

URN: urn:nbn:de:kobv:b4-opus-24415

HANS-CHRISTIAN SCHMITZ, BERNHARD SCHRÖDER & KLAUS-PETER
WEGERA,
Das Bonner Frühneuhochdeutsch-Korpus und das Referenzkorpus
,Frühneuhochdeutsch‘,

in:

*Perspektiven einer corpusbasierten historischen Linguistik und Philologie.
Internationale Tagung des Akademienvorhabens „Altägyptisches Wörter-
buch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften,
12. – 13. Dezember 2011*, herausgegeben von Ingelore Hafemann,
Berlin 2013, S. 205-219.

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Thesaurus Linguae Aegyptiae 4

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

herausgegeben von Ingelore Hafemann

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Thesaurus Linguae Aegyptiae

4

BERLIN 2013

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Perspektiven einer corpusbasierten historischen Linguistik
und Philologie

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

herausgegeben von Ingelore Hafemann

BERLIN

2013

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

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

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



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

INHALTSVERZEICHNIS

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

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

VORWORT

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

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

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

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

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

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

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

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

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

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

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

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

Berlin, Mai 2013

Ingelore Hafemann

DAS BONNER FRÜHNEUHOCHDEUTSCH-KORPUS UND DAS REFERENZKORPUS ‚FRÜHNEUHOCHDEUTSCH‘

HANS-CHRISTIAN SCHMITZ, BERNHARD SCHRÖDER & KLAUS-PETER WEGERA

1. Einleitung

Das Bonner Frühneuhochdeutsch-Korpus

Es gibt zwei Bonner Korpora des Frühneuhochdeutschen, nämlich erstens ein ‚großes‘ Korpus aus 1500 Texten, welches wir hier das Gesamtkorpus nennen. Das Gesamtkorpus ist nicht in maschinenlesbarer Form vorhanden. Zweitens gibt es ein elektronisches Korpus, das aus 40 Texten des Gesamtkorpus besteht. Wir nennen dieses Korpus das elektronische Teilkorpus oder kürzer das Teilkorpus.

Das Gesamtkorpus wurde zwischen 1972 und 1974 an der Forschungsstelle Frühneuhochdeutsch der Universität Bonn zusammengestellt, um „in großer Breite die Textüberlieferung des 14.-17. Jahrhunderts für sprachgeschichtliche Untersuchungen verschiedenster Art bereit[zu]stellen“ (HOFFMANN & WETTER, 1985, XIV). Die Texte des Gesamtkorpus entstammen 22 Sprachlandschaften und sieben Zeitschnitten à 50 Jahre (1350-1700).

Das Auswahlkorpus entstand zwischen 1972 und 1985 im Rahmen des Projekts ‚Flexionsmorphologie des Frühneuhochdeutschen‘ unter Leitung von Werner Besch, Winfried Lenders (ab 1976), Hugo Moser und Hugo Stopp (bis 1981). Es diente als Materialgrundlage zur Analyse der Flexionsmorphologie des Frühneuhochdeutschen und zur Erarbeitung von mehreren Bänden der Grammatik des Frühneuhochdeutschen, nämlich zur Flexion der Substantive (WEGERA, 1987), zur Flexion der starken und schwachen Verben (DAMMERS *et al.*, 1988) und zur Flexion der Adjektive (SOLMS & WEGERA, 1991). Mit Fertigstellung dieser Bände hatte das Korpus seinen primären Zweck erfüllt. Der Datensatz wurde daraufhin nicht mehr genutzt, bis er im Jahr 2002 nach XML und HTML transferiert und über das WWW öffentlich bereitgestellt wurde. Das Korpus ist seitdem in verschiedenen Formen und über verschiedene Schnittstellen allgemein verfügbar. Es wird in Forschung und Lehre genutzt.

Das Referenzkorpus ‚Frühneuhochdeutsch‘

Im Rahmen eines im Oktober 2011 begonnenen Projekts wird an der Ruhr-Universität Bochum in Zusammenarbeit mit den Universitäten

Halle und Potsdam ein Referenzkorpus des Frühneuhochdeutschen erstellt. Dieses Korpus wird mit über 300 Texten vom Umfang her kleiner sein als das Bonner Gesamtkorpus, aber weitaus größer als das elektronische Teilkorpus. Das Bonner Teilkorpus spielt für die Konstruktion des Referenzkorpus eine zentrale Rolle, da ein Großteil seiner Texte übernommen, korrigiert und an die Standards für Referenzkorpora angepasst wird. Das Referenzkorpus wird das Bonner Korpus mittelfristig in Forschung und Lehre ablösen.

Gliederung des Artikels

Der vorliegende Artikel ist wie folgt gegliedert: In Abschnitt 2 referieren wir kurz den Umfang und die Konzeption des Bonner elektronischen Teilkorpus. Daraufhin erläutern wir in Abschnitt 3 die XML-Kodierung und die Zugriffsmöglichkeiten auf das Korpus, insbesondere über eine HTML-Darstellung für die explorative Erschließung (FnhdC/HTML) und über eine Maske für die gezielte Durchsuchung des Korpus (FnhdC/S). In Abschnitt 4 nennen wir beispielhaft einige Anwendungen in Forschung und Lehre, bevor wir in Kapitel 5 den Blick auf das Referenzkorpus ‚Frühneuhochdeutsch‘ und damit die Zukunft des Bonner Korpus richten. Wir übernehmen in diesem Artikel Passagen aus dem Arbeitsbericht zur XML-Kodierung des Bonner Frühneuhochdeutsch-Korpus von DIEL *et al.* (2002).

2. Konzeption und Umfang des Bonner Teilkorpus

Die Anfänge des Bonner Frühneuhochdeutsch-Korpus liegen im Beginn der 1970er Jahre. Die Grundlagen, die Zusammenstellung und die Probleme dieser frühen Phase der computerunterstützten Grammatikographie sind ausführlich dokumentiert in SOLMS & WEGERA (1998).

Das elektronische Teilkorpus besteht aus 40 Texten, respektive Textausschnitten. Die Texte sind nicht immer vollständig in das Korpus aufgenommen, sondern jeweils zu einem Ausschnitt von ca. 30 Normalseiten mit etwa 400 Wortformen. Der relativ kleine Umfang ist durch den damals großen Aufwand der Digitalisierung begründet. Der Umfang reicht allerdings aus, um die flexionsmorphologische Exemplarität des Korpus zu gewährleisten und dadurch seinen Zweck als Materialgrundlage für die Erarbeitung von drei der vier flexionsmorphologischen Bände der Grammatik des

Frühneuhochdeutschen zu erfüllen.¹ Die Texte sind durchgängig mit Wortklassen- und Formenbestimmungen versehen (s.u., Abschnitt 3).

Die Korpustexte entstammen zehn verschiedenen Sprachlandschaften, nämlich dem Mittelbairischen (Wien), Schwäbischen, Ostfränkischen (Nürnberg), Obersächsischen, Ripuarischen (Köln), Osthochalemannischen, Oberschwäbischen (Augsburg), Elsässischen (Straßburg), Hessischen und Thüringischen. Die Texte entstammen außerdem vier Zeitschnitten à 50 Jahre, nämlich 1350-1400, 1450-1500, 1550-1600 und 1650-1700. Jede Sprachlandschaft ist mit insgesamt vier Texten vertreten, wobei diese Texte aus den verschiedenen Zeitschnitten stammen; entsprechend ist jeder Zeitschnitt mit zehn Texten aus den zehn verschiedenen Landschaften vertreten.

Das Korpus wurde nachträglich ergänzt durch insgesamt 20 Texte für die Zeiträume 1500-1550 und 1600-1650 (mit jeweils einem Text pro Sprachlandschaft); darüber hinaus wurden 22 weitere Texte für die zusätzlichen Sprachlandschaften (Nord-)Hessisch, Schlesisch, Westhochalemannisch, Mittelbairisch 2 (München) und den norddeutschen Raum zusammengestellt. Diese Ergänzungen wurden allerdings nicht digitalisiert und sind ergo im elektronischen Teilkorpus nicht vorhanden. Auch nachträgliche Korrekturen wurden nicht immer in die elektronische Fassung eingearbeitet.

Das Korpus stellt eine authentische Mischung verschiedener Textsorten dar: Insgesamt sind fünf kirchlich-theologische Fachtexte, zwölf chronikalische und Berichtstexte, zwölf erbauliche Texte, ein Bibeltext, sechs unterhaltsame Texte (‘schöne Literatur’) und vier Realientexte (Fachprosa) vorhanden. Es fehlen metasprachliche Texte, Texte mit gebundener Sprache, Privattexte, Rechts- und Geschäftstexte. Die Authentizität der Textsortenmischung gilt für das gesamte Teilkorpus. Innerhalb der Sprachlandschaften und der Zeitschnitte liegt allerdings eine weitgehend zufällige Textsortenmischung vor. Wenn also beispielsweise die Texte eines Zeitschnitts verglichen werden, dann können festgestellte Unterschiede nicht ohne weiteres als Unterschiede zwischen den verschiedenen Sprachlandschaften gedeutet werden. Es ist möglich, dass die Unterschiede vielmehr textsortenbedingt sind. Wollte man nun den Einfluss der Textsorte überprüfen, müssten für jede Sprachlandschaft und jeden Zeitschnitt systematisch alle verschiedenen Textsorten vorliegen. Das ist nicht der Fall.

¹ Band V der Grammatik des Frühneuhochdeutschen – WALCH (1988) – basiert nicht auf dem Korpus.

Die Auswahl der Texte wurde an strenge Qualitätskriterien gebunden, die allerdings nicht immer eingehalten werden konnten. Grundsätzlich galt, dass ein Text nur dann aufzunehmen sei, wenn erstens seine Datierung und Lokalisierung belegt oder zu ermitteln war und zweitens der Verfasser bzw. Hersteller aus der Entstehungslandschaft stammte oder wenigstens zehn Jahre lang in ihr tätig war. Schließlich sollten zwischen dem Verfassen des Originaltexts und der Herstellung der gegebenen Textfassung nicht mehr als 50 Jahre liegen.

Für die Digitalisierung der Texte – hierin liegt zweifellos ein Manko des Korpus – wurden teilweise Editionen verwendet, ohne dass diese mit den originalen Handschriften abgeglichen wurden.

Die Zusammenstellung des Korpus hat sich für die Flexionsmorphologie insbesondere der Substantive bewährt. Die Belegdichte für Adjektive und Verben ist hingegen geringer und manchmal sogar unbefriedigend. SOLMS & WEGERA (1998: 27) stellen fest, dass das Korpus „bezüglich der Belegverhältnisse in seiner Gesamtheit zumeist hinreichend [ist]; unzureichend wird es bei grundsätzlich oder textsortenbezogen wenig frequenten Phänomenen. Die Erfahrung zeigt jedoch auch, daß solche Probleme durch entsprechende Zusatzerzertion gelöst werden können.“ Während die flexionsmorphologischen Phänomene relativ unabhängig von Textsorten vorkommen, sind Wortbildungsphänomene stark textsortenabhängig. Die fehlende Textsortenorientierung wird dabei bei der Untersuchung von Wortbildungsphänomenen spürbar. Das Korpus kann hier nur bedingt als Materialgrundlage dienen.

Fassen wir die Vor- und Nachteile des Korpus zusammen: Der entscheidende Vorteil des Bonner Teilkorpus ist, dass es existiert. Es ist frei verfügbar, kann gelesen und durchsucht werden und ist insofern schon mangels Alternativen eine wichtige Ressource der Sprachwissenschaft. Es handelt sich bei ihm ferner um einen gut durchdachten, wohl strukturierten Bestand, der sowohl einen hohen Qualitätsstandard hinsichtlich der Quellenauswahl als auch der Exemplarität bezüglich flexionsmorphologischer Phänomene beanspruchen kann. Das Korpus aus den 1970er Jahren wird allerdings nicht den Standards moderner Referenzkorpora gerecht: Seine Texte basieren z.T. auf Editionen, und sie sind nicht vollständig, sondern nur in Ausschnitten im Korpus vorhanden. Das Korpus hat mit 40 Teiltextrn einen relativ geringen Umfang, so dass es zur systematischen Untersuchung von Phänomenen jenseits der Flexionsmorphologie nur bedingt geeignet ist. Hier fällt insbesondere die umfangs-

bedingt geringe Repräsentanz einzelner Textsorten ins Gewicht, aufgrund derer es nicht die Materialgrundlage für die erschöpfende Untersuchung von textsortenabhängigen Phänomenen wie solchen der Wortbildung bieten kann. Schließlich wurden Korrekturen und Ergänzungen des Korpus nur teilweise eingearbeitet, so dass die vollständige Richtigkeit des Korpus nicht immer garantiert ist.

3. Kodierung und Zugriffsmöglichkeiten

Kommen wir zur Kodierung der digitalisierten Texte und den Möglichkeiten des Zugriffs auf die Daten: Die erste Version des elektronischen Teilkorpus wurde beginnend 1972 auf Lochkarten gespeichert, dann in den Zeichensatz der MS-DOS Codepage 437 übertragen und auf Disketten gespeichert. Die durch diese Übertragung entstandene zweite Version nennen wir die Original-kodierte Version.² Die Lochkarten sind unseres Wissens nicht erhalten. 2002 wurde die Original-kodierte Version vollständig nach XML transformiert. Maßgabe der Transformation war, die Kodierung in den neuen Standard zu überführen und dabei inhaltlich so wenig wie möglich zu verändern. Offensichtliche Fehler wurden jedoch korrigiert und Kodierungen in den Originaltexten, deren Bedeutung trotz Konsultation von LENDERS & WEGERA (1982) und Befragung ehemaliger Projektmitglieder nicht festgestellt werden konnte, wurden gelöscht.³ Auf Grundlage der XML-kodierten Version wurde im gleichen Jahr eine ‚lesbare‘, an Information reduzierte HTML-Version erstellt. Im Jahre 2007 wurde die XML-Version einer Korrektur unterzogen. Darüber hinaus wurde eine neue HTML-Version erzeugt (FnhdC/HTML, s.u.) und eine Suchmaschine für das Korpus implementiert (FnhdC/S, s.u.). Alle vorhandenen Versionen des Korpus – die Original-kodierte Version, die XML-Version und die beiden HTML-Versionen – wurden als komprimierte Archive über die Webseite korpora.org verfügbar gemacht.⁴

Die XML-Kodierung ist in einer kommentierten Document Type Definition (DTD) und in DIEL *et al.* (2002) beschrieben.⁵ In den

² Sie stimmt nicht in allen Punkten mit der von BERG (1982) beschriebenen ersten Version überein.

³ Den ehemaligen Projektmitgliedern war klar, dass nicht alle Korrekturen in das elektronische Teilkorpus eingearbeitet wurden und dass es ergo nicht vollständig richtig ist. Daher rührte ihre Skepsis bezüglich der Veröffentlichung des Korpus.

⁴ <http://www.korpora.org/Fnhd>

⁵ Die DTD befindet sich zusammen mit den XML-Dateien im komprimierten Archiv. Ferner steht sie im Anhang von DIEL *et al.* (2002).

Texten sind Seiten, Blätter, Kapitel und Zeilen markiert und nummeriert.⁶ Darüber hinaus sind Eingriffe (Bearbeitungen), Hervorhebungen, Überschriften, Zitate und Namen als solche annotiert. Endlich sind Substantive, Adjektive und Verben mit morphosyntaktischen Angaben versehen. Dazu gehören erstens die Wortklassen. Zweitens werden für Adjektive und Substantive, deren Formen bestimmt werden konnten, Kasus, Numerus und Genus angegeben. Entsprechend werden von Verben, deren Formen bestimmt werden konnten, Tempus, Modus, Person und Form (Infinitiv, Partizip, finite Form) genannt. Drittens sind Flexive, Lemmata, Prä- oder Suffixe in normalisierter Form und die Vokale der Stammsilbe gesondert aufgeführt.

Zu mehreren Wortformen sind im Annotationsteil des Original-kodierten Korpus zwei Wortklassen angegeben. Dabei handelt es sich zumeist um substantivierte Infinitive, die sowohl als Substantive als auch als Infinitive annotiert sind. Beide Angaben, für den Erst- und den Zweittyp, wurden in die XML-Version überführt.

Die Zeichen der XML-Version sind gemäß dem Unicode™-Standard in UTF-8 kodiert. Auf Sonderzeichen wird mit Referenzen verwiesen. Alle Zeichen mit Diakritika sind aufgespalten. Sie bestehen aus Grundzeichen und darauf folgendem Diakritikum. Ein „ä“ ist somit nicht als einzelnes Zeichen kodiert (wie in ISO-8859-1), sondern als „a“ gefolgt vom Diakritikum „̈“; ein „å“ ist kodiert als „a“ gefolgt vom Diakritikum „̊“.

Alle Texte des Korpus sind über ein Quellenverzeichnis erschlossen, das wie die Texte selbst in XML- und HTML-kodierten Fassungen vorliegt. Die XML-kodierten Daten sind aufgeteilt in Basis- und Zusatzdaten. Die Basisdaten umfassen Kurztitel, Titel, Autoren, Quellen und Herausgeber, Erscheinungsorte, -jahre und andere Editionsspezifika, Sprachlandschaften, Entstehungszeiten sowie die Anzahl der aufgenommenen Seiten. Die Zusatzdaten umfassen editorische Anmerkungen, eventuelle Vorlagen, Biographien der Verfasser, Übersetzer, Schreiber und Drucker, die Textarten sowie eventuelle editorische Eingriffe (z.B. Kürzungen).

Aus der 2002er HTML-Version des Korpus wurden alle linguistischen Annotationen entfernt, um die Texte leicht lesbar zu machen. Diese Lösung war nicht optimal, weil die für die Rezeption wertvolle linguistische Zusatzinformation verloren ging. Deshalb wurde 2007

⁶ Genauer: Zur Vermeidung konkurrierender Strukturen sind Seiten-, Blatt-, Kapitel- und Zeilenwechsel markiert.

eine neue Version erstellt (FnhdC/HTML, vgl. FISSENI *et al.*, 2007), die sowohl lesbar ist als auch die linguistischen Annotationen enthält. In ihr werden die linguistischen Annotationen für eine Wortform angezeigt, wenn man den Mauszeiger auf die entsprechende Wortform legt. Abbildung 1 zeigt einen Ausschnitt von Wilhelm Durandus' Rationale (Text 111⁷ des Korpus) in dieser Darstellung. Der Mauszeiger steht hier auf dem Wort „chünig“. Neben dem Text werden die zugehörigen morphosyntaktischen Informationen wie der Stammvokal, die Wortart und wortartsspezifische Angaben, hier Kasus, Numerus und Genus, angezeigt. Außerdem zeigt die Lemmaangabe das Wort in der neuhochdeutschen Schreibung.

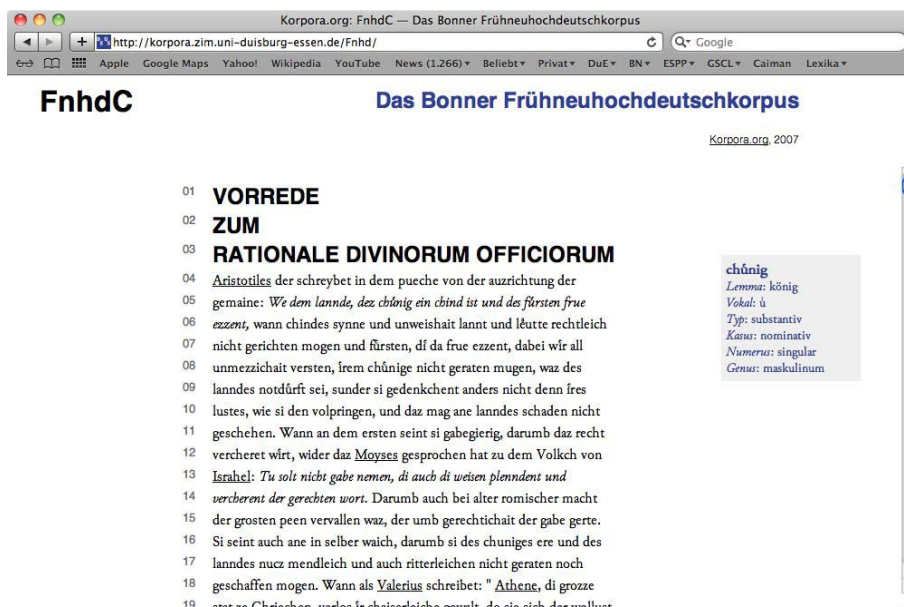


Abbildung 1: Darstellung eines Ausschnitts von Text 111 (Wilhelm Durandus: Rationale, Wien 1384) in FNHD/HTML.

⁷ Zur Nummerierung der Korpustexte: Das Korpus ist in zwei Hälften à 20 Texte geteilt. Der erste Teil trägt die Nummer 1, der zweite die Nummer 2. In jedem Teil sind Texte aus fünf Sprachlandschaften vertreten, jeweils nummeriert von 1-5. Schließlich wurden die Zeitschnitte mit den Zahlencodes 1, 3, 5, und 7 versehen. Die Nummer eines Textes ist eine dreistellige Ziffernfolge, bestehend aus dem Code für den Teil des Korpus, die Sprachlandschaft und den Zeitschnitt. Durandus' Rationale trägt die Nummer 111, weil der Text dem ersten Teil, darin der ersten Sprachlandschaft (Mittelbairisch, Wien) und dem ersten Zeitschnitt (1350-1400) entstammt.

Zusätzlich wurde eine Suchmaschine implementiert, mittels derer das Korpus und seine einzelnen Texte durchsucht werden können (FnhdC/S, vgl. FISSENI *et al.*, 2007). Die Herausforderung bestand darin, einerseits komplexe Suchanfragen auch unter Einbezug der Annotationen zuzulassen und andererseits die effektive Nutzbarkeit auch für technisch wenig affine Linguisten zu gewährleisten, die Bedienung also einfach zu machen. Diese Aufgabe wurde mithilfe einer Suchmaske gelöst, die die Eingabe von (Mustern von) Wortformen oder Lemmata und die Auswahl linguistischer Eigenschaften (Kasus, Numerus, etc.) ermöglicht. Abbildung 2 zeigt einen Ausschnitt der Suchmaske mit Eingaben für die Suche nach Substantiven im Dativ Singular mit dem Lemma „erde“. Abbildung 3 zeigt einen Ausschnitt des entsprechenden Suchergebnisses.

The screenshot shows the search interface for the FnhdC (Das Bonner Frühneuhochdeutschkorpus). The browser address bar shows <http://korpora.org/Fnhd/>. The page title is "FnhdC Das Bonner Frühneuhochdeutschkorpus".

The search form includes the following fields and options:

- Word Class (Wortklasse):** A dropdown menu with options: (egal), Potentiell, Substantiv (selected), Verb, Adjektiv. Buttons: "Formular an Wortklasse anpassen", "Suchen".
- Word Form (Wortform):** An empty text input field. Checkboxes: "Wortform als Muster angeben" (checked), "Wortform-Feld anpassen (Muster/Liste)". Button: "Suchen".
- Lemma:** A text input field containing "erde". Checkboxes: "Lemma als Muster angeben" (checked), "Lemma-Feld anpassen (Muster/Liste)". Button: "Suchen".
- Numerus:** A dropdown menu with options: (egal), singular (selected), plural, unbekannt.
- Kasus:** A dropdown menu with options: (egal), nominativ, genitiv, dativ (selected), akkusativ.
- Gender:** A dropdown menu with options: (egal), femininum.

Warning messages are displayed in green boxes:

- "Wenn Sie eine Liste der 70449 Wortform-Typen sehen wollen, kann es lange dauern, bis die Seite geladen und dargestellt ist."
- "Wenn Sie eine Liste der 8238 Lemma-Typen sehen wollen, kann es lange dauern, bis die Seite geladen und dargestellt ist."

Abbildung 2: Suchanfrage nach den Flexionsformen von Erde im Dativ Singular in FnhdC/S.

The screenshot shows a web browser window with the URL <http://korpora.org/Fnhd/>. The page title is "FnhdC Das Bonner Frühneuhochdeutschkorpus". Below the title, there is a search bar and a list of search results for the word "erden".

Search results:

- zur Bedeutung von Hervorhebungen s. ([fnhd_c.html_Auszeichnungen.nimj](#)).
- Gefunden in [Sigmund Herberstein: „Moscouia, Wien 1557“ \(Gegend: Mittelbairisch \(Wien\), Zeitstufe: 1550–1600\)](#):
[Blatt 3 Lage B Position B] ³⁷ muessen/ So hat er sich mit seinem Satl vnder seinem haubt auf der **erden** ligund
- Gefunden in [„DEO GRATIAS“ \(Gegend: Mittelbairisch \(Wien\), Zeitstufe: 1650–1700\)](#):
[Seite 29 Teil Deo Gratias] ⁰⁶ mit ganz bestürtztem Herzen sich zur **erden**
- Gefunden in [„Buch Altväter, Stuttgart 14. Jahrhundert“ \(Gegend: Schwäbisch, Zeitstufe: 1350–1400\)](#):
[Blatt 74] ⁰³ da er an der **erd** sinu knu zaichen vand da knüwet
- Gefunden in [„Buch Altväter, Stuttgart 14. Jahrhundert“ \(Gegend: Schwäbisch, Zeitstufe: 1350–1400\)](#):
[Blatt 91] ²³ an der **erd** vnd als got wolt do wir
- Gefunden in [„Buch Altväter, Stuttgart 14. Jahrhundert“ \(Gegend: Schwäbisch, Zeitstufe: 1350–1400\)](#):
[Blatt 91] ¹⁰ näch gerennet vnd spurten an der **erd** das

On the right side, a tooltip for the word "erden" is visible:

erden
Lemma: erde
Vokal: e
Typ: substantiv
Kasus: dativ
Numerus: singular
Genus: femininum

Abbildung 3: Ergebnisdarstellung der Suchanfrage von Abbildung 2.

4. Nutzung in Forschung und Lehre

An einigen Nutzungsbeispielen der XML-Fassung des Bonner Frühneuhochdeutsch-Korpus in Forschung und Lehre lässt sich illustrieren, welcher Mehrwert sich bereits aus der wortbezogenen morphosyntaktischen Etikettierung des Korpus und seinen diachronen und diatopischen Auswahlprinzipien ergibt.

Das Bonner Frühneuhochdeutsch-Korpus stellt die Materialgrundlage für flexionsmorphologische Untersuchungen des Frühneuhochdeutschen dar, wie sie in oben genannten Bänden der von MOSER *et al.* (1970 ff.) herausgegebenen Grammatik des Frühneuhochdeutschen ihren Niederschlag fand. Die linguistische Etikettierung der Texte war genau auf diese Untersuchungszwecke abgestellt. Die Korpusannotation erweist sich allerdings auch in anderen Bereichen der Morphologie als hilfreich. So arbeiteten PRELL (1991), PRELL & SCHEBBEN-SCHMIDT (1996) und DOERFERT (1994) mit diesem Korpus im Kontext von Untersuchungen zur frühneuhochdeutschen Wortbildungsmorphologie.

Die XML-Annotation ist gegenüber über die Wortgrenze hinausgehenden syntaktischen oder semantischen Etikettierungen offen. Helmut Weiß und Anna Voldina (Goethe-Universität Frankfurt, IDS

Mannheim) untersuchen die diachrone Entwicklung von Null-Argumenten im Deutschen anhand mittelhochdeutscher und frühneuhochdeutscher Texte. So zeigt sich, dass frühere Sprachstufen des Deutschen offenbar ausgeprägtere Möglichkeiten der Pronominalellipse kannten als das Neuhochdeutsche. VOLODINA & ONEA (2012) zitieren das Beispiel aus dem Korpus-Text 113 *Die Denkwürdigkeiten der Helene Kottannerin*, das verdeutlicht, dass hier im mit *ob* eingeleiteten Nebensatz das Personalpronomen der ersten Person in Subjektposition nach einem ebensolchen Subjekt im Hauptsatz im Gegensatz zum Neuhochdeutschen nicht realisiert zu werden braucht:

- (1) ich solt auf das haws vnd solt versüehen, ob ø ír kran vnd ander ir klainat mocht hinab zu ír bringen [...]⁸

Zum Zweck dieser Untersuchung planen Weiß und Volodina, ein Korpus mittel- und frühneuhochdeutscher Texte, um die entsprechenden Markierungen von Pro-Formen und Nullrealisierungen von Argumenten zu ergänzen, darunter auch Texte des Frühneuhochdeutsch-Korpus. Eine Pilotstudie verwendet die bereits erwähnten *Denkwürdigkeiten der Helene Kottannerin* aus dem Bonner Teilkorpus.

Das elektronische Korpus in der gegenwärtigen Präsentationsform erweist sich auch in der universitären Lehre als überaus nutzbringend. Die Anzeige der Lemmata zumeist in der neuhochdeutschen Schreibung hilft Studierenden bei der Erschließung der Texte, wie sich in verschiedenen Seminaren zu Themen des Sprachwandels und der Sprachgeschichte an der Universität Duisburg-Essen gezeigt hat. Hier erweist sich insbesondere die Darstellungsform einer ‚Glossierung on-demand‘ als günstig. Im Gegensatz zum Angebot einer Übersetzung bleibt das Augenmerk auf den Originaltext gelenkt. Dass die Glossierung mithilfe des Mauszeigers für Einzelwörter angefordert werden muss und nicht ständig für alle Wortformen eingeblendet ist, stellt einen Anreiz dar, sich auf die historische Schreibung einzulassen (vgl. Abbildung 1 im vorherigen Abschnitt 3).⁹

⁸ KOTTANNERIN, Bonner Frühneuhochdeutschkorpus, Text 113, S. 13, Z. 13f., Markierung des Nullsubjekts mit „ø“ nach VOLODINA & ONEA (2012).

⁹ Die vereinheitlichte Lemmaschreibung wird auch im Forschungskontext genutzt: M. FISCHER (2003) beispielsweise greift in seinen Untersuchungen zum Gebrauch von „zweifeln“ im Frühneuhochdeutschen ausdrücklich darauf zurück.

Die wortformenbezogene Suche nach morphosyntaktischen Merkmalen wurde in verschiedenen Seminaren zum Sprachwandel an der Universität Duisburg-Essen illustrativ und explorativ eingesetzt. Dies bietet sich beispielsweise bei Kurseinheiten zum Flexionswandel an. So lässt sich die diachrone und diatopische Verteilung der Endungen *-(e)* und *-(en)* auf die Singular-Kasus von *Erde* leicht durch FnhdC/S-Abfragen für die einzelnen Kasus, wie in den in Abschnitt 3 gezeigten Abbildungen 2 und 3 für den Dativ-Singular gezeigt, erkunden. Das unterschiedlich schnelle Fortschreiten der Kasusnivellierung bei den einzelnen Kasus wird unmittelbar veranschaulicht.

Auch für die didaktisch motivierte Erkundung von Phänomenen des Syntaxwandels hat sich FnhdC/S schon als nützlich erwiesen, obgleich weder nach Phrasen noch nach syntaktischen Funktionen gesucht werden kann. Für die tendenziell fortschreitende Rechtsverschiebung der rechten Satzklammer im vom Korpus abgedeckten Zeitraum findet man allein durch die Suche nach Infinitiven und Partizipien leicht eine Fülle an illustrierendem Material. Die folgende Abbildung 4 zeigt das Ergebnis der Suchabfrage nach Infinitiven und Partizipien.

Korpora.org: FnhdC — Das Bonner Frühneuhochdeutschkorpus

http://korpora.org/Fnhd/

FnhdC **Das Bonner Frühneuhochdeutschkorpus**

Korpora.org 2007

- Gefunden in [Wilhelm Durandus: „Rationale. Wien 1384“ \(Gegend: Mittelbairisch \(Wien\), Zeitstufe: 1350–1400\)](#):
²¹ halber vers, von dem **gesprochen** wirt am 6 tail vom sampcztag der
 [Seite 38]
- Gefunden in [Wilhelm Durandus: „Rationale. Wien 1384“ \(Gegend: Mittelbairisch \(Wien\), Zeitstufe: 1350–1400\)](#):
²⁴ ingang wirt **georden** zw dem werich und lob **Christi**, daz di **forerwelten**
 [Seite 38]
- Gefunden in [Wilhelm Durandus: „Rationale. Wien 1384“ \(Gegend: Mittelbairisch \(Wien\), Zeitstufe: 1350–1400\)](#):
²⁵ werden **gefordert** zw dem warn goczdinst. In ettleichen chirchen
 [Seite 38]
- Gefunden in [Wilhelm Durandus: „Rationale. Wien 1384“ \(Gegend: Mittelbairisch \(Wien\), Zeitstufe: 1350–1400\)](#):
²⁷ pabst, zw **bedewtten** ain grozzew frewd der zwchunft **Christi**. Sunder
 [Seite 38]
- Gefunden in [Wilhelm Durandus: „Rationale. Wien 1384“ \(Gegend: Mittelbairisch \(Wien\), Zeitstufe: 1350–1400\)](#):
²⁹ **gesungen** wirt vor dem ingang czw der mess, recht sam ain vorspil
 [Seite 38]

gesprochen
 Lemma: sprechen
 Typ: verb
 Klasse: stark_4a
 Tempus: praeteritum
 Form: partizip

Abbildung 4: Ergebnis der Suchabfrage nach Infinitiven und Partizipien in Text 111 (Wilhelm Durandus: *Rationale*, Wien 1384).

5. Ausblick: Das Referenzkorpus ‚Frühneuhochdeutsch‘

Im Rahmen des umfassenden Historischen Referenzkorpus des Deutschen (früher DDD = Deutsch Diachron Digital) wurde das Bonner Frühneuhochdeutsch-Korpus als vorhandenes, teillemmatisiertes und teilannotiertes Textarchiv erneut interessant. Die Planung des Historischen Referenzkorpus – eigentlich ein Archiv, das als Korpus genutzt werden kann¹⁰ – sieht ein Textarchiv vor, das von den Anfängen der deutschsprachigen Überlieferung bis ca. 1800 reicht. Dabei werden die überlieferten (und erhaltenen) Handschriften bis 1200 komplett erfasst,¹¹ für die Zeiträume danach wird jeweils eine strukturierte Textauswahl für das Archiv aufbereitet.

Für die Erarbeitung des Frühneuhochdeutsch-Referenzkorpus spielt das Bonner Frühneuhochdeutsch-Korpus eine zentrale Rolle. 31 der 40 Texte¹² der zweiten Hälfte des 14. Jahrhunderts bis 1600 – erweitert um die (nicht-digitalisierte) Auswahl 1600-1650 – werden aus dem Bestand übernommen, korrigiert und an die Standards der Referenzkorpora angepasst. Die Textausschnitte werden von 12.000 Wortformen auf max. 20.000 Wortformen erweitert. Die 14 Texte, die bisher nur nach der jeweiligen Edition eingelesen sind, sollen zusätzlich nach der Handschrift beziehungsweise dem Druck bearbeitet werden.

Die Annotation wird umfangreicher sein als dies bisher der Fall ist; es werden *alle* Wortformen annotiert. Die Annotation wird neben

¹⁰ Vgl. WEGERA (2013) zur Unterscheidung von Archiv und Korpus.

¹¹ Dies ist mit den derzeitigen technischen Mitteln innerhalb eines finanzierten Zeitrahmens möglich. Die Überlieferung des Althochdeutschen (und Altsächsischen) wird in Zusammenarbeit der Humboldt-Universität Berlin (Karin Donhauser, Anke Lüdeling), der Goethe-Universität Frankfurt (Jost Gippert) und der Friedrich-Schiller-Universität Jena (Rosemarie Lühr) erarbeitet; das Mittelhochdeutsche an der Ruhr-Universität Bochum (Klaus-Peter Wegera, Stefanie Dipper) und der Friedrich-Wilhelms-Universität Bonn (Thomas Klein, Claudia Wich-Reif), das Frühneuhochdeutsche ebenfalls an der Ruhr-Universität Bochum (Klaus-Peter Wegera, Stefanie Dipper), an der Martin-Luther-Universität Halle (Hans-Joachim Solms) und der Universität Potsdam (Ulrike Demske). Geplant sind das Mittelniederdeutsche an der Westfälischen Wilhelms-Universität Münster (Robert Peters) und der Universität Hamburg (Ingrid Schröder) und das Frühe Moderne Deutsch (1650-1800) in Zusammenarbeit deutscher Universitäten (Bochum, Halle, Potsdam) mit der University of Manchester (Martin Durrell) und der Universität i Oslo (Kirsten Bech). Zu diesem Zweck wird das Manchester-Korpus (GerManC; vgl. DURRELL *et al.*, 2007) an die Standards der übrigen Referenzkorpora angepasst und erweitert.

¹² Die Texte aus dem Zeitraum 1650-1700 werden evtl. im Rahmen der Überarbeitung des GerManC Teil des Referenzkorpus Frühes Modernes Deutsch (1650-1800) berücksichtigt.

den Wortarten (POS) im Falle der flektierten Wortarten jeweils eine Angabe der Flexionsklasse und Angaben zu dem flexionsmorphologischen Merkmalen Kasus, Numerus und Genus (Substantive, Adjektive und Pronomina), Person, Numerus, Modus und Tempus (Verben) beinhalten. Eine syntaktische Annotation erfolgt für eine kleinere, noch nicht festgelegte, Auswahl von Texten des Korpus.¹³ Für das Referenzkorpus wurde das STTS für Erfordernisse der Annotation historischer Texte modifiziert (STTS(H)). Mit diesem Tagset annotierte Texte werden in das XML-Standoff-Format PAULA (Potsdamer Austauschformat für linguistische Annotationen)¹⁴ überführt und über die linguistische Datenbank ANNIS (Annotation von Informationsstruktur)¹⁵ verfügbar gemacht.

Es stellt sich die Frage, ob und in welcher Form das Bonner Frühneuhochdeutsch-Korpus neben dem Referenzkorpus weiter gepflegt werden soll. Da es sich um einen gut durchdachten strukturierten Bestand handelt, ist eine weiterhin separate Nutzung sicherlich möglich. Als Form wäre allerdings die erweiterte Fassung wünschenswert. Diese wäre dann aber auch über ANNIS möglich, da auf strukturierte Teilkorpora als solche ein separater Zugriff möglich sein wird.

Es lässt sich vorhersehen, dass das Bonner Korpus für die Forschung und Lehre weitgehend vom Referenzkorpus ‚Frühneuhochdeutsch‘ abgelöst wird. Es bleibt bestehen als Beispiel für ein exemplarisches und hinsichtlich der Textauswahl hochwertiges Korpus, und möglicherweise besteht auch ein anhaltender Bedarf an der spezialisierten Oberflächen des Bonner Korpus (FnhdC/HTML und FnhdC/S). Bis zur Ablösung durch das Referenzkorpus fungiert das Bonner Korpus als Labor für die Untersuchung verschiedener Zugriffsmöglichkeiten und Analysen, zur Erhebung von Anforderungen für die eHumanities, als Prototyp zur Erprobung spezialisierter Suchmöglichkeiten und für den Einsatz in der Lehre.

¹³ Zur Methode vgl. DEMSKE (1996).

¹⁴ <http://www.sfb632.uni-potsdam.de/en/paula-en.html>

¹⁵ <http://www.sfb632.uni-potsdam.de/annis/>; vgl. DIPPER (2005), CHIARCOS *et al.* (2008).

LITERATUR

- BERG, E., 1982: Entwicklung eines Kodierungssystems am Beispiel frühneuhochdeutscher Texte, in: LENDERS, W. & K.-P. WEGERA (Hrsg.), *Maschinelle Auswertung sprachhistorischer Quellen*, Sprache und Information 3, Tübingen, 19-50.
- CHIARCOS, C. *et al.*, 2008: A flexible framework for integrating annotations from different tools and tagsets, in: *Traitement Automatique des Langues* 49 (2), 217-246.
- DAMMERS, U. *et al.*, 1988: *Flexion der starken und schwachen Verben*, Band IV der Grammatik des Frühneuhochdeutschen, hrsg. von MOSER, H. *et al.*, Heidelberg.
- DEMSKE, U., 1996: Bestandsaufnahme zum Untersuchungsbereich ‚Syntax‘, in: FRITZ, G. & E. STRASSNER (Hrsg.), *Die Sprache der ersten deutschen Wochenzeitungen im 17. Jahrhundert*, Tübingen, 70-125.
- DIEL, M. *et al.*, 2002: *XML-Kodierung des Bonner Frühneuhochdeutschkorpus*, IKP-Arbeitsbericht NF 02, Bonn.
- DIPPER, S., 2005: XML-based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation, in: *Proceedings of Berliner XML Tage 2005 (BXML 2005)*, Berlin, 39-50.
- DOERFERT, R., 1994: *Die Substantivableitung mit -heit, -keit, -ida, -î im Frühneuhochdeutschen*, Berlin.
- DURRELL, M. *et al.*, 2007: ‘GerManC: A historical corpus of German 1650-1800’, in: *Sprache und Datenverarbeitung* 31, 71-80.
- FISCHER, M., 2003: Ein Zweifelsfall: *zweifeln* im Deutschen, in: ECKARDT, R. (Hrsg.), *Questions and Focus*, ZAS Papers in Linguistics 30, Berlin.
- FISSENI, B. *et al.*, 2007: FnhdC/HTML und FnhdC/S, in: *Sprache und Datenverarbeitung* 31, 67-69.
- GRASER, H. & K.-P. WEGERA, 1978: Zur Erforschung der frühneuhochdeutschen Flexionsmorphologie, in: *Zeitschrift für Deutsche Philologie* 97, 74-91.
- HOFFMANN, W. & F. WETTER, 1985: *Bibliographie frühneuhochdeutscher Quellen. Ein kommentiertes Verzeichnis von Texten des 14.-17. Jahrhunderts (Bonner Korpus)*, Frankfurt am Main [u.a.].

- LENDERS, W. & K.-P. WEGERA (Hrsg.), 1982: *Maschinelle Auswertung sprachhistorischer Quellen. Ein Bericht zur computerunterstützten Analyse der Flexionsmorphologie des Frühneuhochdeutschen*, Tübingen.
- MOSER, H. *et al.* (Hrsg.), 1970 ff.: *Grammatik des Frühneuhochdeutschen. Beiträge zur Laut- und Formenlehre*, Heidelberg.
- PRELL, H.-P., 1991: *Die Ableitung von Verben aus Substantiven in biblischen und nichtbiblischen Texten des Frühneuhochdeutschen*, Frankfurt am Main.
- PRELL, H.-P. & M. SCHEBBEN-SCHMIDT, 1996: *Die Verbableitung im Frühneuhochdeutschen*, *Studia linguistica Germanica* 41, Berlin.
- SOLMS, H.-J. & K.-P. WEGERA, 1991: *Flexion der Adjektive*, Band VI der Grammatik des Frühneuhochdeutschen, hrsg. von MOSER, H. *et al.*, Heidelberg.
- SOLMS, H.-J. & K.-P. WEGERA, 1998: Das Bonner Frühneuhochdeutsch-Korpus. Rückblick und Perspektiven, in: BERGMANN, R. (Hrsg.), *Probleme der Textauswahl für ein elektronisches Thesaurus. Beiträge zum ersten Göttinger Arbeitsgespräch zur historischen deutschen Wortforschung 1. und 2. November 1996*, Stuttgart, Leipzig.
- VOLODINA, A. & E. ONEA, 2012: Am Anfang war die Lücke, in: BÄR, J. & M. MÜLLER (Hrsg.), *Geschichte der Sprache – Sprache der Geschichte. Probleme und Perspektiven der historischen Sprachwissenschaft des Deutschen*, Berlin, 207-237.
- WALCH, M., 1988: *Flexion der Pronomina und Numeralia*, Band V der Grammatik des Frühneuhochdeutschen, hrsg. von MOSER, H. *et al.*, Heidelberg.
- WEGERA, K.-P., 1987: *Flexion der Substantive*, Band III der Grammatik des Frühneuhochdeutschen, hrsg. von MOSER, H. *et al.*, Heidelberg.
- WEGERA, K.-P., 2013, im Druck: Language Data Exploitation. Design and Analysis of Historical Language Corpora, in: BENNETT, P. *et al.* (Hrsg.), *New Methods in Historical Corpus Linguistics*, *Corpus Linguistics and Interdisciplinary Perspectives on Language* 3, Tübingen.