur sehr stiefmütterlich behandelt. Exemplarisch kann hier der umfangreiche Ergänzungsband zum HSK-Werk „Wörterbücher/Dictionaries“ (2013) genannt werden, in dem die digitale historische Lexikographie des Deutschen mit nur wenigen Zeilen erledigt wird.

Die Ziele des Arbeitspakets 7 bezogen sich auf drei Schwerpunkte der digitalen historischen Lexikographie:

- Systematisierung und Weiterentwicklung integrativer wort- und wortschatzgeschichtlicher Darstellungsformen;
- Systematisierung und Weiterentwicklung von Formen und Verfahren der Visualisierung;
- Systematisierung und Weiterentwicklung von mehrdimensionalen Formen des Zugriffs und der Abfrage in digitalen historischen Wörterbuchsystemen.

II.7.2 Systematisierung von digitalen Darstellungsweisen und Praktiken

Im Hinblick auf die Systematisierung von Darstellungsformen, Zugriffsweisen und Formen der Visualisierung in gegenwärtigen digitalen historischen Wörterbuchsystemen wurden zu den großen historischen Wörterbuchunternehmen und den historischen Portalen sog. Steckbriefe erstellt, in denen nach einem festen und vergleichbaren Schema die jeweilige Praxis der Informationsdarbietung und der Erschließung charakterisiert wurde. Die drei wesentlichen Charakterisierungsdimensionen sind: die Hauptkomponenten eines Angebots (z.B. Artikel, Quellenverzeichnis, Stichwortliste, Volltexte, Versionierungsangaben), die vorgesehenen Zugriffs- und Suchweisen (z.B. Stichwortsuche, komplexe Suche, Suche in Artikel-Elementen, Vernetzung mit anderen Wörterbüchern)

sowie die Artikeldarstellung und darauf bezogene Funktionen (z.B. Mittel der Übersichtlichkeit, Ein- und Ausblendtechniken, Gliederungsübersichten, Wortverlaufskurven, Wortwolken).

Die untersuchten historischen digitalen Wörterbuchunternehmen sind: das Wörterbuchnetz und die darin organisierten Wörterbuchangebote, das Grimmsche Wörterbuch (online, CD), die Online-Angebote des Deutschen Rechtswörterbuchs, des Frühneuhochdeutschen Wörterbuchs, des Mittelhochdeutschen Wörterbuchs, des Oxford English Dictionary, des Schweizerischen Idiotikon, des Südhessischen Wörterbuchs, des Dictionary of the Scots Language sowie das Online-Wortschatz-Informationssystem des IDS, der Trésor de la Langue Française informatisé, das Ordnet, die geintegrierte Taalbank und das DWDS.

Die Steckbriefe und die darauf bezogene vergleichende Synopse (Pietsch 2020)⁷ bieten eine breit angelegte Momentaufnahme zu den zentralen Aspekten der Artikelgestaltung, der Vernetzung, zu Formen der Visualisierung und der Organisation von Zugriffsweisen in der digitalen historisch-lexikographischen Praxis im Zeitraum 2018/19.

II.7.3 Narrative Formen der Bedeutungsgeschichte

Ausgehend von Beobachtungen zu Darstellungsformen im historischen Wörterbuch, hier vor allem der Neubearbeitung des Deutschen Wörterbuchs (²DWB), wurden von Volker Harm zunächst traditionelle lexikographische Präsentationsformen auf Defizite hin analysiert und der Nutzen narrativer Formate als Alternative bzw. Ergänzung für die historische Lexikographie erörtert. Hierbei hat sich herauskristalliert, dass dieses Format weder für alle Wörterbuchtypen noch für alle Lemmata gleichermaßen geeignet ist; einen unbestreitbaren Nutzen haben narrative Darstellungsformen indes für Wörterbücher mit entwicklungsbezogener Fragestellung bzw. für Lemmata mit hoher geschichtlicher Dynamik. In einem weiteren Schritt wurden in Printwörterbüchern vorliegende narrative Darstellungen im Hinblick auf ihre Strukturen geprüft; hier hat sich vor allem die Integration unterschiedlicher lexikographischer Informationspositionen (Belegtexte, Bedeutungsangaben) in eine narrative Darstellung als problematisch erweisen. Von narrativen lexikographischen Formaten im Internet, so ist zu folgern, ist also sowohl eine hohe Modularisierung als auch eine starke Verknüpfung der einzelnen lexikographischen Informationspositionen zu erwarten. Da sich sowohl Print- als auch Online-Lexikographie grundsätzlichen Fragen des „Erzählens“ in der historischen Lexikologie zu stellen haben, sind zudem Grundzüge einer „Historik“ der Wortgeschichte erarbeitet worden, die sowohl unterschiedliche Modi des wortgeschichtlichen Erzählens als auch eine Typologie wortgeschichtlicher Fragestellungen umfasst.

⁷ https://zhistlex.de/papiere/pietsch_2020_digitale_woerterbuecher_ZHistLex.pdf

Als Anwendungsbeispiel wurde zunächst das Adjektiv *völkisch* (ein politisch stark aufgeladenes sog. „Fahnenwort“) auf der Basis des entsprechenden Artikels im ¹DWB in narrativer Darstellung bearbeitet (vgl. Harm 2020⁸). Hieran sollte u.a. demonstriert werden, a) wie Wortgeschichte und Belegpräsentation im Interesse einer besseren Lesbarkeit aufeinander bezogen werden können, b) wie im entsprechenden ¹DWB-Artikel verdeckt bleibende Informationen zur Bezeugungsgeschichte anzusprechen sind und c) wie Kontextualisierungen sprachlicher Entwicklungen eine wesentlich prominentere Stellung in der Wortgeschichte verliehen werden kann. Am Beispiel von *völkisch* wurde zugleich die Verknüpfung des Artikels mit Visualisierungen erprobt (Hoenen 2018). Als zweites Anwendungsbeispiel wurde ein Artikel zum Stichwort *Team* entworfen; damit sollte der Nutzen narrativer Formate außerhalb des im engeren Sinne politischen Wortschatzes überprüft werden. Eine Implementierung dieses Artikels wurde dann im Rahmen des BMBF-Verbundprojekts „Zentrum für digitale Lexikographie der deutschen Sprache (ZDL)“ nachgenutzt und ausgebaut. Auch die konzeptionellen Erträge zu narrativen Formaten in der Wortgeschichte sind in die ZDL-Praxis mit eingeflossen. Volker Harm hat Ergebnisse seiner Arbeit zu narrativen Darstellungsformen in der Wortgeschichte und der historischen Lexikographie auf Tagungen in Mannheim, Bullay, Saarbrücken und Göttingen zur Diskussion gestellt.

Thomas Gloning stellt am Beispiel der Bedeutungsgeschichte von *gefälligst* (Gloning 2020⁹) einen weiteren Vorschlag für ein narratives Beschreibungsformat zur Diskussion und hinterlegt den Vorschlag mit Beschreibungsvarianten. Das Wort *gefälligst* ist deshalb ein sehr guter Gegenstand, weil dabei funktionale und pragmatische Aspekte des Wortgebrauchs und seiner Entwicklung besonders berücksichtigt werden müssen (im Unterschied zu *völkisch*, wo ‚ideologische‘ Entwicklungen sehr viel stärker im Vordergrund stehen). Mit Hilfe der gut 200 Belege aus dem Deutschen Textarchiv und weiterer DWDS-Quellen wird gezeigt, wie sich das Wort von einem Mittel der Höflichkeit, das im 18. und 19. Jahrhundert ähnlich verwendet wurde wie heute *freundlicherweise*, zu einem Mittel des kommunikativen Nachdrucks werden konnte („Mach gefälligst das Fenster zu!“). Auch in diesem Beitrag werden Möglichkeiten diskutiert, wie man in einer digitalen Umgebung den narrativen Kern einer *Bedeutungsgeschichte* mit der Dokumentation von Belegen (und der Diskussion ihrer teilweise nicht unproblematischen Deutung) verbinden kann.

Ralf Plate und Roland Schuhmann erproben in einer Fallstudie aus dem Zusammenhang der Arbeiten in AP6 die Ergiebigkeit einer epochenübergreifenden Wortfamilienstruktur für Fragen der Wort- und Bedeutungsgeschichte und die Aufbereitung lexikographischer Erkenntnisse zur wortgeschichtlichen Entwicklung in diskursiver Darstel-

⁸ https://zhistlex.de/papiere/harm_2020_narrative-darstellungsformen_ZHistLex.pdf

⁹ https://zhistlex.de/papiere/gloning_2020_gefaelligst_ZHistLex.pdf

lung. Die Wortfamilie *bauen* erscheint für diese Fallstudie geeignet, weil für das Kernwort selbst die Befunde zum älteren Deutsch in den Epochenwörterbüchern (AWB, MWB, FWB) erarbeitet sind und eine entwicklungsbezogene Darstellung im Artikel des ²DWB^{A-F} vorliegt. Diese Befunde zeigen eine starke Ausweitung des Gebrauchsspektrums im Mhd. und Frnhd. und einen grundlegenden Wandel hin zum Nhd., als deren Ergebnis der gegenwartssprachliche Gebrauch mit jenem des Ahd. ('wohnen') keine Überschneidung mehr zeigt. Die Fallstudie fragt nach der Chronologie der Ausbreitung der neuen Gebrauchsweisen / Bedeutungen im Verbund der seit dem Mhd. stark anwachsenden Wortfamilie. Der entwicklungsbezogene Befund ergibt sich nicht schon durch das Nebeneinander der Einzelartikel der Epochenwörterbücher, er muss durch genaue Interpretation der Wörterbuchartikel selbst erarbeitet und darstellend vermittelt werden. Plate/Schuhmann 2018 bieten im [Abstract](#) den Aufriss der Fragestellung der Fallstudie und eine Skizze der einschlägigen lexikographischen Forschungsliteratur, die in einem [Reader](#) beigegeben ist; die ahd./mhd. [Wortfamilienübersichten](#) (mit Verknüpfung in die Epochenwörterbücher AWB und MWB) schließlich, die nach Morphologie, Überlieferungsart (im Ahd.: Glossen/Texte), Überlieferungschronologie und Semantik geordnet sind, bilden die Grundlage (und Begleitung) für die geplante diskursive Darstellung. Ihre Ausarbeitung und Online-Publikation ist geplant. Sie soll begleitet werden von interaktiven dynamischen Visualisierungen, wie sie in AP7 erprobt worden sind (vgl. unten unter II.7.4).

Gerd Fritz (JLU Gießen), der ZHistLex assoziiert ist, hat inzwischen eine Monographie „Darstellungsformen in der historischen Semantik“ fertiggestellt, in der mehrere wort- und bedeutungsgeschichtliche Fragestellungen eingehend behandelt werden, die auch zum Arbeitsprogramm des Projekts gehören. Die Monographie ist unmittelbar relevant für die Frage nach narrativen Darstellungsformen in der Historischen Semantik und nach dem Zusammenhang von Erzählen und Erklären. Darüber hinaus sind in diesem Buch auch die Ausführungen zu Formen der Visualisierung in wortgeschichtlichen Kontexten einschlägig. Das Buch wird im März oder April 2020 in der Reihe „Linguistische Untersuchungen“ der Gießener Elektronischen Bibliothek im Open-Access-Modus erscheinen.¹⁰

II.7.4 Formen der Visualisierung in der digitalen historischen Lexikographie

Formen der Visualisierung sind ein wesentlicher Charakterisierungsaspekt in den oben genannten Steckbriefen zu digitalen Wörterbucheinträgen. Wenn man die Steckbriefe „quer“ liest mit einem Fokus auf Visualisierungen, bekommt man einen guten Überblick über das Repertoire von Visualisierungsformen und -techniken, die gegenwärtig in der

¹⁰ <http://geb.uni-giessen.de/geb/volltexte/2009/7327/>

digitalen historischen Lexikographie genutzt werden. In der soeben erwähnten Monographie von Gerd Fritz werden Formen der Visualisierung in der Historischen Semantik und der Historischen Lexikographie in einem eigenen Kapitel (Kap. 5) behandelt. Wortwolken, Sterne, Entwicklungspfade usw. werden hier im Hinblick auf die jeweils eigenen Aspekte der wortgeschichtlichen Entwicklung, der dabei vorausgesetzten Konzepte (z.B. „semantische Nähe“) und auch ihrer Grenzen und Probleme diskutiert.

Armin Hoenen (AdW Göttingen) hat Visualisierungsformen im Rahmen des Projekts sowohl allgemein als auch mit Bezug auf die Anwendung als Zugriffsform in der historischen Lexikographie untersucht. Dabei wurde sowohl die allgemeine Forschungslage (vgl. Visualisierungen wie von Geeraerts 1997) wie Anwendungen in digitalen historisch-lexikographischen Angeboten (OED, DWDS usw.) betrachtet, vor allem wurden teilweise eigene Umsetzungen konzipiert, erzeugt und im ZHistLex Plenum (zahlreiche Videokonferenzen) vorgestellt und besprochen. In Hoenen (2020¹¹) wird über diese Vorschläge berichtet.

Als eine neue Zugriffsform wurde eine annotierte interaktive Timeline konzipiert und prototypisch entwickelt, mit der eine Wortgeschichte exploriert werden kann: Timeline+.¹² Sie zeigt den Frequenzverlauf eines Lexems (Y-Achse) als Kurve der relativen Korpusfrequenzen über einer Zeitleiste (X-Achse) an. Frequenzverlaufskurven werden in einer Vielzahl bestehender Online-Lexika bereits angeboten. Momentan unterstützt der Prototyp der Timeline+ den Import von DWDS-Beleglisten. In dieser erweiterten interaktiven Timeline erscheinen als Infoboxen oder farblich hinterlegte Flächen a) Ereignisse aus der Wortgeschichte, die entweder semi-automatisch extrahiert (wenn der Artikel schon besteht oder bestimmte Strukturmerkmale es möglich machen) oder vom Experten manuell annotiert sind (z.B. wenn diese Wörterbuchstrecke noch nicht erschlossen ist) und b) allgemeine Ereignisse wie Erfindungen, Naturkatastrophen und historisch bedeutende Publikationen. Die Expertenereignisse können auch mit Aspekten der narrativen Wortgeschichten korreliert und als Nebenprodukt der Artikelarbeit in einem gesonderten listenhaften Dateiformat (z.B. CSV) festgehalten werden. Allgemeine Ereignisse werden als historischer Kontext vermittelt und bieten dem Nutzer, der nicht unbedingt über ein profundes historisches Kontextwissen verfügt (bzw. verfügen muss), die Möglichkeit, eigene Zusammenhänge zu entdecken.

Kritisch anzumerken ist, dass zu jedem Zeitpunkt einerseits sehr viele Ereignisse stattgefunden haben können, so dass eine visuelle und epistemologische Informationsüberladung droht. Andererseits können Ereignisse fehlen, die die Wortgeschichte entscheidend beeinflusst haben. Eine repräsentative quantitative Analyse der für Wortgeschichten prägenden Ereignisse und von deren Charakter, sowie deren Zusammenhänge mit historischen Ereignissen aus der Forschung abzuleiten ist jedoch kaum möglich.

¹¹ https://zhistlex.de/papiere/hoenen_2020_visualisierungsformen_ZHistLex.pdf

¹² <https://zhistlex.de/timeline/>

Unbestritten ist, dass z.B. neue Erfindungen oder soziale Ereignisse Wortschatzentwicklungen bedingen und beeinflussen. Inwieweit jedoch eher wortindividuelle Ereignisse prägend für Bedeutungswandel sind, bleibt schwer zu sagen. Sicherlich bietet die historische Kontextualisierung historisch weniger erfahrenen Nutzern Anhaltspunkte für den historischen Wortgebrauch eines Lexems. Gegen die informatorische Überladung können die Ereignisse zudem kategorisiert und gezielt nach Kategorien ein- und ausgeblendet werden. Fehlende Ereignisse sind nur durch Expertenannotationen möglich. Die Timeline+ kann als z.B. Zugriffsform für den Einstieg in den entsprechenden Artikel dienen.

Im Rahmen des Projektes wurde ein funktionsfähiger Prototyp entwickelt. Dieser wurde auf dem ELexis-Workshop der ersten Konferenz der Europäischen Assoziation der Digital Humanities (EADH) vorgestellt und online publiziert, siehe Hoenen (2018b). Eine eher „künstlerische“ Herangehensweise an individuelle Visualisierungen wurde ebenfalls anhand eines Beispiels erarbeitet, vgl. Hoenen (2018a). Erkenntnisse aus den Reflexionen und experimentellen Umsetzungen von Visualisierungen in der Lexikographie waren, dass neue Zugriffs- und Präsentationsformen auch im Falle historischer Lexeme denk- und umsetzbar sind, dass aber viel Expertenwissen dazu nötig ist. Wortgeschichten müssen trotz einer ggf. lückenhaften Beleglage verantwortungsvoll rekonstruiert, rekontextualisiert und schließlich visuell umgesetzt werden. Nicht jedes Lexem besitzt zudem eine visualisierungsfähige Wortgeschichte. Visualisierungen in der historischen Lexikographie benötigen daher ggf. mehr Kuration (Kontextualisierung, Disambiguierung, Kenntlichmachen der Unsicherheiten), als dies bei modernen Lexemen der Fall wäre. Mit solchen im Rahmen von ZHistLex gewonnenen allgemeineren Reflexionsergebnissen zu Visualisierungen in der Lexikographie, die auch auf Erfahrungen aus den oben beschriebenen technischen Umsetzungen beruhen, beschäftigt sich Hoenen (2020) eingehender.

Enge Beziehungen ergeben sich in Bezug auf Formen der Visualisierung auch zum AP6, dort wurden – am Beispiel von mhd. *buwen* ‚bauen‘ – Visualisierungsformen für komplexe historische Wortfamilienzusammenhänge und für Formen der historischen Entwicklung von Wortfamilien entwickelt.

II.7.5 Zugriffs- und Erschließungsweisen

In der traditionellen gedruckten Lexikographie ist der Zwang zur Entscheidung für ein primäres Zugriffsprinzip der Ausgangspunkt für grundlegende Kontroversen gewesen: Soll man das Material nach der alphabetischen Form des Lemmas, nach Wurzeln oder nach onomasiologischen Kategorien anordnen? Und wenn man sich für ein Anordnungsprinzip entschieden hat, wie kann man ggf. andere Zugriffe über Register vorse-

hen? Das alphabetische Zugriffsprinzip hat sich in der Geschichte der historischen Lexikographie weitgehend durchgesetzt. Die oben erwähnten „Steckbriefe“ dokumentieren die gegenwärtige Anordnungs- und Zugriffspraxis.

Im ZHistLex-Projekt wurden darüber hinaus zwei weitere digitale Zugriffsstrategien ausgearbeitet. Die erste beruht auf der in AP2 entwickelten Schnittstelle (API), mit der sich gezielt Artikelpositionen ansprechen lassen. Hierbei geht es z.B. um maschinenlesbare Anfragen, die sich folgendermaßen umschreiben lassen: „Wirf alle Lesarten von Wörtern aus, zu denen es Belege im Zeitraum von 1630-1640 gibt“. Hierfür müssen Felder für Belege, für die Datierung von Belegen, für Lesarten und die darauf bezogenen Belege ansprechbar sein, was der Sinn der Abfragbarkeit von Artikelpositionen mit Hilfe der API-Spezifikation ist. Oder: „Wirf alle Artikel im Althochdeutschen Wörterbuch aus, bei denen in einem der Felder mit Bedeutungsangaben das Verb XY steht“. Anfragen dieser Art sind gebunden an vorhandene Daten und Formen der Datenauszeichnungen. Wenn, um ein Beispiel zu geben, in einem Textteil zur Wortgeschichte eines Fremdworts das Datum des Erstbelegs nicht explizit als solches ausgezeichnet ist, kann man auch keine Abfrage dazu stellen. Eine Maschine kann nicht erkennen, ob „1624“ ein Erstbelegdatum oder eine Seitenzahl in einem voluminösen Handbuch ist. Im Projekt haben wir deshalb unterschiedlich komplexe Abfrageszenarien entwickelt und sie darauf hin beurteilt, ob sie mit dem verfügbaren Stand der Datenauszeichnung umsetzbar wären. Dabei wurden schnell enge Grenzen erkennbar. Komplexere Anfragen würden eine sehr viel tiefere Annotation der bislang verfügbaren Wörterbuchdaten voraussetzen.

Für eine solche Annotation wurde im Rahmen von zwei wortschatzgeschichtlichen Vorhaben ein Verfahren entwickelt und erprobt, mit dem sich lexikologische Markierungen zur Erschließung des Vokabulars in den Daten selbst verankern lassen. Dabei wird der TEI Lex0-Standard als Grundlage verwendet. Die Markierungen werden mit Hilfe von <usg>-Elementen innerhalb des Elements <sense> formuliert, sie sind dann mit Hilfe der XQuery-Syntax auch abfragbar. Dieses Verfahren wird derzeit in einem Projekt zur Geschichte des deutschen Jazz-Vokabulars (Thomas Gloning) und in einem Dissertationsprojekt zum Wortgebrauch in Texten der ersten Frauenbewegung (Anna Pfundt) verwendet und weiterentwickelt. Auf der Lexikographie-Tagung in Göttingen und auf dem Germanistentag in Saarbrücken (2019) wurde das Projekt Jazzvokabular vorgestellt, das elektronische Vokabular zum Wortschatz des Diskurses um das Frauenwahlrecht und das dazugehörige Textcorpus im DTA wurde auf der CLARIN-Konferenz in Leipzig (2019) vorgestellt. Ein entsprechender Beitrag für den Konferenzband ist eingereicht. Eine XML-Kodierung mit den internen Markierungen sieht derzeit so aus:

```

<entry>
<form type="lemma"><orth>Hardbop-Pianist</orth></form>
<gramGrp><pos>NN</pos></gramGrp>
<sense>
<def>JazzmusikerIn, der/die Piano spielt und der Stilrichtung Hardbop zuzuordnen ist.</def>
<usg type="textsorten">Sachbuch</usg>
<usg type="systematik">Person</usg>
<usg type="systematik">Jazzmusiker</usg>
<usg type="systematik">Instrumentalist</usg>
<usg type="wortbildung">N+N</usg>
<usg type="herkunft">Siehe die Komponenten</usg>
<cit type="quotation">
<bibl><date>2016</date>Kampmann, Jazz<biblScope>S. 104</biblScope></bibl>
<quote>1956 verschlimmerte sich sein Zustand abermals, als sein jüngerer Bruder, der gefei-
erte <mentioned>Hardbop-Pianist</mentioned> Richie Powell, bei einem Autounfall zu Tode
kam.</quote></cit>
</sense>
</entry>

```

Es zeichnet sich ab, dass für die Aufgabe einer Stand-alone-Erschließung historischer Wörschätze auch Linked Open Data- bzw. Linguistic Linked Open Data-Technologien (LOD, LLOD) einsetzbar wären, wie sie u.a. von Sabine Tittel und Christian Chiarcos am Beispiel des Dictionnaire Etymologique de l’Ancien Français (DEAF) vorgeschlagen wurden. Eine Markierung wie `<usg type="systematik">Jazzmusiker</usg>` lässt sich auch als RDF-Tripel aus einem Identifikator für eine `<sense>`-Einheit, für eine Relation („gehört zur Jazzwortschatz-Systemstelle“) und einen Wert (Jazzwortschatz-Systemstelle ‚Bezeichnung für Arten von JazzmusikerInnen‘) formulieren.

II.7.6 Nutzen und Verwertbarkeit

Das Projekt verfolgte keine direkt wirtschaftlich verwertbaren Ziele. Seine Ziele und die darauf bezogenen Resultate beziehen sich auf die aktuelle digitale historisch-lexikographischen Praxis, ihre Systematisierung und auch ihre Weiterentwicklung. Insofern sind die Resultate des Projekts ein Beitrag zur Systematisierung von Darstellungsweisen, Zugriffsformen, Weisen der Visualisierung usw. in der digitalen historisch-lexikographischen Praxis sowie zu Perspektiven der Weiterentwicklung dieser Praxis.

Im Hinblick auf viele automatisierte Verfahren stellt derzeit die Differenzierung von Verwendungsweisen ein schwer überwindbares Nadelöhr dar. Automatisch erstellte Wortverlaufskurven beziehen sich derzeit auf die Einheit „Wort“, ohne dass die unterschiedlichen „Lesarten“ bzw. „Bedeutungspositionen“ hierbei berücksichtigt werden

können. Im Hinblick auf polyseme Wörter und bei Wörtern mit einer komplexen Bedeutungsgeschichte wird diese digital erzeugte Darstellungsform quasi wertlos, ohne dass diese gravierende Einschränkung in der bisherigen Diskussion um digitale Darstellungsformen bislang eingehend diskutiert worden wäre.

Es gehört deshalb mit zu den Aufgaben der weiteren Schritte, auch die Grenzen digitaler Darstellungsmittel und -verfahren ggf. sehr viel deutlicher zu formulieren. Die Formulierung von aktuellen Grenzen ist möglicherweise auch ein wichtiger Ausgangspunkt für zukünftige Forschungen. Ein Beispiel: Historische Kookkurrenzprofile sind sicherlich ein guter heuristischer Anknüpfungspunkt für wortgeschichtliche Entwicklungen, aber die Frage, wie historische Kookkurrenzprofile und ihre Entwicklung genau mit den eigentlichen Gebrauchsprofilen von Wörtern zusammenhängen, ist erst noch genauer zu klären.

II.7.7 Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen

Insgesamt war in den drei Jahren eine hohe Dynamik der Forschung im Bereich der elektronischen Lexikographie zu verzeichnen, aber nur sehr wenig Aktivitäten mit direktem Bezug zur historischen Lexikographie (des Deutschen). Ausnahmen sind die oben genannten Arbeiten von Sabine Tittel und Christian Chiarcos, aber auch die konzeptionellen und technischen Arbeiten im Projekt „Zentrum für Digitale Lexikographie“ (ZDL).

II.8 Erfolgte und geplante Veröffentlichungen, Vorträge, Präsentationen

II.8.1 Erfolgte und geplante Veröffentlichungen

Wesentliche Ergebnisse des Projekts sind auch auf der ZHistLex-Webseite dokumentiert:

<https://zhistlex.de>

Hier vor allem die Spezifikation der Webservices und der ZHistLex-APIs.

[AP2](#), [AP3](#), [AP4](#), [AP5](#), [AP6](#), [AP7](#), [ZHistLex](#)

Fritz, Gerd: Darstellungsformen in der historischen Semantik. Gießen: Gießener Elektronische Bibliothek 2020. (Monographie, im Erscheinen.)

[AP7](#)

Gloning, Thomas (2020): „gefälligst“ – Eine neue Bedeutungsgeschichte. Zugleich ein Beitrag zu narrativen Formaten in der historischen Semantik und Lexikographie.

https://zhistlex.de/papiere/gloning_2020_gefaelligst_ZHistLex.pdf

[AP7](#)

Gloning, Thomas (2020; in Vorbereitung): Perspektiven für ein digitales „Sudetendeutsches Wörterbuch“. Ausarbeitung zum Vortrag auf der Tagung des Verbunds LexikoNet 2, JLU Gießen, 16.11.2018. Für diesen Beitrag spielen die ZHistLex-Webservices und konzeptionelle Aspekte aus dem ZHistLex-Projekt eine wichtige Rolle.

[ZHistLex](#), [AP7](#)

Harm, Volker (2020): Narrative Darstellungsformen der digitalen Lexikographie.

https://zhistlex.de/papiere/harm_2020_narrative-darstellungsformen_ZHistLex.pdf

[AP7](#)

Hoenen, Armin (2018): “Annotated Timelines and Stacked Area Plots for Visualization in Lexicography” in Proceedings of the Elexis workshop, collocated with EADH Gallway.

https://lexdhai.insight-centre.org/Lex_DH_AI_2018_paper_2.pdf

[AP7](#)

Hoenen, Armin (2018): “Attempts at Visualization of Etymological Information” in Proceedings of the 11th International Conference on Language Resources and Evaluation, 2018.

https://lrec-conf.org/workshops/lrec2018/W33/pdf/9_W33.pdf

[AP7](#)

Hoenen, Armin (2020): Einige neue explorative Visualisierungsformen in der digitalen historischen Lexikologie und Lexikographie.

https://zhistlex.de/papiere/hoenen_2020_visualisierungsformen_ZHistLex.pdf

[AP6](#), [AP7](#)

Pfundt, Anna, Melanie Grunt Suárez & Thomas Gloning (2020): Word usage in German texts on women’s suffrage around 1900. Corpus building, lexical documentation and the CLARIN-D infrastructure. Ausarbeitung eines Beitrags zur Clarin Annual Conference,

Leipzig 2019. (Status: accepted für den Tagungsband)

AP7, ZHistLex und Clarin-D, ZHistLex und Nachwuchsförderung

Pietsch, Andre (2020): Historische Wörterbücher und Wörterbuchsysteme in digitalen Umgebungen. Komponenten, Verknüpfungen, Darstellungsmittel, Nutzungsoptionen.

https://zhistlex.de/papiere/pietsch_2020_digitale_woerterbuecher_ZHistLex.pdf

AP7

II.8.2 Fachöffentliche Vorträge und Präsentationen

Gloning, Thomas (Koord.): Panel „Wörterbücher – Netze und Nachnutzung“. Workshop „Nachnutzung“ der AG eHumanities der Akademienunion, 11.11.2017.

Vorstellung des ZHistLex-Projekts

ZHistLex

Gloning, Thomas: Historische Lexikographie im digitalen Medium. Neue Forschungen zu Repräsentationsformen zeitlicher Dynamik des Wortgebrauchs. Eröffnungsbeitrag zum ZHistLex-Panel beim [26. Deutscher Germanistentag, Saarbrücken, 22.-25.9.2019.](#)

ZHistLex

Gloning, Thomas: Historisches Vokabular des Jazz. Wort(schatz)geschichte in kommunikativ-kulturellen Zusammenhängen und die Frage nach digitalen Arbeitsumgebungen. Vortrag auf der Tagung [“Historische Lexikographie: Potentiale, Perspektiven. Zur Zukunft der deutschen Wörterbücher in den Akademien der Wissenschaften”](#), Göttinger Akademie der Wissenschaften, 18.9.2019, und in leicht revidierter Fassung im ZHistLex-Panel beim [26. Deutscher Germanistentag, Saarbrücken, 22.-25.9.2019.](#)

AP7

Gloning, Thomas: Historische Wörterbücher im digitalen Zeitalter. Perspektiven aus dem Projekt „ZHistLex“. Vortrag auf der Tagung des Verbunds LexikoNet 2, JLU Gießen, 16.11.2018. Ein schriftlicher Beitrag zum Tagungsband ist in Vorbereitung.

ZHistLex, AP7

Harm, Volker: Das *Deutsche Wörterbuch* der Brüder Grimm. Perspektiven der Nachnutzung. Vortrag beim Workshop „Nachnutzung“ der AG eHumanities der Akademienunion, 11.11.2017.

AP3, AP7

Harm, Volker: Erzählte Wortgeschichte als Darstellungsform der historischen Lexikographie. Vortrag 28.4.2018, [Arbeitsgespräch zur historischen Lexikographie 2018, Bullay.](#)

AP7

Harm, Volker: Narrative Darstellungsformen für Bedeutungsgeschichten. Vortrag 12.6.2019, [“Workshop Darstellungsformen und Quellengrundlagen, Zugriffs- und Wortschatzstrukturen, Visualisierung im Wörterbuch” am 12.6.2019 am Institut für deutsche Sprache Mannheim](#)

AP7

Harm, Volker: *Wortgeschichte digital* – Grundzüge des Projekts. Vortrag auf der Tagung [“Historische Lexikographie: Potentiale, Perspektiven. Zur Zukunft der deutschen Wörterbücher in den Akademien der Wissenschaften”](#), Göttinger Akademie der Wissenschaften, 18.9.2019

AP7

Harm, Volker: Narrative Darstellungsformen für Bedeutungsgeschichten. Vortrag [26. Deutscher Germanistentag, Saarbrücken, 22.-25.9.2019](#) (Panel „Historische Lexikographie im digitalen Medium. Neue Forschungen zu Repräsentationsformen zeitlicher Dynamik des Wortgebrauchs“)

AP7

Herold, Axel: TEI für Wörterbücher. Vortrag auf der 25. Jahrestagung der ITUG, 3.–6. Oktober 2018, Potsdam.

AP2

Herold, Axel: Lehrtätigkeit im Rahmen der Lexical Data Masterclass 2018 (<https://lexmc18.sciencesconf.org/>), 3.–7. Dezember 2018, Berlin.

AP2, ZHistLex

Hinkelmanns, Peter: Der onomasiologische (Aus-)Blick. Verzweigungen, Zusammenführungen, Neubildungen. Bedeutungswandel untersucht im digitalen Korpus der Mittelhochdeutschen Begriffsdatenbank. Vortrag Vortrag [26. Deutscher Germanistentag, Saarbrücken, 22.-25.9.2019](#) (Panel „Historische Lexikographie im digitalen Medium. Neue Forschungen zu Repräsentationsformen zeitlicher Dynamik des Wortgebrauchs“)

Kooperation ZHistLex/Mittelhochdeutsche Begriffsdatenbank

Hinkelmanns, Peter: Althochdeutsches Wörterbuch *digital*. Die Digitalisierung des althochdeutschen Wörterbuchs und dessen Integration in das Trierer Wörterbuchnetz. Vortrag beim Workshop „Nachnutzung“ der AG eHumanities der Akademienunion, 11.11.2017.

AP3

Hoenen, Armin, Roland Mittmann & Ralf Plate, Zeitliche Entfaltung von Polysemie und Wortfamilien-Ausbau. Digitale historische Wortfamilien-Lexikographie am Beispiel von *bauen* im Alt- und Mittelhochdeutschen. Vortrag auf dem [26. Deutscher Germanistentag, Saarbrücken, 22.-25.9.2019](#)

AP5, AP7

Mittmann, Roland (2018): [Neue Nutzungsmöglichkeiten des Referenzkorpus Altdeutsch und seiner Materialien](#). Arbeitsgespräch zur historischen Lexikographie am 28.04.2018 auf der Marienburg bei Bullay (Mosel).

AP5

Mittmann, Roland (2019): Berichte und Diskussion über die Arbeiten des eHumanities-Zentrums für historische Lexikographie (zhistlex.de): Lesekorpus Altdeutsch (LeA) und epochenübergreifende Wortfamilienstrukturen. [Arbeitsgespräch zur historischen Lexikographie](#) am 27.04.2019 auf der Marienburg bei Bullay (Mosel).

AP5, AP6

Mittmann, Roland & Ralf Plate: Online Dictionaries and Digital Text Corpora on Older German. On their Consolidation and combined Application

[42. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Hamburg, 4-6.3.2020](#)

AP2, AP5, AP6

Mittmann, Roland & Ralf Plate, Epochenübergreifende Wortfamilienstrukturen Althochdeutsch-Mittelhochdeutsch. Vortrag auf dem

[“Workshop Darstellungsformen und Quellengrundlagen, Zugriffs- und Wortschatzstrukturen, Visualisierung im Wörterbuch” am 12.6.2019 am Institut für deutsche Sprache Mannheim](#)

AP6

Mittmann, Roland & Ralf Plate, [Das ‚Referenzkorpus Altdeutsch‘ als Lesekorpus. Grammatisch annotierte und mit Wörterbüchern verknüpfte Texte für Lehre und Selbststudium](#), in: [Das Mittelalter](#) 24/1 (2019) [Themenheft: Digitale Mediävistik], S. 173-187.

AP5

Plate, Ralf: Philologische Theorie und Praxis der digitalen historischen Lexikographie. Am Beispiel des Mittelhochdeutschen Wörterbuchs. Vortrag auf der Tagung

[“Historische Lexikographie: Potentiale, Perspektiven. Zur Zukunft der deutschen Wörterbücher in den Akademien der Wissenschaften”](#), Göttinger Akademie der Wissenschaften, 18.9.2019

AP5

Plate, Ralf: Die Vernetzung der digitalen Angebote zur historischen Lexikographie des mittelalterlichen Deutsch. Vortrag auf der Tagung

[“Mittelalter im Fokus. Chancen und Perspektiven des Akademienprogramms für die mediävistische Forschung im 21. Jahrhundert”](#) der Bayerischen Akademie der Wissenschaften, 12.9.2019.

AP2, AP5, AP6

Plate, Ralf & Roland Schuhmann, Wortgeschichte und Wortfamilie am Beispiel *bauen* im Alt- und Mittelhochdeutschen. Referat auf dem [Arbeitsgespräch zur Historischen Lexikographie 2018, Bullay \(Mosel\), 27.-29.4.2018](#). [Abstract](#)|[Handout](#)

AP6, AP7

Recker-Hamm, Ute: Wie finden Wörterbücher und Korpora im Internet zusammen? Technische Lösungen und ihre Folgen für die Konzeption der lexikographischen Vernetzung. Referat auf dem [Arbeitsgespräch zur Historischen Lexikographie 2018, Bullay \(Mosel\), 27.-29.4.2018](#). [Abstract](#) | [Folien](#)

[Abstract](#) | [Folien](#)

AP2, AP4, AP5

Recker-Hamm, Ute: Bericht über die Arbeiten des eHumanities-Zentrums für historische Lexikographie. Kurzreferat auf dem [Arbeitsgespräch zur Historischen Lexikographie 2019](#), Bullay (Mosel), 26.-28.4.2019.

AP2, AP4, AP5

Anlagen

Anlage III Erfolgskontrollberichte der beteiligten Projekt-Partner

Anlage IV Berichtsblatt und Document Control Sheet zum Schlussbericht