

**Bochumer
Linguistische
Arbeitsberichte
24**



**Dokumentation
zum Referenzkorpus Frühneuhochdeutsch
und Referenzkorpus Deutsche Inschriften**

**Birgit Herbers, Sylwia Kösser, Ilka Lemke, Ulrich Wenner,
Juliane Berger, Sarah Kwekkeboom und Frauke Thielert**

Bochumer Linguistische Arbeitsberichte



Herausgeberin: Stefanie Dipper

Die online publizierte Reihe „Bochumer Linguistische Arbeitsberichte“ (BLA) gibt in unregelmäßigen Abständen Forschungsberichte, Abschluss- oder sonstige Arbeiten der Bochumer Linguistik heraus, die einfach und schnell der Öffentlichkeit zugänglich gemacht werden sollen. Sie können zu einem späteren Zeitpunkt an einem anderen Publikationsort erscheinen. Der thematische Schwerpunkt der Reihe liegt auf Arbeiten aus den Bereichen der Computerlinguistik, der allgemeinen und theoretischen Sprachwissenschaft und der Psycholinguistik.

The online publication series “Bochumer Linguistische Arbeitsberichte” (BLA) releases at irregular intervals research reports, theses, and various other academic works from the Bochum Linguistics Department, which are to be made easily and promptly available for the public. At a later stage, they can also be published by other publishing companies. The thematic focus of the series lies on works from the fields of computational linguistics, general and theoretical linguistics, and psycholinguistics.

© Das Copyright verbleibt beim Autor.

Band 24 (May 2021)

Herausgeberin: Stefanie Dipper
Sprachwissenschaftliches Institut
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum

Erscheinungsjahr 2021
ISSN **2190-0949**

**Birgit Herbers, Sylwia Kösser, Ilka Lemke, Ulrich Wenner,
Juliane Berger, Sarah Kwekkeboom und Frauke Thielert**

**Dokumentation
zum Referenzkorpus Frühneuhochdeutsch
und Referenzkorpus Deutsche Inschriften**

2021

Bochumer Linguistische Arbeitsberichte

(BLA 24)

Abstract

Dieser Band dokumentiert die Zusammenstellung und Aufbereitung des Referenzkorpus Frühneuhochdeutsch und des Referenzkorpus Deutsche Inschriften. Erläutert werden die Korpusstruktur sowie die Transkriptions- und Annotationskonventionen.

This volume documents the compilation and processing of the Reference Corpus Early New High German and the Reference Corpus German Inscriptions. The corpus design as well as the transcription and annotation conventions are explained.

Dokumentation zum

**Referenzkorpus Frühneuhochdeutsch und
Referenzkorpus Deutsche Inschriften**

Birgit Herbers, Sylwia Kösser, Ilka Lemke, Ulrich Wenner,
Sarah Kwekkeboom, Juliane Berger und Frauke Thielert

Stand: März 2021

Inhaltsverzeichnis

0.	Vorbemerkungen.....	1
I.	Projekt.....	2
I.1	Korpusstruktur.....	2
I.2	Das Bonner Frühneuhochdeutsch-Korpus	3
I.3	Referenzkorpus Deutsche Inschriften (ReDI).....	4
II.	Transkription.....	4
II.1	Organisatorisches.....	5
II.1.1	Header	5
II.1.2	Ordnungsteil und Zählung.....	6
II.2	Kodierung von Abkürzungen und Superskripten/Diakritika.....	6
II.2.1	Superskribierte Buchstaben.....	6
II.2.2	Superskripte über <y> und <i>.....	7
II.2.3	Subskribierte Buchstaben	7
II.2.4	Kürzung von Prä- und Suffixen/Endungen sowie einzelnen Buchstaben und Buchstabenverbindungen	8
II.2.5	Anders modifizierte Buchstaben, besondere Zeichen, Ligaturen, Eingabekonventionen und Sonstiges	9
II.2.6	Spezielle lateinische Kürzungen in Halle.....	10
II.2.7	Initialen und Majuskeln.....	12
II.2.8	Interpunktion in der Handschrift/im Druck.....	13
II.2.9	Klammern.....	14
II.2.10	Bildbeischriften.....	17
II.2.11	Informationen zum Text außerhalb des Haupttextes	18
II.3	Zur Nachbereitung (Sekundärzählung).....	20
III.	Präedition	20
III.1	Wichtige Präeditionszeichen	21
III.2	Neuhochdeutsche Interpunktion	21
III.3	Fremdsprachliches Material.....	22
III.4	Verschmelzungsformen	22
III.5	Partikelverben.....	22
III.5.1	Markierung von Partikelverben.....	22
III.5.2	Komposita aus Adverb und Partizip.....	23
III.6	Pronominaladverbien in Distanzstellung.....	23
III.7	Negationspartikel en-.....	23
IV.	Annotation.....	23
IV.1	Generelles.....	23
IV.1.1	Lemmaansatz.....	24
IV.1.2	Fremdsprachliche Wörter	25
IV.1.3	Abkürzungen und Abkürzungen	25

IV.2	Tagset (Abweichungen/Ergänzungen zu HiTS)	25
IV.2.1	Entfallene Tags.....	26
IV.2.2	Einzelne/Neue Tags.....	26
IV.3	Annotation nach lexikalischen Kategorien (alphabetisch sortiert)	26
IV.3.1	Adjektive	26
IV.3.2	Adverbien.....	28
IV.3.3	Eigennamen.....	29
IV.3.4	Fragewörter	30
IV.3.5	Junktionen	30
IV.3.6	Artikelwörter und Pronomina.....	31
IV.3.7	Substantive	35
IV.3.8	Verben.....	37
IV.4	Verschmelzungsformen aus Präposition und Artikel	43
IV.5	Zahlen	43
IV.6	Zirkumpositionen	43
V.	Literatur	44
VI.	Anhang	46
VI.1	Eigennamen	46
VI.1.1	Vor- und Nachnamen	46
VI.1.2	Fremdsprachliche Eigennamen	46
VI.1.3	Nomina sacra (Heiligennamen).....	47
VI.2	Heiligtage	47
VI.3	Ortsnamen	47
VI.4	Annotation ‚Jahr‘ in Alters-/Zeitangaben in ReDI	48
VI.5	Header-Beispiele	49
VI.5.1	Referenzkorpus Frühneuhochdeutsch	49
VI.5.2	Referenzkorpus Deutsche Inschriften.....	50
VI.6	Tagset	52

0. Vorbemerkungen

Diese Dokumentation erfasst die Entscheidungen und Konventionen, die bei der Erstellung des Korpus befolgt wurden und sich in den Rohdaten widerspiegeln. Für die Nutzung in ANNIS wurden die Daten konvertiert, so dass beispielsweise statt '\$' für das Schaft-S in ANNIS in der Ebene "tok_dipl" (diplomatisches Token) das Zeichen 'ſ' steht und auf der Ebene "tok_anno" (annotiertes Token) das normalisierte Zeichen 's' verwendet wird. Die Informationen aus den Rohdaten (inkl. der originalen Transkriptionen sowie der Kommentare zu den Transkriptionen) sind in der XML-Version des Korpus enthalten (im Format CorA-XML).

Lizenz und Zitierweise

Das Referenzkorpus Frühneuhochdeutsch ist lizenziert unter einer Creative Commons Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.

Wenn Sie das Korpus zitieren möchten, bitte das folgende Format verwenden:

Wegera, Klaus-Peter; Solms, Hans-Joachim; Demske, Ulrike; Dipper, Stefanie (2021). Referenzkorpus Frühneuhochdeutsch (1350–1650), Version 1.0, <https://www.linguistics.ruhr-uni-bochum.de/ref/>. ISLRN: 918-968-828-554-7.

I. Projekt

Ziel des Projektes war die Erstellung eines Referenzkorpus für die frühneuhochdeutsche (frnhd.) Sprachstufe im Rahmen des ‚Korpus historischer Texte des Deutschen‘ (ehemals ‚Deutsch Diachron Digital‘), in dem hochdeutsche Sprachdenkmäler des spätmittelalterlichen und frühneuzeitlichen Zeitraums von ca. 1350-1650 in strukturierter Auswahl zusammengefasst, digital verfügbar transkribiert, lemmatisiert sowie grammatisch annotiert wurden; die Auswahl ist durch die Kategorien Raum, Zeit, Textbereich und Überlieferungsform bestimmt. Ziel war also eine hinlänglich umfangreiche, verlässliche und handschriften-/druckausgabengetreue Datenbasis des Frnhd., die Recherchen zur Historiolinguistik in einem Maße erlaubt, das weit über das bisher Mögliche hinausgeht; das Korpus ist zudem als Arbeitsinstrument für die mediävistische und Frühneuzeitforschung nutzbar. Das Projekt schließt an die bereits laufenden Referenzkorpusprojekte ‚Altdeutsch‘ und ‚Mittelhochdeutsch‘ an und nutzt alle dort entwickelten Methoden und Werkzeuge zur Digitalisierung (digital verfügbare Transkription/Annotation).

Die Transkription und morphologische Annotation der Texte erfolgten in den Arbeitsstellen in Bochum und Halle (im Folgetext kurz mit Bochum und Halle benannt), die syntaktische Annotation erfolgte in der Arbeitestelle Potsdam.

Alle Texte wurden je nach Länge in einem Umfang von bis zu 20.000 Wortformen transkribiert und gemäß einem den anderen Referenzkorpora gemeinsamen Tagset HiTS¹ z.T. manuell und z.T. automatisch annotiert. Die Annotation erfolgte über die online-Annotationstools CorA² und LAKomp³, die Daten wurden nach der Annotation in CorA-XML umgewandelt und sind über die linguistische Datenbank ANNIS⁴ online und frei verfügbar.

Anknüpfend an die bis 1350 reichende differenzierte Struktur des Referenzkorpus ‚Mittelhochdeutsch‘ wurde ein Korpus des Frnhd. gestaltet, das in seiner Struktur die dynamische Entwicklung der Periode hinsichtlich ihrer differenzierbaren diachronischen, diatopischen und medial-/überlieferungsformen-/textbereichsbezogenen Varianz (Hs./Dr., Vers/Prosa, Textsorte) abbildet.

Zum Spezialkorpus ReDI (Referenzkorpus Deutsche Inschriften) s. I.3.

I.1 Korpusstruktur

Das Korpus zeigt die dynamischen sprachgeschichtlichen Entwicklungen im Frnhd. auf. Das zugrunde liegende Korpusgerüst besteht aus verschiedenen mit Texten gefüllten Rasterfeldern.

Diese Rasterfelder decken den Zeitraum 1350-1650 ab, die ‚linke Achse‘ der Rasterfelder besteht also aus sechs zeitlich aufeinanderfolgenden Intervallen, die im Abstand von 50 Jahren angesetzt werden.

¹ Dipper u.a. (2013).

² Bollmann u.a. (2014).

³ Medek (*Gießler) u.a. (2015).

⁴ s. <https://linguistics.rub.de/annis/annis3/REF/>.

Zeitraum I	1350-1400	214
Zeitraum II	1400-1450	115
Zeitraum III	1450-1500	215
Zeitraum IV	1500-1550	116
Zeitraum V	1550-1600	216
Zeitraum VI	1600-1650	117

Die ‚rechte Achse‘ zeigt die Sprachräume an. Insgesamt wurden sieben sprachliche Großräume bearbeitet, die in Abhängigkeit von der jeweiligen Zeitspanne und der notwendigen Berücksichtigung u.a. kleinräumlicher Schriftdialekte ihrerseits weitergehend in maximal 18 dialektal orientierte Sprachlandschaften differenziert werden:

- A Ostoberdeutsch (sbair., mbair., nbair.)
- B Westoberdeutsch (oschwäb., schwäb., hchal., oberrhein.)
- C Westmitteledeutsch (südrhfrk., hess., msfrk., rip.)
- D Ostmitteledeutsch (schles., obs., thür.)
- E Ostfränkisch
- F Böhmisch-Mährisch
- G Norddeutsch (norddt., preuß.)⁵

Bis in die 2. Hälfte des 15. Jh.s (also bis Zeitraum III) wurde versucht, abhängig von der Überlieferungslage in allen Korpusfeldern jeweils einen Prosa- und einen Verstext zu berücksichtigen. Für die 1. Hälfte des 16. Jh.s ist diese Unterscheidung aufgegeben worden.

In Zeitraum I und Zeitraum II wurden nur Handschriften, in Zeitraum III und Zeitraum IV sowohl Handschriften als auch Drucke verwendet.

Innerhalb des Korpus wurden die verschiedenen und für das Frnhd. charakteristischen Textsorten berücksichtigt. Die sehr breite Forschungsdiskussion zeigt, dass eine für jeden Text anwendbare, eindeutige und konsensuelle Textsortenzuweisung nach wie vor problematisch ist. Die Texte wurden in einem induktiv-pragmatischen Verfahren folgenden sechs Textbereichen zugewiesen:

- RG Rechts- und Geschäftstexte
- CB chronikalische und Berichtstexte
- RE Realientexte
- UN unterhaltende Texte
- KT kirchlich-theologische Texte/Bibeln
- EB erbauliche Texte

I.2 Das Bonner Frühneuhochdeutsch-Korpus

Das Bonner Frühneuhochdeutsch-Korpus entstand zwischen 1972 und 1985 an der Bonner Forschungsstelle ‚Frühneuhochdeutsch‘ im Rahmen des Projekts ‚Flexionsmorphologie des Frühneuhochdeutschen‘ unter Leitung von Werner Besch, Winfried Lenders (ab 1976), Hugo

⁵ F Böhmisches-Mährisch und G Norddeutsch (norddt., preuß.): keine weitere Bearbeitung.

Moser und Hugo Stopp (bis 1981). Es diene als Materialgrundlage für die Erarbeitung der Bände 3, 4 und 6 der Grammatik des Frühneuhochdeutschen⁶.

Das Korpus besteht aus 40 Quellen, die nach Sprachlandschaften und Zeitschnitten (1350-1400, 1450-1500, 1550-1600 und 1650-1700) angeordnet sind. Es handelt sich um Auswahltexte mit einem Umfang von jeweils ca. 30 Normalseiten, d.h. ca. 12.000 Wortformen. Die annotierten Texte sind online einseh- und abrufbar.⁷

Diese Texte wurden in das neue Referenzkorpus integriert und dabei kollationiert und jeweils um ca. 8.000 Wortformen ergänzt.

I.3 Referenzkorpus Deutsche Inschriften (ReDI)

Das Referenzkorpus Deutsche Inschriften (ReDI)⁸ stellt ein besonderes Korpus dar, weil es mit den Inschriften eine Textsorte/Überlieferungsform bietet, die in den anderen Referenzkorpora nicht systematisch berücksichtigt wurde.

In ReDI wurden alle deutschsprachigen (sowie gemischt-sprachliche, bestehend aus deutschen und fremdsprachlichen Wörtern), nichtkopialen Inschriften, die in der Editionsreihe *Die Deutschen Inschriften* ediert sind, bis 1650 erfasst und vollständig grammatisch annotiert. Dem Projekt standen zur Projektlaufzeit (2014-2016) 64 Inschriftenbände (der mittlerweile rund 100 Bände der Inschriftenreihe *Die Deutschen Inschriften*) für die Bearbeitung zur Verfügung. Der weitaus größte Teil der Inschriften stammt aus spätmittelhochdeutscher und frühneuhochdeutscher Zeit, es gibt nur wenige frühere mittelhochdeutsche Inschriften. Das Korpus enthält sowohl hochdeutsche als auch niederdeutsche sowie hochdeutsch-niederdeutsch gemischte Inschriften und stellt somit ein Bindeglied zwischen ReF (Referenzkorpus Frühneuhochdeutsch) und ReN (Referenzkorpus Niederdeutsch) dar.

II. Transkription

Der folgende Abschnitt beschäftigt sich mit der digitalen Erfassung, also der Transkription der Korpustexte, und weist auf Besonderheiten in der Kodierung heute nicht mehr gängiger Sonderzeichen hin. Sämtliche Abbildungen stammen aus den Handschriften und Drucken der Korpustexte oder aus Capelli.⁹

Besonderheiten in ReDI: Von den 64 in ReDI bearbeiteten Inschriftenbänden wurden 16 Bände nach den gleichen Richtlinien wie die Texte in ReF manuell transkribiert (Deutsche Inschriften Bd. 1-16). Die anderen 48 Bände wurden von ‚Deutsche Inschriften Online‘ (Interakademisches Projekt der Akademien zu Göttingen und Mainz mit dem Ziel der Digitalisierung und Online-Bereitstellung der Inschriftenbände der gesamten DI-Reihe; s. www.inschriften.net) als Datensätze zur Verfügung gestellt. Diese Datensätze wurden an die

⁶ Gr. d. Frnhd.

⁷ s. <http://korpora.zim.uni-due.de/Fnhd/>.

⁸ Herbers (2016).

⁹ Capelli (1928).

Transkriptionsstandards der Referenzkorpora angepasst (bspw. projekteigene Kodierungen von Symbolen und Sonderzeichen und Präedition).

II.1 Organisatorisches

II.1.1 Header

Jeder ReF-Text erhält einen Header (s. Anhang; VI.5), der grundlegende Informationen zu dem jeweiligen Text enthält. Die Headerdaten dienen in ANNIS der Auffindung gewünschter Textstellen nach unterschiedlichen Suchvorgaben.

Die wichtigsten Metainformationen sind:

Korpussigle: projektinterne Sigle

Titel: Kurztitel des Werkes

Verfasser: soweit bekannt

Textart: Prosa/Vers (Druck oder Handschrift)

Textsorte: Textbereich (s. I.1)

Bibliothek/Archiv

Signatur: Signatur der lagernden Bibliothek/des lagernden Archivs

Auswahl Bonner Frnhd. Korpus: Angabe der Textstellen aus dem Bonner Frnhd. Korpus, sofern vorhanden

Auswahl Referenzkorpus: Angabe der ausgewählten Textstellen

Umfang in Wortformen: Umfang in Wortformen (Bonner Frnhd. Korpus, ReF und gesamt)

Datierung/ Lokalisierung/ Druck- oder Schreibort/ Drucker oder Schreiber

verwendete Edition: sofern vorhanden

Literatur: Sekundärliteratur, sofern vorhanden

Sonstiges: Anmerkungen zur Transkription (z.B. Schreiber- oder Druckerspezifika), Lemmatisierung, Annotation des spezifischen Textes

In ReDI gibt es grundsätzlich zwei Header: einen zum Gesamtband und jeweils einen zu jeder Inschrift. Aus dem Header für den Gesamtband geht jeweils hervor, welche Edition zugrunde liegt, welche Besonderheiten dieser Inschriftenband bezüglich einzelner Inschriften aufweist, welche Symbole in welcher Weise kodiert wurden und wer die einzelnen Bearbeitungsstufen vorgenommen hat.

Die Header für die einzelnen Inschriften enthalten detaillierte Informationen zur Inschrift. Die ID besteht aus der Nr. des jeweiligen Editionsbandes der Reihe *Deutsche Inschriften* sowie der jeweiligen Katalognummer (= Inschriftennummer). Desweiteren werden der Standort und der Überlieferungsträger (bspw. Grabstein, Holzkreuz, Hauswand) jeder Inschrift angegeben, die Sprache wird nach ‚hochdeutsch‘, ‚niederdeutsch‘ oder ‚hochdeutsch-niederdeutsch gemischt‘ unterschieden. Die Kategorien ‚Textart‘ (Prosa oder Vers) und ‚Textbereich‘ (Alltag, Poesie, Recht, Religion, Wissenschaft) wurden an ReF angepasst. Wichtig für ReDI sind die Angaben zur Abbildung, da der Großteil der Bände nach der Edition aufgenommen

wurde. Bei Vorliegen einer Abbildung wurde der von der Edition gegebene Text mit der Abbildung kollationiert und ggf. abgeändert. Wenn möglich, werden Links zu den Abbildungen aufgeführt.

II.1.2 Ordnungsteil und Zählung

Alle Texte wurden nach der Originalzählung des Textes eingegeben. Jede Textzeile beginnt mit dem Ordnungsteil; dieser besteht aus der F-Sigle (korpusinterne Ordnungsnummer), Bindestrich, Folio-/Seitenangabe, Komma, Zeile. Seiten und Zeilenangaben wurden hierbei mit Nullen aufgefüllt. Nach dem Ordnungsteil wird jeweils ein Tab gesetzt, dann beginnt der Text.

Bsp.: F108-036r,01

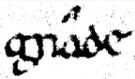
Es kommt v.a. bei paginierten Drucken vor, dass Vorrede, Index usw. auch eine Zählung erhalten und der eigentliche Text wieder bei Seite 1 beginnt. In diesem Falle wurden die Seiten der Vorrede in einer Weise bezeichnet, die nicht im eigentlichen Text Verwendung findet.

II.2 Kodierung von Abbrüviaturen und Superskripten/Diakritika

In diesem Abschnitt wird die Kodierung von Abbrüviaturen, Superskripten und Diakritika erläutert. Textspezifische Abkürzungen und Sonderzeichen wurden mit dem variablen Zeichen `&1` bis `&n` kodiert, im Header beschrieben und in ihrer Funktion bestimmt. Gibt es in einer Handschrift mehrere außergewöhnliche Abbrüviaturen, wurden entsprechend mehrere Kürzungszeichen definiert und vermerkt. Diese variablen Zeichen stehen in der Regel für Abbrüviaturen, Kürzungen usw., für die keine Kodierung vereinbart ist. Für die Hallenser Texte wurden zum Teil ergänzend spezifische Konventionen vereinbart, die in II.2.6 gesondert gekennzeichnet werden.

II.2.1 Superskribierte Buchstaben

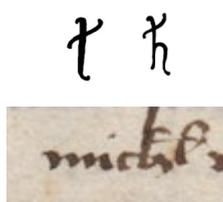
Abbildung	Kodierung	Beschreibung
	<code>%x</code>	Hochstellung eines Buchstabens <code><x></code> ; wird (wenn vertretbar) an der Stelle im Wort eingefügt, an die er sinnvollerweise gehört, auch wenn er leicht nach rechts oder links versetzt ist. Bsp.: <code>wi\$%heit</code>
	<code>%%x</code>	Kürzung von <code><r></code> durch Superskription des Folgevokals. Bsp.: <code>\$p%%ach = fp^ach = sprach</code> Bsp.: <code>g%%oaz = g^oz = groz</code>
	<code>y'</code>	<code><y></code> mit Superskript (egal ob Punkt, Akzent, Trema etc.)
	<code>\-</code>	Nasalstrich (in verschiedenen Ausprägungen)

	\x	Buchstabe als Superskript Zu beachten ist, dass nur Vokalmodifikatoren als Superskripte transkribiert werden – etwa <i\e>, <u\o>. Alle anderen superskribierten Buchstaben werden als hochgestellt aufgefasst und mit % kodiert.
	â	Diakritikon über Vokal Bsp.: <i>gnâde</i>
	x\^, x\' , x\`	Diakritikon über Konsonant (Zirkumflex, Akut, Gravis)
	x\:	Doppelpunkt (Trema) über Zeichen; <ä>, <ö> und <ü> können direkt transkribiert werden.
	x*	Kodierung für alle Diakritika ohne klare Funktion und unspezifizierte Zeichen für Vokalmarkierung, z.B. ein nach links offenes, übergeschriebenes <o>
	p\~	Nasalstrich für <i>pre</i> , <i>prae</i>

II.2.2 Superskripte über <y> und <i>

Superskripte über <y> und <i> werden, sofern in der Transkription berücksichtigt, im Header beschrieben.

II.2.3 Subskribierte Buchstaben

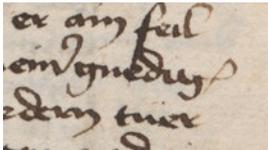
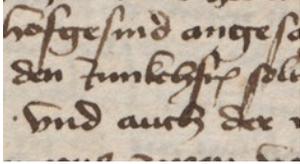
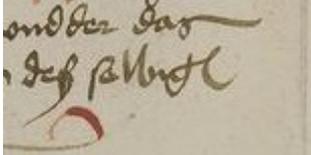
Abbildung	Kodierung	Beschreibung
	e\,	für <i>ε</i> (e-caudata)
	∨	Schrägstrich durch die oder Modifikation der Oberlänge. Häufig als kontraktive Kürzung oder allgemeines Abbrechungszeichen verwendet und bei Nomina sacra. Bsp.: Joh∨ = <i>Johannes</i> Bsp.: IV = <i>lat. vel</i> Bsp.: ap∨s = <i>apostolus</i>
	∟	Schrägstrich durch die oder Modifikation der Unterlänge. Häufig als kontraktive Kürzung oder allgemeines Abbrechungszeichen verwendet, vor allem bei <q> und <p> (q\∟ = <i>que</i> , <i>qu</i> , <i>qui</i> oder <i>quod</i>) und p\∟ = (<i>pro</i> , <i>per</i> , selten <i>pre/prae</i>)

II.2.4 Kürzung von Prä- und Suffixen/Endungen sowie einzelnen Buchstaben und Buchstabenverbindungen

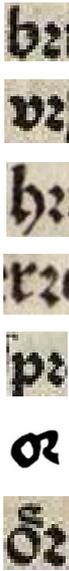
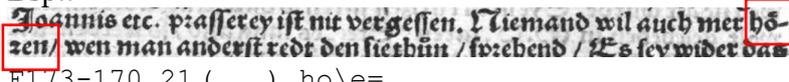
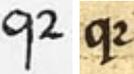
Abbildung	Kodierung	Beschreibung
	'	Kürzung von <er>, <r>, <ir>, <re>, <ri>. Unabhängig von der graphischen Realisation in der Handschrift als <^>, <s>, <z> realisiert. Das Zeichen <^> wird auch dann verwendet, wenn es in der Handschrift übergeschrieben ist. Es wird dort eingefügt, wohin es sinnvollerweise gehört.
	%2	Kürzung von <ur> durch ein Zeichen ähnlich wie <~> oder <2>
	%9	Kürzung von v.a. auslautend <us> durch <9>
	9	Kürzung von anlautend <con>, <com> Anmerkung: Im Unterschied zu %9 steht 9 am Wortanfang, Bsp.: 9munio = <i>communio</i> Bsp.: 9t%%a = <i>contra</i>
	3	Kürzung von auslautend <et> durch <3> Bsp.: gewill3 = <i>gewillet</i>
	4	Kürzung von auslautend <rum> Bsp.: tuo4 = <i>tuorum</i>
	_3	Kürzung von v.a. auslautend <us>, hauptsächlich in der lat. Endung -ibus Bsp.: incordib_3 = <i>incordibus</i>

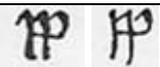
Konventionen spezifisch für Halle

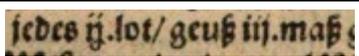
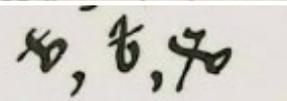
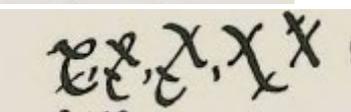
Abbildung	Kodierung	Beschreibung
	\&1	meist für ein e vor Nasal Bsp.: vngri\$chn&1 Bsp.: fragn&1
	\&2	meist für auslautend -en/-em gnad&2 = <i>gnad(en)</i> auch bei latein. Wörtern mit dieser Kürzung verwenden; s. II.2.6 Lateinische Kürzungen

 	&3	kann für verschiedene Endungen stehen; meist jedoch für auslautend <i>-en</i> Bsp. Bild 1: gnedig&3 = <i>gnedig(en)</i> Bsp. Bild 4: Junkhfr&3 = <i>Junkhfr(au)</i> auch bei latein. Wörtern mit dieser Kürzung verwenden; s. II.2.6 Lateinische Kürzungen
	&4	meist für auslautend <i>-en</i> <i>Selbig&4 = Selbig(en)</i>

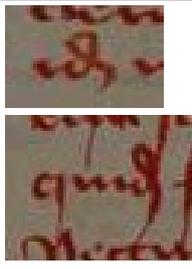
II.2.5 Anders modifizierte Buchstaben, besondere Zeichen, Ligaturen, Eingabekventionen und Sonstiges

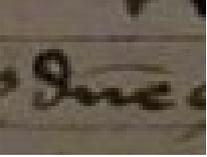
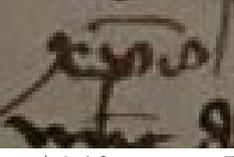
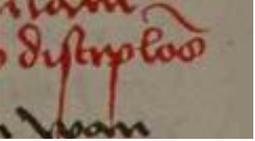
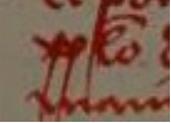
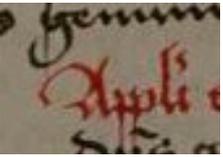
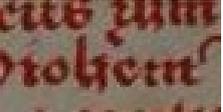
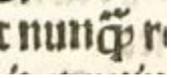
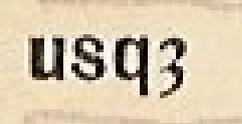
Abbildung	Kodierung	Beschreibung
	\$	Schaft-S im Gegensatz zu Rund-S Bsp.: <i>Selbe</i>
	d_e	Ligatur aus <d> und an der Oberlänge angebundenem <e>
	a_e	Ligatur <a_e>
	o_e	Ligatur <o_e>
	b_r v_r h_r r_r p_r o_r o_r_e	Ligatur aus unterschiedlichen Buchstaben und einem runden <r> Entsprechend wird bei allen (auch den nicht hier aufgeführten) r-Ligaturen verfahren: Bsp.: <i>ho_rte</i> beachte: Ligatur vor Superskript! (<i>o_r\e</i>) und Trennung) Sonderfall: Sollte eine Ligatur durch einen Zeilenumbruch getrennt werden (zu erkennen an dem kleinen runden <r>), wird auf die Markierung der Ligatur verzichtet. Bsp.:  F173-170,21 (...) <i>ho\e=</i> F173-170,22 <i>ren/ +K ho\e=ren/</i> : {r} ist im Druck Ligatur-r; die Ligatur ist durch Zeilenumbruch getrennt @K <i>wen man anderst</i> (...)
	q_2	Zusammensetzung aus einem <q> und einem <2>-ähnlichen Symbol (Kürzung für <i>quia</i>)

	*f7	Kürzung für lat. <et> = <i>und</i>
	p_p	pp-Ligatur
	p ^v _p	pro-p-Ligatur Bsp.: p ^v _pter = <i>propter</i> Bsp.: p ^v _pheta = <i>propheta</i>
	c'	Zeichen ähnlich dem er-Kürzel über einem <c>. Abkürzung für <i>cetera</i> (häufig mit *f7, also *f7 c' für <i>et cetera</i>)
	§	Schaft-<s> oder langes <s> (im Gegensatz zu rundem <s>)
	&&	Kaufmanns-Und (,ampersand')
	-	Bindestriche wie <-> oder <=>, die Komposita oder andere Wortbildungsprodukte innerhalb einer Zeile verbinden, werden durch einen einfachen Bindestrich dargestellt, Bsp.: Druck: <i>Hof-Haltung</i> → Transkript: <i>Hof-Haltung</i> Bsp.: Druck: <i>Ehe=Vertrag</i> → Transkript: <i>Ehe-Vertrag</i>

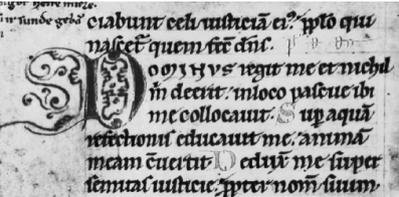
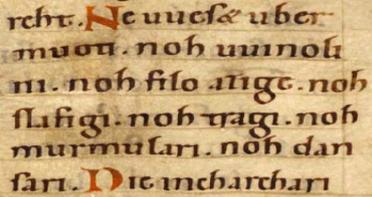
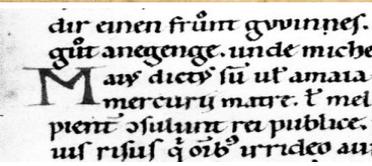
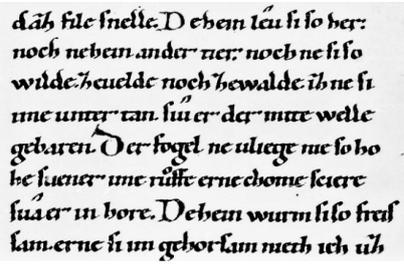
Konventionen spezifisch für Halle		
Abbildung	Kodierung	Beschreibung
  	j\ x\ v\	Modifikation des senkrechten Teils einer römischen Zahl (i.d.R. j, v, x), links oder rechts, zum Anzeigen eines halben Zahlenwertes „gestrichenes“ <j> = einhalb, „gestrichenes“ <v> = viereinhalb, „gestrichenes“ x = neuneinhalb. Bsp. Bild 1: ij\ (= 1 1/2)

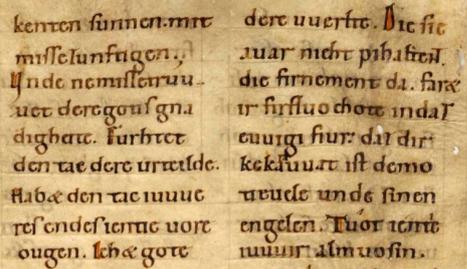
II.2.6 Spezielle lateinische Kürzungen in Halle

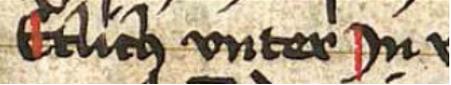
Abbildung	Kodierung	Beschreibung
	\&2	Kürzung für auslautend -em/-en; s. II.2.4. *fid\&2 *quid\&2

	\&3	im Lat. für Endung <i>-is</i> oder auch <i>-ig</i> ; s. II.2.4.
 *fdne\&10 = domine  xps\&10 = xps = Χριστός, Christus  *fdi\$cp\&10los = di\$cipulos  Ihm\&10  *fpp\&10	\&10	Diese Kürzung sieht aus wie Nasalstrich, ist aber meist keiner; kann im Lateinischen für alle mögl. Abkürzungen stehen; da oft die Stelle der Kürzung nicht klar ist, wird im Zweifelsfall die Kürzung an das Ende des Wortes gehängt Bsp.1: *fdne\&10 = domine Bsp.2: xps\&10 = xps = Χριστός, Christus Bsp.3: *fdi\$cp\&10los = di\$cipulos Bsp.4: Ihm\&10 Bsp.5: *fpp\&10o = populo
  	\&11	sieht aus wie eine r-Kürzung, ist aber keine Appli → Appostoli → *fAppli\&11 Johanem → ioh\&11em Sequi → *fSeq\&11
	\&12	quam-Kürzung
	\&15	lateinische Kürzung <i>ue</i> bei angehängtem <i>que</i>

II.2.7 Initialen und Majuskeln

Abbildung	Kodierung	Beschreibung
 <p>Abb. 1</p>  <p>Abb. 2</p>  <p>Abb. 3</p>	<p>*{X*}</p>	<p>(Schmuck-)Initialen</p> <p>Die Höhe (in Zeilen) der Schmuckinitialen wird durch Zahlen in der Codierung wiedergegeben.</p> <p>Bsp.: *{D*3}OMINVS (Abb. 3)</p> <p>Diese Regelung betrifft nur Zeilen, die von der Initiale tatsächlich beeinflusst werden.</p> <p>In Abb. 2 gehört also keine Zahl in die Klammer!!</p>
 <p>Abb. 4</p>  <p>Abb. 5</p>  <p>Abb. 6</p>	<p>*{X*}</p> <p>*{x*}</p>	<p>Satzmajuskeln, Versalien und Lombarden</p> <p>Die Größe des Buchstabens soll durch eine der Zeilenzahl entsprechende Zahl vermerkt werden. Dabei werden die Zeilen gezählt, die durch die Majuskel ‚eingerrückt‘ werden.</p> <p>Bsp.: *(M*2)aivs (Abb. 5)</p> <p>Ist die Majuskel nur eine Zeile hoch, ist die Codierung mit Zahlen überflüssig.</p> <p>Bsp.: *(N*)e (Abb. 4)</p> <p>Es kommt häufig vor, dass Minuskeln in (Satz-)Majuskelfunktion stehen (sog. überhöhte Minuskel). Diese Minuskeln werden *(x*) codiert. Auch hier ist ggf. eine Zahl einzusetzen.</p> <p>Bsp.: *(d*)er (Abb. 6)</p> <p>Da eine definitive Unterscheidung, ob eine Majuskel oder Minuskel</p>

		vorliegt, oft schwierig ist (z.B. bei <i>o</i> und <i>i</i>), wird im Zweifelsfall stets die Majuskel angesetzt.
 <p style="text-align: right;">Abb. 7</p>	X	<p>Einfache Majuskeln</p> <p>Einfache Majuskeln werden als lateinische Großbuchstaben codiert, d.h. A, B, C, D. Bsp.: <i>Die</i></p> <p>Sind einfache Majuskeln zusätzlich rubriziert, können bzw. müssen je nach Textbefund und Frequenz ein Vermerk im Header gemacht werden (wenn es durchgängig so ist) oder Kommentarklammern (+K...@K) für jede Majuskel einzeln (wenn nicht durchgängig rubriziert) angesetzt werden.</p>

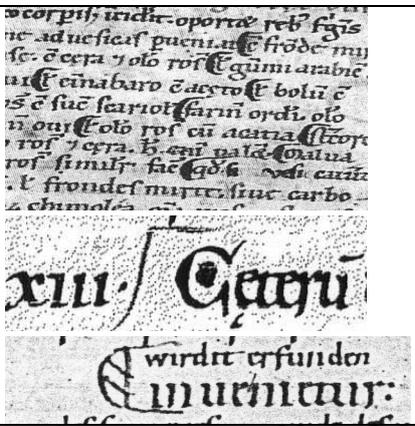
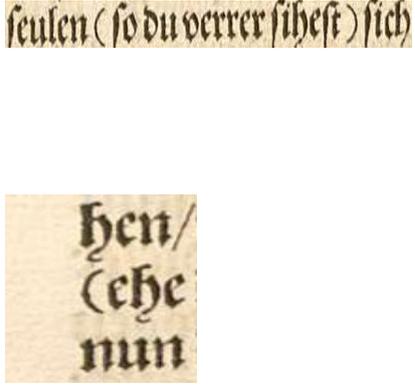
Konventionen spezifisch für Halle		
Abbildung	Kodierung	Beschreibung
	*<x*>	<p>Lombarde</p> <p>im Satz befindliche Buchstaben; wesentlich weniger ausgeschmückt als Initialen, häufig rubriziert (entweder andere Tinte oder Strich durch den Buchstaben)</p>

II.2.8 Interpunktion in der Handschrift/im Druck

Alle ‚originalen‘ Interpunktionszeichen werden ohne Spatium direkt an das Wort angeschlossen, auf das sie folgen.

Ausnahmen hierzu sind einzig die Punkte, die Wörter umschließen, die nur aus einem Buchstaben bestehen, wie z.B. *.S.* = *sanctus* und *.e.* = *ê* („Gesetz“). Bei den Ausnahmen wurde der Punkt nicht an das vorausgehende Wort angeschlossen, sondern wie in den beiden Beispielen vorangestellt.

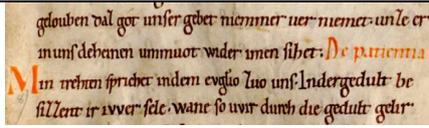
Abbildung	Kodierung	Beschreibung
	.	tiefer („normaler“) Punkt
	%.	<p>mittelhoher (o. halbhoher) Punkt</p> <p>Wenn in der Handschrift/dem Druck ausschließlich mittelhohe und keine normalen Punkte vorzufinden sind, wird im Transskript der normale Punkt verwendet und im Header auf den halbhohen Punkt als Interpunktion hingewiesen.</p>

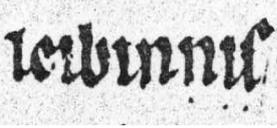
	/	Virgel
	,	Komma
	!	Ausrufezeichen
	?	Fragezeichen
	*C	Alinea/Absatzzeichen <¶> jeglicher Form Bsp.: *C wirdit erfunden
	&(… &)	Runde Klammern werden behandelt wie Interpunktionszeichen. Sie werden an das vorhergehende Wort angeschlossen, indem ein & ohne Spatium vor die Klammer bzw. zwischen Klammer und Wort gesetzt wird. Bsp.: \$eulen&(\$o du verrer sihest&) \$ich Steht die öffnende Klammer am Anfang einer Zeile, so wird sie an das Ende des letzten Wortes der vorhergehenden Zeile gesetzt. Bsp.: F190-10,01 ...mit den dieff\$ten gedancken& F190-10,02 ehe der men\$ch ...

II.2.9 Klammern

Unleserliche Textabschnitte wurden als solche gekennzeichnet.

Abbildung	Kodierung	Beschreibung
	<...>	Bei Transkription schwer lesbare(r) Text, Wörter, Buchstaben.
	[...]	Nicht mehr lesbar und aus Edition übernommen (nicht verwendbar, wenn keine Edition vorliegt!)
	[[...]]	Nicht mehr lesbar und vom Editor erschlossen (nicht verwendbar, wenn keine Edition vorliegt!)
	{2,1}	Editionszählung {Seite, Zeile} oder {Vers}
	+R ... @R	Rubrizierung Bsp.: +R De patientia @R

 <p>gedouben dal got unfer gubet niemmer uer niemet unte er in uns dehanen ummuot wider imen siber: <i>De patientia.</i> Man mehen spricher inden ewigio Iuo uns Lindergedult be sillzen ir zwer sele wane so univ durch die gedult gehir</p>		
	<p>+Ü ... @Ü</p>	<p>Überschrift (unabhängig von der tatsächlichen Größe) und Zwischenüberschrift im Text. Fremdsprachliches Material wird in Überschriften konsequent mit *f gekennzeichnet, unabhängig vom Umfang, da eine Klammerschachtelung nicht erwünscht ist) (Zwischenüberschriften am Rand, s. II.2.II.4 zu Marginalien)</p>
	<p>+K ... @K</p>	<p>Sämtliche Befunde und allg. Anmerkungen zum Überlieferungsträger/zum Text etc., die möglicherweise von Bedeutung sind, aber nicht durch die Transkription dargestellt werden können oder aber zusätzlich zur Transkription der Erläuterung bedürfen, z.B. Erläuterungen zu Streichungen oder Ergänzungen, Textverlust durch Beschnitt oder sonstige Beschädigung, Bemerkungen zu etwaigem Schreiberwechsel. Für den Fall, dass keine Edition vorliegt, aber editorische Anmerkungen eingefügt werden sollen, ist die Abkürzung „evtl. {...}“ ohne Angabe eines Editors/Bearbeiters zu verwenden. Bsp.: Text: oder +K oder: evtl. {über} @K → Hinweise zu allgemeinen Abkürzungen in Klammern</p>
	<p>+E ... @E</p>	<p>Anmerkungen zur Edition. Diese Klammern dürfen nur verwendet werden, wenn eine Edition existiert. Liegt keine Edition vor, sind K-Klammern (s. oben) für z.B. Korrekturvorschläge gegenüber dem Textbefund zu verwenden. Die E-Klammern nach Möglichkeit nur dort verwenden, wo sie als Verständnishaile dienen. Bsp.: Text: drvnter und drrber +E drrber: 1. {drüber}, Editor @E In Editionsclammern sind folgende Abkürzungen häufiger gebraucht: str. = streichen!</p>

		<p>str. NN = streicht NN l. = lies l. NN = liest NN</p> <p>Bsp.: ein rtter +E rtter: l. {ritter}, Lachmann @E</p>
	+L ... @L	Lateinische Passagen, die mehr als drei Wörter umfassen und nicht Teil des deutschen Satzes sind, werden in Latein-Klammern gesetzt.
	+F ... @F	Passagen, die mehr als drei fremdsprachige (nicht lateinische) Wörter umfassen, werden in F-Klammern gesetzt. Dabei müssen fremde (z.B. griechische) Grapheme in eine latinisierte Schreibung übertragen werden. Zusätzlich muss das Ursprungsalphabet in einem nachgestellten K-Kommentar beschrieben werden.
	*[...*]	<p>In der Handschrift durch Rasur, Unterpungierung, Unterstreichung getilgte Wörter und Buchstaben (hier ist eine K-Klammer anzusetzen).</p> <p>Wörter, die in der Handschrift überliefert sind, aber dem Sinn oder der Syntax nach überflüssig sind, z.B. Dittographien, werden übernommen; in einer K-Klammer wird aber darauf hingewiesen, dass das betreffende Wort überflüssig ist.</p> <p>Bsp.: leibin*[ni\$*] +K leibin*[ni\$*]: {ni\$} unterpung. @K</p>

II.2.9.1 Anmerkung zu Klammern

Jedes geklammerte Wort verfügt über eigene Klammern, d.h. dass im Text zusammengeschiedene, jedoch getrennt angesetzte Wörter nicht als Gruppe geklammert werden dürfen. Bei der Klammersetzung von <...>, [...], [[...]] gilt daher folgende Regelung:

richtig: [der][rîter] **falsch:** [der|rîter]

Es kann vorkommen, dass Text innerhalb von Latein-, Rubrizierungs- oder Ü-Klammern kommentiert werden muss. In diesem Fall wurde die Kommentarklammer **hinter** L-, R-, oder Ü-Klammer gesetzt.

Bsp.: +L Nomen *[est*] omen. @L +K *[est*]: gestr. @K

II.2.9.2 Abkürzungen für häufige Bemerkungen in K-Klammern und Absatzmarkierungen

üdZ	über der Zeile
üdZ*	über der Zeile mit Verweiszeichen
untdZ	unter der Zeile
untdZ*	unter der Zeile mit Verweiszeichen
alR/arR	am linken/rechten Rand
alR*/arR*	am linken/rechten Rand mit Verweiszeichen
rad.	Getilgt durch Rasur
unterpung.	Getilgt durch Unterpungierung
gestr.	Getilgt durch Streichung
<	korrigiert aus
AdZ	Abschnittsmarkierung durch Zeichen in der Hs.
AdZLZ	Abschnittsmarkierung durch Zeichen in Hs. und Rest der Zeile leer
AdL	Abschnittsmarkierung durch vergrößerten Abstand zweier Wörter auf einer Zeile (also durch eine Lücke)
Abs	Absatz (das, was wir heute klassisch unter einem Absatz verstehen)
AbsLZ	Absatz durch Leerzeile; die Anzahl der Leerzeilen kann fakultativ angegeben werden
AdHR	Abschnittsmarkierung durch Herausrücken, Einrücken bzw. Hervorheben des ersten Buchstabens bzw. Wortes
AdU	Abschnittsmarkierung durch Unterstreichen des ersten Wortes

II.2.9.3 Nicht lesbarer und verlorener Text

Wenn eine Ergänzung, z.B. aus einer Edition, nicht möglich ist, wurden entsprechend der vermuteten Anzahl nicht lesbarer Buchstaben Punkte in eckigen Klammern gesetzt. Im Header muss auf den Textverlust und den Umgang damit hingewiesen werden.

Bsp.:



01v,02 <h>az v<n>te [ni]<h>t. <d>a<z> [...] [.]<w>[i]

II.2.10 Bildbeischriften

Abbildungen innerhalb eines Drucks oder einer Handschrift, die über Bildbeischriften verfügen, wurden in das Transkript integriert. Sie erhalten **keine** eigene Zeile, sondern werden in einer Kommentarklammer (ähnlich wie Marginalien) erfasst. Zusätzlich müssen sie als Bildbeischriften identifiziert werden mit der Angabe, ob sie sich ober- oder unterhalb des jeweiligen Bildes befinden.

Bsp.: Zwei Bildbeischriften an einem Bild

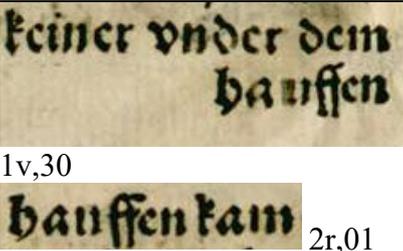
F094-004,03 ge\$etz j\'n früntz\$chaft zu\'* lei\$t\&1 +K Bildbeischrift über dem Bild,
 rubr.: hie lei\$tet ma*- ein fru*ntlichen tag mit denn eigno\$\$en @K +K Bildbeischrift
 unter dem Bild, rubr.: der uo*- zu\'*rich clag @K

II.2.11 Informationen zum Text außerhalb des Haupttextes

II.2.11.1 Kustoden

In frnhd. Texten (besonders Drucken) finden sich sog. Kustoden (manchmal auch Reklaman-
 ten genannt), die auf das erste Wort oder Silbe der nachfolgenden Seite verweisen, indem sie
 es vorwegnehmen.

Beispiele mit Transkription:

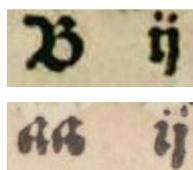
 <p>al= ren 3v ren vnn 4r</p>	<p>3v,24 al= 4r,01 ren vnn</p>
 <p>keiner vnder dem hauffen 1v,30 hauffen kam 2r,01</p>	<p>1v,30 keiner vnder dem 2r,01 hauffen kam</p>

Die Kustode selbst wurde bei der Transkription ignoriert, um die Doppelung des betroffenen
 Wortes im Fließtext zu vermeiden und die Zeilenzählung nicht zu korrumpieren. Die Informa-
 tion, dass der Text Kustoden enthält, wurde im Header zu vermerken.

II.2.11.2 Lagenzählung

Neben der Foliozählung kann auch eine Lagenzählung (meist Buchstaben weit außerhalb des
 Textes) in einem Textträger eingetragen sein.

Bsp.:

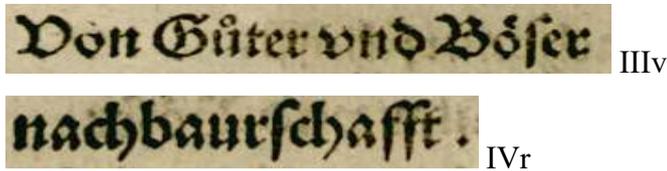


Die Lagenzählung wurde bei der Transkription und im Ordnungsteil ignoriert; auch hier wur-
 de im Header zu vermerkt, dass der Text eine eingetragene Lagenzählung aufweist.

II.2.11.3 Laufende Kolumne

Manche Texte enthalten auf jeder Seite oberhalb des Textspiegels den Namen oder Teile des Namens des Gesamtwerkes. Man nennt diese Information laufende Kolumne.

Bsp.: Von Gu\oter vnd Bo\er nachbaur\$chafft.



Die laufende Kolumne wurde bei der Transkription ignoriert, damit sie weder Zeilenzählung noch Fließtext beeinflusst; auch hier wurde im Header vermerkt, dass der Text eine eingetragene Lagenzählung aufweist.

II.2.11.4 Marginalien

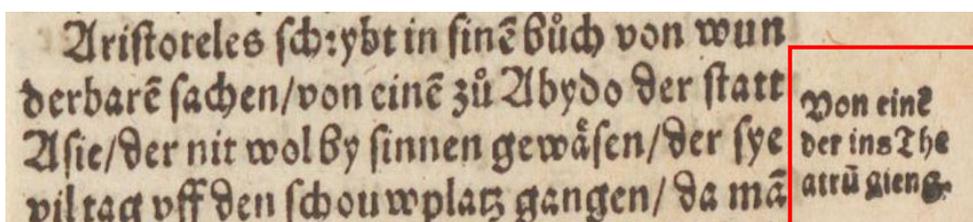
Marginalien sind sehr zahlreich und fassen z.B. häufig zusammen, was im Fließtext, den sie begleiten, steht oder sie verweisen auf Bibelstellen, Daten etc.

Marginalien wurden mit einem eigens dafür konventionalisierten Kommentar ans Ende derjenigen Zeile gestellt, auf deren Höhe die Marginalie beginnt (im Zweifelsfall immer eine Zeile nach oben gezogen): +K Marginalie alR / arR: [Text] @K. Es wurde vermerkt, ob sie am linken oder rechten Rand der Seite steht (stehen Marginalien innerhalb eines Drucks konsequent nur am linken oder nur am rechten Rand, so wurde dies im Header vermerkt und muss nicht in jedem Marginalien-Kommentar einzeln angegeben werden). In der Marginalie realisierte Trennzeichen und präeditorisch zu ergänzende Trennzeichen wurden ohne Spatien an der entsprechenden Stelle eingefügt (s. Bsp.).

Wenn das Ende der Zeile ein getrenntes Wort betrifft, so steht die Marginalienangabe am Ende des jeweiligen Wortes, also in der nächsten Zeile. Im Normalfall gelangt sie dadurch fälschlicherweise in die nächste Zeile, was kommentiert werden muss.

Fremdsprachlicher Text in der Marginalie wurde nicht gekennzeichnet.

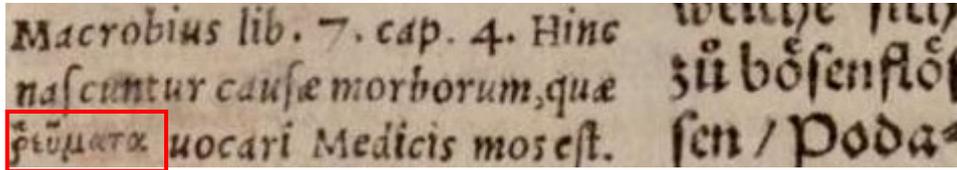
Bsp.:



- | | |
|--------|---|
| 13r,11 | Ari\$toteles \$chrybt in \$ine\ - bu\och von wun(=) |
| 13r,12 | derbare\ - \$achen/ von eine\ - zu\o Abydo der \$tatt +K Marginalie arR:
Von eine\ - der ins The(=)atru\ - gieng. @K |
| 13r,13 | A\$ie/ der nit wol by \$innen gewa\esen/ der \$ye |

Es kann vorkommen, dass Marginalien Text mit nicht-lateinischen Graphemen beinhalten. In diesem Fall wurden die Kommentare wie folgt angesetzt: +K Marginalie: [Text in lateinische Buchstaben übertragen]; (Kommentierung des Alphabets) @K

Bsp.:



+K Marginalie: Macrobius lib. 7. cap. 4. Hinc nascuntur cause morborum, que rheumata uocari Medicis mos est.; (rheumata ist in griechischen Graphemen gedruckt) @K

II.3 Zur Nachbereitung (Sekundärzählung)

Nach (oder wenn möglich auch schon während) der Eingabe des Textes wurde die sog. Sekundärzählung eingefügt, d.h. die Zählung der ggf. verwendeten bzw. vorliegenden Edition. Dies dient der Auffindbarkeit von Textstellen besonders bei Verstexten.

Wie auch schon in ReM steht die Sekundärzählung im laufenden Text in {...} geschweiften Klammern. Bei Prosatexten ist grundsätzlich immer nur der Beginn einer neuen Seite mit der entsprechenden Seitenzahl angegeben, bei Verstexten wird nach Versen gezählt.

Bsp.: F193-901r,01 {003} wir zu dießen geSchwinden vnd be=

F154-03v,09 <z>erzijt/ vnd ande' bij#Si<tz>'e' vp die {423} Rentk<a\>-mer geSchicht vnd geSat wart van|des Raitz

Bei der Trennung eines Wortes durch die Sekundärzählung wurde die Trennung im Transkript nicht in das Wort gesetzt, sondern vor die entsprechende Wortform.

Bsp.: **richtig:** im was {034,02} gedanchet
falsch: im was ge(=){034,02}danchet

III. Präedition

An die Texterfassung schließt sich als nächster Arbeitsschritt die Präeditierung an, d.h. die Aufbereitung der Transkription für die dann folgende computerunterstützte Lemmatisierung und grammatische Bestimmung.

Zur Präeditierung gehören u.a. das Einfügen der nhd. Interpunktion, die Kennzeichnung von fremdsprachigem Material sowie die Überprüfung und ggf. Kennzeichnung der Worteinheiten.

III.1 Wichtige Präeditionszeichen

Die folgende Übersicht enthält die wichtigsten Präeditionskodierungen:

Kodierung	Beschreibung
(=)	nicht sichtbar realisierte Trennung am Zeilenende
=	sichtbar realisierte Trennung am Zeilenende
=	überflüssiges originales Trennzeichen am Zeilenende
#	in der Hs./im Dr. getrennt stehende, vom Bearbeiter zusammengezogene Wörter/Wortteile Bsp.: Hs.: barm vnge → Transkript: barm#vnge
	in der Hs./im Dr. zusammen stehende, vom Bearbeiter getrennte Wörter/Wortteile Bsp.: Hs.: derrîter → Transkript: der rîter
*f	Einzelne fremdsprachliche (häufig lateinische oder griechische) Wörter erhalten die Markierung *f direkt am Anfang des Wortes. *f wird auch gesetzt, wenn fremdsprachliche Wörter Teil eines deutschen Satzes sind. Bsp.: mit dir i\$† ein *fnom\-. von\-. ze evven ze evven <u>Achtung:</u> ‚Item‘ wird immer als fremdsprachlich markiert. Geht eine Alinea voran, steht diese vor der Markierung *f: *C*fItem Bei Wörtern, die im griechischen Alphabet abgedruckt sind, ist die latinisierte Schreibung zu verwenden; zusätzlich muss das Ursprungsalphabet kommentiert werden. Bsp.: rheumata +K rheumata: in griech. Graphemen @K Fremdsprachliche Eigennamen sind, unabhängig von der Flexion, von dieser Kennzeichnung ausgenommen, z.B. <i>Rio de Janeiro</i> , <i>Petrus</i> , <i>Katharina di Medici</i> usw.

III.2 Neuhochdeutsche Interpunktion

Für die Kennzeichnung von Satzgrenzen eines Textes können die modernen Interpunktionszeichen Punkt, Fragezeichen, Ausrufezeichen, Semikolon, Doppelpunkt, Komma und Anführungszeichen verwendet werden.

Nicht möglich ist die Verwendung des Geviertstrichs (‚langer Gedankenstrich‘). Zur Unterscheidung von möglichen handschriftlichen oder im Druck befindlichen Interpunktionszeichen werden die modernen Interpunktionszeichen in runde Klammern gesetzt.

Die Interpunktion folgt im Wesentlichen den gleichen Regeln, die auch in der neuhochdeutschen Schriftsprache gelten.

Die moderne Interpunktion muss sich wie die handschriftliche der Wortform unmittelbar anschließen: Es darf kein Leerzeichen oder sonstiges zwischen Wortform und Interpunktion treten.

In den jüngeren Drucken können auch Satzzeichen auftauchen, die es in ReM noch nicht gab und somit eine strukturierende Funktion erhalten haben, so z.B. runde Klammern oder Anführungszeichen. Da sie nun aber als echte Satzzeichen in frnhd. Drucken vorkommen können, wurden sie mit <&> wie folgt markiert:

Bsp.

seulen (so du verrer sihest) sich

Seulen&(\$o du verrer \$ihe\$t&) \$ich

III.3 Fremdsprachliches Material

Personen- und Ortsnamen (auch Landschaftsbezeichnungen und Völkernamen) werden – auch mit lateinischer Endung – nie mit *f markiert; lemmatisiert wird das deutsche Äquivalent, sofern ermittelbar.

Wörter mit lateinischer Flexionsendung (\neq Orts- und Personennamen) sind als Fremdwörter zu markieren. (s. IV.1.2)

In ReDI werden alle lateinisch flektierten Namen, auch Personen- und Ortsnamen, mit *f markiert.

III.4 Verschmelzungsformen

Verschmelzungen von zwei aufeinander folgenden Wortformen werden durch | aufgelöst und als zwei eigenständige Wortformen behandelt, z.B. durch enklitischen Antritt eines Personalpronomens an das vorhergehende Verb:

Bsp.: *saget|er* \rightarrow *saget er*, *maht|u* \rightarrow *maht du*, *bi\$t|u* \rightarrow *bi\$t du*, *der|s* \rightarrow *der es*

Verschmelzungen von Präposition und Artikel, also Fälle wie *im*, *am*, *zum/zur*, *vom*, werden anders als in ReM dagegen nicht mehr getrennt.

III.5 Partikelverben

Laut HiTS-Aufsatz können als Verbpartikel nur Partikel fungieren, die sowohl Präposition als auch Adverb sind. Zu beachten ist hierbei, dass einige Präpositionen, die heute nur noch als Präp. fungieren, früher auch Adverb sein konnten, wie bspw. *bei* oder *vor*.

Trennbare Präfixe in ReF (nhd. Form):

ab-, *an-*, *auf-*, *aus-*, *bei-*, *durch-*, *ein-*, *mit-*, *nach-*, *nieder-*, *über-*, *um-*, *unter-*, *vor-*/*vür(e)-*, *zu-*, *wider-*

III.5.1 Markierung von Partikelverben

Partikelverben werden mit *1 und *2 markiert, wobei die Basis *2 und die Partikel *1 erhält.

Bsp.: *Er muss an*1|fangen*2 und den Pfeil ab*1|schneiden*2(.)*

Substantivierte Partikelverben werden wie Verben behandelt, d.h. ebenso mit *1 und *2 versehen und getrennt.

Bsp.: *das Auf*1|schneiden*2*

III.5.2 Komposita aus Adverb und Partizip

Es gibt Fälle, die auf den ersten Blick wie Partizipien von Partikelverben aussehen, bei denen aber keine Partikel, sondern ein Adverb vorliegt. Das Adverb ist anders als die Partikel vom Verb (Partizip) zu trennen und als eigene Wortform anzusetzen!

Bsp.: *e|genant*, *ob|genant* (,oben genannt‘), *vor|gemel(de)t*, *vor|genant*, *vor|gesagt*,
vor|geschrieben (,zuvor/vorher geschrieben‘), *nach|geschrieben*
das vor|genommen Mittel = ,das vorher/zuvor genommene Mittel‘

III.6 Pronominaladverbien in Distanzstellung

Pronominaladverbien in Distanzstellung werden mit *1 und *2 markiert, wobei das **Adverb mit *1** und die **Präposition mit *2** gekennzeichnet ist/wurde.

Bsp.: *Das Zimmer da*1 er inne*2 liegt.*

Pronominaladverbien in Kontaktstellung werden (anders als in ReM) mit # zusammengezogen und bei Zusammenschreibung nicht durch | getrennt.

III.7 Negationspartikel *en-*

Negationspartikeln, die der Satzverneinung dienen, werden **abgetrennt**, wenn sie enklitisch oder proklitisch z.B. mit finitem Verb stehen. Dies trifft häufig auf hochfrequente Wörter wie *sein*, *können*, *wollen*, *wissen*, *machen* und *tun* zu.

Bsp.: *di en|salt du nit schouwen* (,die sollst du nicht anschauen‘)

IV. Annotation

Die Annotation erfolgt über die web-basierten Annotationstools CorA (Corpus Annotation) und LAKomp (LemmatisierungAnnotationKomparation).

IV.1 Generelles

Grundsätzlich wurden alle im Transkript aufgenommenen Wortformen annotiert. Dies betrifft auch erschlossene Wortformen (s. [[...]] in Kap. II.2.9), wie zum Beispiel *ir* [[*bruder*]].

Die grammatische Bestimmung im Morpho-Tag erfolgt je nach Beleg. Sofern eine grammatische Bestimmung möglich ist, wurde sie im Morpho-Tag angegeben.

So wird bspw. im Gegensatz zu attributiv gebrauchten Adjektiven bei prädikativ verwendeten Adjektiven i.d.R. nur die Gradationsstufe angegeben.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
schön sein	schön	ADJ	ADJD	Pos.
ein halbes blat	halb	ADJ	ADJA	Pos., Neutr. Nom., Sg.

IV.1.1 Lemmaansatz

Im Allgemeinen wurde auf der Grundlage des DWBs¹⁰ annotiert bzw. lemmatisiert, selbst wenn dadurch Uneinheitlichkeiten übernommen werden können.

Bsp.: *halbtheil – vierteil; beleiben – bleiben, vorcht – furcht, hier – hie;*
 Bereiche *sz, t/th; fern – ferr, ferre → sofern*

Sofern ein Lemma nicht im DWB verzeichnet ist, werden einschlägige Lexika (z.B. Lexer¹¹, landschaftlich gebundene Wörterbücher) herangezogen. Dies wird als Kommentar vermerkt.

Der Lemmansatz bei Varianten ist abhängig von dem Eintrag im DWB; es werden keine Lemmata angesetzt, die im DWB nur Verweisfunktion haben:

s.: es wird der Eintrag gewählt, auf den verwiesen wird (Bsp.: *etzlich, s. etslich* DWB, Bd. 3, Sp. 1189)

vgl.: Eintrag wird als eigenständig angesehen. Lemma kann verwendet werden.

Token	Einträge Grimm, DWB	
nit	nit, s. nicht.	falsch
	nicht, das negative icht	richtig

Sofern das DWB dialektale Formen als Lemma verzeichnet, wurden diese auch als Lemma angesetzt und mit einem Verweis zur hochdeutschen Form versehen.

Bsp.: *treuge* ‚trocken‘ + Verweis zu *trocken*

Lemmatisiert wird i.d.R. nach der Form des Belegs, nicht nach dessen Bedeutung.

Bsp.: *nichts unbändiger doch **denn** die wut des leidigen magens*
 Lemma: *denn*

Bsp.: *wie unsre Vorfahren alle Ehrfurcht **für** ihren Fürsten gehabt, wenn er sie regiert wie er sollte.*
 Lemma: *für* (~vor)

Bei mhd. kontrahierten Verben ist die Langform anzusetzen (Bsp.: *han* → *haben*).

¹⁰ s. DWB.

¹¹ s. Lexer.

Diminutive (auch dialektale Varianten) werden nach dem im DWB verzeichneten Lemma angesetzt. Sofern sie nicht im DWB verzeichnet sind, werden sie neu angesetzt.

IV.1.2 Fremdsprachliche Wörter

Längere lateinische Textpassagen, im Transkript durch sog. Lateinklammern (+L @L; s. II.2.9) gekennzeichnet, wurden nicht annotiert.

IV.1.3 Abkürzungen und Abkürzungen

Abkürzungen werden, wenn möglich (d.h. bei eindeutigem Bezug), aufgelöst. Annotiert wird wie die Vollform, eine morphologische Bestimmung entfällt allerdings, d.h. im POS-Tag wird ausgesternt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
FV\;RSTL.	fürstlich	ADJ	ADJA	*.*.*.*
WIRTTEMB.	Württembergisch [kein Eintrag im DWB]	ADJ	ADJA	*.*.*.*
RAT	rat	NA	NA	Mask.Nom.Sg.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Die H. [=heilige] frouwe	heilig	ADJ	ADJA	*.*.*.*
der fürstl. [= fürstliche] herr	fürstlich	ADJ	ADJA	*.*.*.*

Abkürzungen, die für einzelne Grapheme/Graphemverbindungen stehen, wurden beim Lemmatisieren aufgelöst, so dass die Vollform des Wortes als Lemma erscheint (*vn*-). In der Regel lassen sich die Wortart und die grammatischen Kategorien bestimmen (*lu\$tlíche*-).

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
<i>vn\</i> -	und	KO	KON	
in den lu\$tlíche\ - garten	lustlich	ADJ	ADJA	Mask., Akk., Sg.

Auch bei Abkürzungen, bei denen eine eindeutige Identifizierung des Lexems vorgenommen werden kann, wurde das entsprechende Lemma in Vollform aufgeführt. Eine Wortartenbestimmung auf Lemma- und Beleg-Ebene kann vorgenommen werden, eine grammatische Bestimmung entfällt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
ca	kapitel	NA	NA	–

IV.2 Tagset (Abweichungen/Ergänzungen zu HiTS)

Annotiert wurde nach dem Tagset HiTS (zum Tagset s. VI.6), das im Rahmen der Erstellung der Referenzkorpora erarbeitet wurde. Einzelne Neuerungen und Abweichungen werden im Folgenden aufgeführt.

IV.2.1 Entfallene Tags

DPOSGEN: entfällt für ReF und ReDI

AVD-KO: entfällt für ReF und ReDI

IV.2.2 Einzelne/Neue Tags

POS-Tag	Beschreibung	Beispiel
TRUNC	Kompositionserstglied (1. Glied von Bindestrichkomposita)	Ein- und Ausgang
SPELL	Buchstabierfolge (bisher nicht verwendet)	Meister schreibt man so: M E I S T E R
SYM	Symbol (v.a. bei Inschriften)	Kreuz, Paragraphenzeichen o.ä.
UNK	Wörter deren Sinn nicht erschließbar bzw. nachvollziehbar ist	,snatra, snatra‘ M. (Meister, Magister?) Han\$z

IV.3 Annotation nach lexikalischen Kategorien (alphabetisch sortiert)

IV.3.1 Adjektive

Adjektive werden mit dem Lemma-Tag ADJ versehen. Das POS-Tag richtet sich nach der jeweiligen syntaktischen Funktion: attributiv (ADJA), prädikativ (ADJD), adverbial (AVD) und substituierend (ADJS). Bei der morphologischen Bestimmung wird zudem noch die Gradationsstufe angegeben: Positiv (pos.), Komparativ (komp.) und Superlativ (sup.).

Zur Substantivierung vgl. IV.3.7.2.

IV.3.1.1 Adjektivkomposita mit *aller-*

Verbindungen von *aller-* und direkt folgendem Adjektiv (im Superlativ) werden als Einheit betrachtet (Univerbierung) und somit als Ganzes lemmatisiert sowie als Adjektiv (ADJ) bestimmt. Auf der Prädicationsebene wurden sie durch # zusammengezogen.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
aller#lieb\$ter vater mein	allerlieb\$st	ADJ	ADJA	Sup., Mask., Nom., Sg.

IV.3.1.2 Negation des Partizip Präteritums

Für die Negation des Partizips Präteritum gilt Folgendes: Durch die Präfigierung mit *un-* wird ein Wortartwechsel zum Adjektiv herbeigeführt. Im Korpus finden sich zwei Varianten der Annotation:

Token	Lemma	Lemma-Tag	POS-Tag
unaufgeregt	unaufgeregt	ADJ	ADJx
		VVPP	ADJx

Genuine Adjektive mit Negationspartikel *un-* werden zusammen mit *un-* lemmatisiert, wie z.B. *unartig*.

IV.3.1.3 Adjektive als Modalpartikeln

Wenn ein Adjektiv als Modalpartikel fungiert und ein anderes Adjektiv modifiziert, ist es im POS-Tag mit **AVD** getaggt. Bei der morphologischen Bestimmung wird nur die Gradationstufe angegeben.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
viel besser	viel	ADJ	AVD	Pos.
ganz schön stark	schön	ADJ	AVD	Pos.

IV.3.1.4 Adverbial gebrauchte Adjektive im Komparativ

Adverbial gebrauchte Adjektive im Komparativ wurden im POS-Tag mit **AVD** getaggt. Bei der morphologischen Bestimmung wurde nur die Gradationstufe (Komparativ) angegeben, z.B.:

Bi\$st du ge\$albet(,) So bru\en\$tu de\$st lieber in der hell(,) vil na\her/ dan\ - dein Seel/(,) ...

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
lieber	lieb	ADJ	AVD	Komp.
näher	nahe	ADJ	AVD	Komp.

IV.3.1.5 Lemma-Ansatz unregelmäßiger Adjektive

Die Wortformen *viel* - *mehr* - *meist* werden auf das Lemma 'viel' zurückgeführt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
viel	viel	ADJ	ADJx	Pos.
mehr	viel	ADJ	ADJx	Komp.
meist	viel	ADJ	ADJx	Sup.

Steht *viel* mit Genitiv, wie z.B. in *Ich han den lu\etten hie gar fil gu\otz enttragen\-(.)*, wird das Lemma-Tag **ADJ** und POS-Tag **ADJS** angesetzt. Eine weitere grammatische Bestimmung entfällt.

Die Wortformen *gut* – *besser* – *best* werden auf das Lemma ‚gut‘ zurückgeführt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
gut	gut	ADJ	ADJx	Pos.
besser	gut	ADJ	ADJx	Komp.
best	gut	ADJ	ADJx	Sup.

IV.3.2 Adverbien

IV.3.2.1 Konjunkionaladverbien

Abweichend von HiTS wurden Konjunkionaladverbien im Lemma-Tag einheitlich mit **AVD** >**AVD** (nicht: **AVD-KO** >**AVD**) getaggt, z.B.:

De\$zhalb \$ich nyman mag beru\emen/(,)

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
De\$zhalb	deshalb	AVD	AVD	*

IV.3.2.2 Pronominaladverbien

Pronominaladverbien in Kontaktstellung wurden im Lemma-Tag mit **AVDAP**, im POS-Tag mit **PAVDAP** getaggt. Stehen sie getrennt im Transkript, wurden sie mit # zusammengezogen.

Bei Pronominaladverbien in Distanzstellung (*da*1 ... für*2*) wurde (anders als in ReM) ähnlich verfahren wie bei den mehrteiligen Partikelverben. Die vollständige Lemmaangabe befindet sich im Lemma des präpositionalen Teils (*dafür*). Der adverbiale Teil enthält nur das Adverb (*da*).

Der präpositionale Bestandteil wird im Lemma-Tag mit **AP**, im POS-Tag mit **PAVAP** angegeben.

Der adverbiale (Erstbestand-)teil (z.B. *da*) wird im POS-Tag mit **PAVD**, im Lemma-Tag mit **AVD** getaggt.

Token	Lemma	Lemma-Tag	POS-Tag
da#für	dafür	AVDAP	PAVDAP
da*1...	da	AVD	PAVD
kann ich nichts			
...für*2	dafür	AP	PAVAP

Im POS-Tag können zudem in interrogativer und relativischer Verwendung Pronominaladverbien in Kontaktstellung auch als **PAVRELAP** (Pronominaladverb, relativisch) oder **PAVWAP** (Pronominaladverb; interrogativ) getaggt werden. In Distanzstellung ergeben sich entsprechend **PAVREL** (relativisch) und **PAVW** (interrogativ) für den pronominalen Teil und **PAVAP** für den präpositionalen Teil.

*da*1 (AVD > PAVREL) daz volk inn*2 (AP > PAVAP) \$terben \$chu\ell*

IV.3.3 Eigennamen

Für Eigennamen (z.B. Vor-, Nach-, Ortsnamen) wird als Lemma- und POS-Tag **NE** verwendet.

Sofern Eigennamen im DWB aufgeführt sind, fungieren die dort angesetzten Formen als Lemmata.

Ist ein Name nicht verzeichnet, wurde eine dem Nhd. entsprechende normalisierte Form angesetzt bzw. erfolgt der Lemmansatz nach der Edition [ReDI].

Lässt sich der Name nicht mehr auf einen heute geläufigen Namen zurückführen, wurde eine handschriftennahe Form angesetzt (v.a. Nachnamen).

In Halle erfolgte keine morphologische Annotation von Eigennamen. In Bochum wurden morphologische Annotationen i.d.R. vorgenommen, sofern sie anhand der Flexionsformen zugeordnet werden können.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
HEINRICH	heinrich	NE	NE	Mask.Nom.Sg. oder *. *. *.
WECKENESEL	weckenesel	NE	NE	*.Nom.Sg. oder *. *. *.
(das Land) Spanien	spanien	NE	NE	*.Nom.Sg. oder *. *. *.

In ReDI wurden die Eigennamen soweit möglich morphologisch annotiert; bei Nachnamen wurde in der Regel kein Genus vergeben, Ausnahme stellen die mit -in movierten femininen Nachnamen dar. (Zur ausführlichen morphologischen Annotation von Eigennamen in ReDI s. Anhang)

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
FRAU	frau	NA	NA	Fem.Nom.Sg.
MARIA	maria	NE	NE	Fem.Nom.Sg.
BEIERIN	beyer	NE	NE	Fem.Nom.Sg.
HERR	herr	NA	NA	Mask.Nom.Sg.
HEINRICH	wolf	NE	NE	Mask.Nom.Sg.
BEIER	beyer	NE	NE	*.Nom.Sg.

Tritt zu einem Namen die Präposition 'von', die eine Herkunft oder die Zugehörigkeit zum Adel anzeigt, (z.B. *Heinrich von Pfalzpaint*), wurde diese in Bochum und in ReDI als Präposition (APPR), in Halle als Eigenname (NE) angesetzt.

Monatsnamen wurden als NA (nicht: NE) getaggt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
MAI	mai	NA	NA	Mask.Nom.Sg.

IV.3.4 Fragewörter

Fragewörter (z.B. *wo*, *weshalb*) in direkten und indirekten Fragesätzen wurden einheitlich mit **AVW** getaggt. Eine Ausnahme bilden die Interrogativpronomen *wer* und *was*, die als **PW** getaggt wurden.

IV.3.4.1 *Wer?, Was?*

W-Interrogativpronomen werden auf die Lemmata *wer* bzw. *was* zurückgeführt. Sie wurden sowohl bei interrogativer als auch bei relativischer Verwendung im Lemma- und POS-Tag mit **PW** annotiert, z.B.

**C Vnnd wer von dem *(G*)ycht nicht gereden mag(,) der trinck Sta\et verbena#\$afft(.)*
[...] wan\ - daz ir \$chol \$ehen waz vor im i\$t(.)

IV.3.4.2 *Welch-*

Fungiert *welch-* als Relativpronomen, erhält es das Lemma-Tag **DW** und das POS-Tag **DWx** (nicht: **DRELS/DD** oder **DGx/DG**), z.B.:

und es ist ein groß artney das pulver zu den wunden, welcher hande die \$ynt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
welcher	welch	DW	DWA	Gen.Fem. Sg.

IV.3.4.3 Vergleichendes *Wie*

Wenn ein vergleichendes *wie* einen Nebensatz einleitet, wurde unterschiedlich annotiert:

Bei Fragen, z.B. *we mir arme\-(,) wie i\$t mir ge\$che=hen(?)*, wurde im Lemma- und POS-Tag **AVW** verwendet.

In allen anderen Fällen, z.B. *\$o dunkt dem me\-= \$chen(,) wie im ains in den ore\ - pavk*, wurde im Lemma-Tag **KO** und im POS-Tag i.d.R. **KOKOM** angesetzt.

Analog zum vergleichenden *wie* wurde auch *sam* als **KO/KOKOM** getaggt. Bei *sam* handelt es sich ursprünglich um ein Adverb, das in vergleichender Funktion noch in der Frühphase des Frnhd. auftritt, z.B.

So ift auch der svn als alt(,) Sam der vater mag gefein(.)

IV.3.5 Junktionen

Junktionen werden mit dem Lemma-Tag **KO** getaggt, das POS-Tag unterscheidet sich je nach Distribution: nebenordnend (**KON**), vergleichend (**KOKOM**) unterordnend (**KOUS**).

Steht eine Nebensatzeinleitende (subordinierende) Konjunktion in einem V2-Nebensatz, wurde die Konjunktion im POS-Tag mit **KO*** getaggt, z.B.:

*So#lang nun ein jeder Badga\$tu*rd in dem Bade ohne be\$chwa*r=nu\$ _z vnd \$chwachheit baden ko\ennen/(,)*

Mehrteilige Konjunktionen in Kontaktstellung wurden auf der Ebene der Transkription durch # zu einem Lemma zusammengezogen. Die univertierte Form wurde als Lemma neu angesetzt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
auf#dasz	aufdasz	KO	KOUS	–
um#dasz	umdasz			
so#dasz	sodasz			

Dies betrifft auch Fälle wie *soviel* und *sobald*, z.B.: *kind(,) so#bald es auß mutter#leib kompt (KO > KOUS)*, die als Konjunktionen verwendet werden. Bei adverbialen Gebrauch wurden beide Bestandteile hingegen einzeln lemmatisiert und getaggt.

Bei mehrteiligen Konjunktionen, wie z.B. *sowohl... als auch, wie wenn, als wie, ob ... wohl, auch wenn, als ob* etc., sind die Erstbestandteile im Lemma-Tag als **KO**, die weiteren Bestandteile im Lemma-Tag als **KO** oder **AVD** annotiert.

Als Konjunktion wurde das Lemma *dieweil(e)* (im Sinne von ‚solange‘) auf der Präeditionsebene (mit #) zusammengezogen. Getaggt wurde **KO > KOUS**. Fungiert *dieweil(e)* dagegen als Adverb, wurde **AVD > AVD** getaggt.

IV.3.6 Artikelwörter und Pronomina

(Flektierte) Artikelwörter und Pronomina werden im Lemma i.d.R. auf die maskuline Form im Nominativ Singular zurückgeführt. Wird im DWB bspw. die flektierte Form eines Personalpronomens mit eigenem Eintrag aufgeführt (z.B. ‚uns‘ statt der Grundform ‚wir‘), wurde dennoch die Grundform ‚wir‘ als Lemma angesetzt. Die Genusangabe in der morphologischen Annotation folgt dem Beleg.

IV.3.6.1 Definitartikel

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
das	der	DD	DDART	Neutr.Nom./Akk.Sg.
Buch	buch	NA	NA	Neutr.Nom./Akk.Sg.

IV.3.6.2 Demonstrativpronomina

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
diese	dieser	DD	DDA	Fem.Nom./Akk.Sg.
Frau	frau	NA	NA	Fem.Nom./Akk.Sg.

Vnd drinck deß nuchtern

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
deß	der	DD	DDS	Gen.Sg.

Die zu den Demonstrativpronomen gehörenden Genitivformen *dessen* (Mask./Neutr.) / *deren* (Fem./Pl.) kongruieren nicht mit dem Substantiv, vor dem sie stehen, sondern beziehen sich anaphorisch auf ein vorher im Text genanntes Substantiv, nach dem sich auch das Tagging (Numerus und Genus) richtet, z.B.:

der junge Mann und **dessen** Schwägerin

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
dessen	der	DD	DDS	Mask.Gen.Sg.

zu **deren** gedechtnus

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
deren	der	DD	DDS	*.Gen.Pl.

IV.3.6.3 Indefinitpronomina

Indefinitpronomina werden danach unterschieden, ob sie generell attributiv verwendet werden können (Lemma-Tag: DI) oder nicht (Lemma-Tag: PI). Die morphologische Annotation erfolgt nach jeweiligem Beleg – außer bei indeklinablen Wörtern (z.B. *etwas*, Wörter auf *-lei*, *man*).

Indefinitpronomina, die nicht attributiv verwendet werden können (Pronominalsubstantive) wie bspw. *jedermann*, *irgendwer*, *jemand*, wurden im Lemma- und POS-Tag als **PI** getaggt, z.B.:

nit **yedermans** ding ißt

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
yederman	jedermann	PI	PI	M.Gen.Sg.

Attributiv verwendete Indefinitpronomina, wie bspw. *ander-*, *manch-*, *einig-*, wurden i.d.R. im Lemma-Tag als **DI**, im POS-Tag als **DIx** getaggt, z.B.:

vnd chain **andere** creatur

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
ander	ander	DI	DIA	F.Nom.Sg.

Sonderfälle wie bspw. *man* oder *nichts* werden folgendermaßen getaggt:

Token	Lemma	Lemma-Tag	POS-Tag
man	man	NA	PI
nicht	nicht	PI	PTKNEG

nichts	nichts	PI	PNEG
niemand	niemand	PI	PNEG
icht	icht	PI	PI (positive Funktion) PTKNEG oder PNEG (negative Funktion)

IV.3.6.4 Interrogativpronomina

Interrogativpronomina werden folgendermaßen getaggt:

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Wer	wer	PW	PW	Mask/Fem.Nom.Sg.
Was				Neutr.Nom.Sg.
Wessen				*.Gen.Sg.
Wem				*.Dat.Sg.
Wen				*.Akk.Sg.

In der Arbeitsstelle Halle erfolgt die Angabe des Numerus nur im Nominativ.

IV.3.6.5 Personalpronomina

Das Genus von Personalpronomina wurde ausschließlich in der 3.Sg. angegeben.

Der Lemma-Ansatz richtet sich wie in ReM und MiGraKo nach der **Nom.Sg./Pl.-Form** (,Nennform‘) des Personalpronomens. Als Lemmata angesetzt wurden *ich, du, er, wir, ihr, sie*. In der **3.Sg.** wurde nur die maskuline Form *er* angesetzt, d.h. *sie* (f.) und *es* (n.) entfallen als Angaben im Lemmaansatz analog zur Angabe Mask.Nom.Sg.-Form beim Definitartikel *der*.

Personalpronomina wurden als **PPER** getaggt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
an	an	AP	APPR	–
ihr	er	PPER	PPER	Fem.Dat.Sg.

Ge=lu\$*dich* dez(,) daz \$u\ech...

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
dich	du	PPER	PPER	*.Akk.Sg.

IV.3.6.6 Possessivpronomina

Die Annotation des Possesivpronomens richtet sich nach der KNG-Kongruenz des Bezugsnomens. Analog zum Verfahren in ReM und MiGraKo richtet sich der Lemma-Ansatz nach der **Nom. Sg./Pl.-Form**, d.h. als Lemmata angesetzt werden *mein, dein, sein, unser, euer, ihr*. In der **3.Sg.** wurde nur die maskuline Form *sein* angesetzt, d.h. *ihr* (f.) und *sein* (n.) entfallen im Lemmaansatz.

So tū man **vn\$erm h'ren** we oder vnd i\$st auch un\$'m h'ren vnmar(.)

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
unserm	unser	DPOS	DPOSA	Mask.Dat.Sg.
herren	herr	NA	NA	Mask.Dat.Sg.

Das in ReM vorhandene Tag **DPOSGEN** entfällt für ReDI/ReF¹², z.B.:

seines alters im 47 jahre

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
seines	sein	DPOS	DPOSA (≠ DPOSGEN)	Neutr.Gen.Sg.

IV.3.6.7 Reflexivpronomina

Der Lemma-Ansatz der Reflexivpronomina richtet sich in der **1./2.Sg./Pl.** nach der **Nom.Sg./Pl.-Form** des entsprechenden Personalpronomens, d.h. als Lemmata wurden auch die Personalpronomen **ich, du, wir, ihr** angesetzt.

In der **3.Sg./Pl.** wurde hingegen **sich** als Lemma angesetzt. Es ist genus- und numerusneutral und identisch für die Kasus Dativ und Akkusativ (**PRF**). Seltener finden sich zudem (vorwiegend im Dativ) reflexivisch verwendete Formen des Personalpronomens der **3. Sg./Pl.** (Lemma: Sg.: **er**, Pl.: **sie**), z.B.:

das nam er im (Dat. von er) *zw herczen*

die sy ... bey ir (Dat. von sie) *hete*

damit sie jhnen (Dat.Pl. von sie) *die Zeit kürtzer machen*

Neben dem reflexiven Gebrauch der Personalpronomina können auch andere Pronomina reflexiv verwendet werden. Dies wird im POS- und Lemma-Tag gekennzeichnet.

Token		Lemma	Lemma-Tag	POS-Tag	Morpho-Tag	
1.	Sg.	<i>ich erinnere mich</i>	ich	PPER	PRF	1.Sg.Akk.
2.		<i>du Erinnerst dich</i>	du	PPER	PRF	2.Sg.Akk.
3.		<i>er/sie/es erinnert sich</i> er → ‚ihm‘, sie → ‚ihr‘	sich er	PRF PPER	PRF PRF	3.Sg.Akk. 3. Sg. Dat.
1.	Pl.	<i>wir erinnern uns</i>	wir	PPER	PRF	1.Pl.Akk.
2.		<i>ihr erinnert euch</i>	ihr	PPER	PRF	2.Pl.Akk.
3.		<i>sie erinnern sich</i> sie → ‚ihnen‘	sich sie	PRF PPER	PRF PRF	3.Pl.Akk. 3.Pl.Dat.

¹² Dipper u.a. (2013), 99.

IV.3.6.8 Rezipropronominale

Wie bei den Reflexivpronominale wurde auch bei den Rezipropronominale das Beleg-Tag **PRF** verwendet, z.B.:

*Sie \$ull\|- **einander** dynen.*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
einander	einander	PPER	PRF	3.Pl.Dat.

IV.3.6.9 Relativpronominale

Relativpronominale werden im Lemma-Tag mit DD und im POS-Tag mit DRELS markiert, z.B.:

*vnd in de\|- i\$st der \$el chraft(,) **die** da haizt *ffanta\$tica oder *fymaginaría*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
die	der	DD	DRELS	Fem.Nom.Sg.

In den frnhd. Texten treten häufig verschachtelte Sätze auf (z.B. *do sprachen die pi ir woren*¹³), in denen ein Demonstrativpronomen im Hauptsatz mit einem Relativpronomen des Nebensatzes zusammenfällt. In der Frnhd.Gr., 273 wird diese Konstellation als „Demonstrativrelativ“ bezeichnet. In solchen Fällen wurden **Relativpronomen** angesetzt.

IV.3.6.10 *selb/selber/selbst*

Alle Formen von *selb* (z.B.: *selbs, selbst, selber*) wurden im Lemma mit DD getaggt, als Beleg-Tags fungieren je nach Position: DDA/DDN/DDS. Die grammatischen Angaben der erstarrten nachgestellten Formen wurden ausgesternt.

Geht *selb* der bestimmte Artikel voran, werden beide Bestandteile als lexikalische Einheit aufgefasst und entsprechend lemmatisiert. Dabei wurden die feminine und neutrale Form auf die maskuline zurückgeführt: *dieselbe, dasselbe > derselbe*.

Bei Getrenntschreibung wurde bei der Präedition eine Univerbierung durch # vorgenommen.

IV.3.7 Substantive

Substantive werden durch die Tags NA im Lemma- und im POS-Tag gekennzeichnet. Zu Eigennamen s. IV.3.3.

¹³ Frnhd. Gr. § S 273.

IV.3.7.1 Morphologische Annotation

In ReF wurde das **Genus** von Substantiven nur dann angegeben, wenn es eindeutig anhand der Flexionsendung oder anhand des voranstehenden Artikels/Pronomens erkennbar ist. Das Entscheidungskriterium ist also die Kongruenz in der NP.

In allen anderen Fällen und bei Unsicherheiten wurde das Genus ausgesternt. Steht ein Lemma im Plural, wurde das Genus generell nicht angegeben, z.B.:

*Sein **ler** und sein **leben** sindt schlecht*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
sein	sein	DPOS	DPOSA	*.Nom.Sg.
ler	lehre	NA	NA	*.Nom.Sg.
und	und	KO	KON	–
sein	sein	DPOS	DPOSA	*.Nom.Sg.
leben	leben	VV	NA	*.Nom.Sg.

In ReDI wurde das Genus von Substantiven in der POS-Angabe hingegen in der Regel so getaggt, wie es im Wörterbuch (Deutsches Wörterbuch von Grimm) für das jeweilige Lemma angegeben ist (sowohl im Sg. als auch im Pl.); wenn für den konkreten Beleg morphologisch ein abweichendes Genus bestimmbar ist, wurde dieses angegeben.

Der **Kasus** bei deklinablen Wörtern wurde auch angegeben, wenn durch Zusammenfall oder mangels eines Artikels die Flexion nicht erkennbar ist und sich nur aus dem syntaktischen Zusammenhang ergibt. Als Entscheidungskriterien zählen die Flexionsendung, der Artikel, die Verbrektion und die Satzstellung. In Zweifelsfällen wurde kommentiert.

IV.3.7.2 Substantivierungen

Bei Substantivierungen wurde im Lemma-Tag jeweils die Ausgangswortart angesetzt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
das Siegen	siegen	VVINF	NA	Neutr.Nom.Sg.
der Nächste	nah	ADJ	ADJS	Mask.Nom.Sg.
der Auserwählte	auserwählen	VVPP	NA	Mask.Nom.Sg.

Liegt eine lexikalische Konversion vor, wurde hingegen nicht die Ausgangswortart angegeben.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
der Lauf	Lauf	NA	NA	Mask.Nom.Sg.

Für substantivierte Adjektive und Zahlen gibt es im Tag-Set spezielle Tags: **ADJS** bzw. **CARDS**.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
ER	er	PPER	PPER	3.Sg.Mask.Nom
IST	sein	VA	VAFIN	3.Sg.Präs.Ind.Unr.
ZUR	zur	AP	APPRDDART	–
RECHTEN	recht	ADJ	ADJS	Pos.Fem.Dat.Sg.
GESATZT	setzen	VV	VVPP	–

Liegt eine Genitivkonstruktion vor, wurde beim Lemma-Ansatz bei der Unterscheidung zwischen Kompositum und Genitivkonstruktion im Zweifel für die Analyse als Genitivkonstruktion entschieden.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
der	der	DD	DDART	Fem.Gen.Sg.
sunnen	sonne	NA	NA	Fem.Gen.Sg.
ûfganc	aufgang	NA	NA	Mask.Nom.Sg.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
der	der	DD	DDART	Mask.Nom.Sg.
sunnen#tag	sonnentag	NA	NA	Mask.Nom.Sg.

IV.3.7.3 Bindestrichkomposita

Der abgetrennte Teil erhält das volle Lemma und entsprechend als Lemma-Tag NA, als POS-Tag ist TRUNC angesetzt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
(die)				
GRAF	Grafschaft	NA	TRUNC	Fem.Nom.Sg.
UND				
HERRSCHAFT	Herrschaft	NA	NA	Fem.Nom.Sg.

IV.3.8 Verben

Beim Taggen von Verben wurden Modalverben (VM), Auxiliarverben (VA) und Vollverben (VV) unterschieden. Bei der morphologischen Annotation werden grundsätzlich Person, Numerus, Tempus, Modus und Verbklasse (sw., st., unreg.) angegeben.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
ich	ich	PPER	PPER	1.Sg.*Nom.
gehe	gehen	VV	VVFIN	1.Sg.Präs.Ind.unr.

Die morphologische Annotation aus heutiger Sicht inkongruenter Belege (bspw. dialektaler Formen) orientiert sich i.d.R. an der Person.

*das jr vil lieber \$u\en haben/(,) \$o jhr zu\o dem loblichen vnnd kun\$tlischen handtwerck des Goldt\$chmidens ab*1=|richten*2.(.)*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
jhr	ihr	PPER	PPER	2.Pl.*Nom.
ab*1=	ab	AVD	PTKVZ	–
richten*2 (↔ ‚abrichtet‘)	abrichten	VV	VVFIN	2.Pl.Präs.Ind.sw.

IV.3.8.1 Auxiliarverben

Als Auxiliarverben (VA) wurden alle Formen von *haben*, *sein* und *werden* getaggt, unabhängig davon, ob sie im konkreten Beleg als Voll- oder Kopulaverben fungieren.

Die Verben *sein* und *haben* werden der Klasse der unregelmäßigen Verben zugeordnet, *werden* der Klasse der starken Verben.

Alle Formen, die ursprünglich zum Paradigma von *wesen* gehörten, d.h. auch *wis* und *war/was*, wurden auf *sein* zurückgeführt.

IV.3.8.2 Modalverben

Als Modalverben (VM) werden getaggt: *dürfen*, *können*, *mögen*, *müssen*, *sollen*, *wollen*.

IV.3.8.3 Übersicht POS-Tag-Zuordnung der frnhd. unregelmäßigen Verben

	Vollverb (VV)	Hilfsverb (VA)	Modalverb (VM)
Präteritopräsentien	<i>gönnen, taugen, tur- ren, wissen</i>		<i>dürfen, können, mögen, müssen, sollen</i>
Wurzelerben	<i>gehen (gân), stehen (stân), tun</i>	<i>sein</i>	
urspr. kontrahierte Verben	<i>lassen (lân)</i>	<i>haben</i>	
Modalverben			<i>dürfen, können, mögen, müssen, sollen, wollen</i>

IV.3.8.4 Partikelverben

Partikelverben (auch: trennbare Präfixverben) sind Verbindungen aus einem Basisverb und einer (Verb-)Partikel, d.h. einem (Präpositional-)Adverb.

Als (trennbare) Verbpartikel wurden in Anlehnung an HiTS nur solche Verbbestandteile betrachtet, die gleichzeitig Präposition und Adverb sein können (Adverbien, zu denen es eine gleichlautende Präposition gibt; vgl. auch Kap. III.5). Erweitert um *durch*, *mit*, *unter* sind dies:

<i>an</i>	<i>mit</i>	<i>aus</i>	<i>in</i>
<i>ab</i>	<i>nach</i>	<i>vor</i>	<i>ni(e)der</i>
<i>bei</i>	<i>auf</i>	<i>für</i>	<i>unter</i>
<i>durch</i>	<i>um</i>	<i>zu</i>	<i>wider ‚zurück‘</i>

Trotz möglicher Fernstellung besteht eine feste syntaktische und semantische Einheit, vgl. nhd. *jdn. ansprechen – er spricht jdn. an.*

In der Präedition wurden Partikelverben mit *1 (Verbpartikel) und *2 (Basisverb) markiert; in Kontaktstellung wurden sie durch | getrennt.

Kontaktstellung: *dô er den brief **ane*1|sach*2***

Fernstellung: *er **sach*2** den brief **ane*1***

Bei der Annotation wurden die Bestandteile als selbständige Wörter behandelt und zugleich als zusammengehörig gekennzeichnet: So wurde für jeden Bestandteil, d.h. der Partikel und dem Basisverb, ein eigenes Lemma angesetzt: Der adverbiale Teil erhält hierbei als Lemma-Tag **AVD** und als POS-Tag **PTKVZ**.

Bei Partikelverben mit einem abgetrennten Verbzusatz erhält das Simplex als Lemma das volle Verbkompositum. Der adverbiale Teil erhält als Lemma nur das Adverb/die Präposition.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
sach*2	anschen	VV	VVFIN	3.Sg.Prät.Ind.st.
den				
brief				
ane*1	an	AVD	PTKVZ	–

Bei getrennten Partikelverben mit zwei oder mehr abgetrennten Partikeln erhält der Verbteil als Lemma das Simplex, das durch geschweifte Klammern gekennzeichnet wurde.

Token	Lemma
machen*2	{machen}
den Mund	
auf*1	aufmachen
und	
zu*1	zumachen

V.3.8.4.1 Abgrenzung zwischen trennbarem Partikelverb und untrennbarem Verb

Liegt im Gegensatz zu einem trennbaren Partikelverb ein untrennbares Verb mit formgleichem Erstbestandteil (z.B. häufig bei *durch-*, *über-*, *um-*, *unter-*, *wider-*) vor, wurde der Erstbestandteil nicht abgetrennt, z.B.:

*hieten si dich rehte erkant, si hieten **widerzogen** ir hant (widerziehen ‚zurückziehen‘)*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
si	sie	PPER	PPER	3.Pl.*.Nom.
hieten	haben	VA	VAFIN	3.Pl.Prät.Konj.unr.
widerzogen	widerziehen	VV	VVPP	

V.3.8.4.2 Abgrenzung zwischen Verbpartikeln und Adverbien

Handelt es sich nicht um ein Partikelverb, sondern um eine Zusammensetzung aus Adverb und Verb (z.B. *hinlaufen*), wurden beide Bestandteile gesondert annotiert. In der Prädiction wurde der adverbiale Bestandteil (**hin**|*laufen*) abgetrennt.

Token	Lemma	Lemma-Tag	POS-Tag
hin	hin	AVD	AVD
laufen	laufen	VV	VVINF

Bei Zusammensetzungen aus Adverb und dem Partizip II ist jeweils entschieden worden, ob der adverbiale Teil Verbpartikel oder selbständiges Adverb ist.

Partikelverb: *Sie hat sich etwas vor*1|genommen*2.*

Token	Lemma	Lemma-Tag	POS-Tag
vor*1	vor	AVD	PTKVZ
genommen*2	vornehmen	VV	VVPP

Adverb + Verb: *das vor|genommen Mittel* (= das vorher genommene Mittel)

Token	Lemma	Lemma-Tag	POS-Tag
vor	vor	AVD	AVD
genommen	vornehmen	VVPP	ADJA

V.3.8.4.2 Abgrenzung zwischen Verbpartikeln und Pronominaladverbien

Es gibt Fälle, in denen die Partikel bzw. das Adverb sowohl als Teil des Pronominaladverbs (z.B. *dar an, dar nach, dar zuo, dar aus*) als auch als trennbare Verbalpartikel aufgefasst werden kann, z.B.:

*es wychet*2 balde dar|aus*1*

→ *aus*1 wychen*2* (Verbpartikel)

Token	Lemma	Lemma-Tag	POS-Tag
wychet*2	ausweichen	VV	VVFIN
dar	dar	AVD	AVD
aus*1	aus	AVD	PTKVZ

es wychet balde daraus
 → *daraus* (Pronominaladverb)

Token	Lemma	Lemma-Tag	POS-Tag
wychet	weichen	VV	VVFIN
daraus	daraus	AVDAP	PAVDAP

IV.3.8.5 Infinitive Verbformen

IV.3.8.5.1 Infinitive

Infinitive wurden grundsätzlich mit **VV/VA/VM** im Lemma-Tag und **VVINF/VAINF/VMINF** im POS-Tag gekennzeichnet.

Entsprechend wurden auch Infinitive, die mit *zu* erweitert sind, im POS-Tag mit **VVINF** getaggt; das Tag **VVIZU** wurde hingegen nicht mehr verwendet.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
(um)	um	KO	KOUI	–
zu	zu	PTK	PTKZU	–
leben	leben	VV	VVINF	unflekt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
i\$t	sein	VA	VAFIN	3.Sg.Präs.Ind.unr.
zu\o	zu	PTK	PTKZU	–
wi\$\$ene	wissen	VV	VVINF	flekt.

Wird der Infinitiv eines Verbs allerdings substantiviert, wurden die als POS-Tags vorgesehenen Tags **VVINF/VAINF/VMINF** als Lemma-Tag angesetzt. Im POS-Tag wurde die Substantivierung entsprechend durch **NA** gekennzeichnet, z.B.:

{A}uch *nimt d' me=nsch \$ein narunge mit ezzen vnd mit trinchen* [...]

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
mit	mit	AP	APPR	–
ezzen	essen	VVINF	NA	Neutr.Dat.Sg.

Grundsätzlich wurde nach morphologischen und nicht nach semantischen Kriterien annotiert, d.h. modale Infinitive wurden bei der Annotation nicht gesondert gekennzeichnet.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
das	der	DD	DDS	Neutr.Nom.Sg.
ist	sein	VA	VAFIN	3.Sg.Präs.Ind.unr.
zu	zu	PTK	PTKZU	–
wissen	wissen	VV	VVINF	unflekt.

IV.3.8.5.2 Partizipien

Partizipien werden grundsätzlich mit **VV/VA/VM** im Lemma-Tag gekennzeichnet, sofern sie Teil des Prädikats sind. In prädikativer Verwendung wurden die POS-Tags **VVPS/VAPS/VMPS** (Partizip Präsens) oder **VVPP/VAPP/VMPP** (Partizip Präteritum) angesetzt.

Token	Lemma-Tag	Lemma	POS-Tag	Morpho-Tag
es	PPER	er	PPER	3.Sg.Neut.Nom.
ist	VA	sein	VAFIN	3.Sg.Präs.Ind.Unr.
geschmolzen	VV	schmelzen	VVPP	unflekt.

In adjektivischer Verwendung des Partizips eines Verbs wurden allerdings die als POS-Tags vorgesehenen Tags **VVPS/VAPS/VMPS** bzw. **VVPP/VAPP/VMPP** als Lemma-Tag angesetzt. Im POS-Tag wurden entsprechend die unterschiedlichen Verwendungsweisen angegeben (**ADJx**), z.B.:

*das lebentyg **machent** holcz*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
machent	machen	VVPS	ADJA	Neutr.Nom.Sg.

*der **gelyebte** sun gottes*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
gelyebte	lieben	VVPP	ADJA	Mask.Nom.Sg.

Liegt eine prädikativ verwendete Form im Partizip Präteritum vor, wurde wie folgt annotiert:

*dass es **bekannt** werde*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
bekannt	bekennen	VVPP	ADJD	*.*.*.*

Wenn die adjektivische Funktion des Partizips nicht eindeutig bestimmt werden kann, z.B. bei elliptischen Konstruktionen, wurde **VV > VVPP** (statt: **VVPP > ADJD**) getaggt, z.B.:

fBdelliu\-(,) **gemi\$chet mit nuchtern \$peych vnd darvß gemacht eyn pla\$ter vnd vff den buch gelyeyt vn=der den nabel(,) bricht den \$teyn in der bla\$en vnd in den lenden.(.)*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
*fBdelliu\-		FM	FM	
gemi\$chet	mischen	VV	VVPP	

IV.3.8.6 Negation mit *un-/en-*

Die satzverneinende Negationspartikel *en-/n(e)-* (vorrangig in Zeitraum I und II) wurde in der Prädication vom finiten Verb abgetrennt und als **PTK** (Lemma-Tag) und **PTKNEG** (POS-Tag) getaggt, z.B.:

di en|salt du nit schouwen ‚die sollst du nicht anschauen‘¹⁴

Zur Negation des Partizip Präteritums mit *un-* s. IV.3.1.2

IV.4 Verschmelzungsformen aus Präposition und Artikel

Verschmelzungsformen aus Präposition und Artikel (z.B. *am*, *zum*, *zur*, *beim*) wurden i.d.R. als neue Lemmata angesetzt und im Lemma-Tag mit **AP** und im POS-Tag mit **APPRDDART** gekennzeichnet, z.B.:

Dw \$olt alle ding zum pe\$ten ker|-n.

IV.5 Zahlen

Zahlen wurden im Lemma-Tag als **CARD**, im POS-Tag je nach Position als **CARDx** annotiert. Eine morphologische Annotation erfolgt i.d.R. nur bei ausgeschriebenen Zahlen (bspw. *sieben*).

IV.6 Zirkumpositionen

Bei Zirkumpositionen, wie z.B. *durch/um ... willen/wegen*, wurde der erste Bestandteil im POS-Tag mit **APPR**, der zweite Bestandteil mit **APZR** annotiert. Die Elemente, die etymologisch auf Substantive zurückzuführen sind (*willen/wegen*), wurden nicht mehr als **NA** getaggt, z.B.:

**(D*)arna euer zo eyne|- zijden/ vmb des gro\$\$en(,) verderflichen ha\$\$es vnd n<ij>tz w<i>lle(,)*

Token	Lemma	Lemma-Tag	POS-Tag
um...	um	AP	APPR
... wegen	wegen	AP	APZR
... willen	willen	AP	APZR

Liegen hingegen keine ‚echten‘ Zirkumpositionen (z.B. *um ... herum*, *an ... entlang*) vor (der Zweitbestandteil ist etymologisch nicht auf ein Substantiv zurückzuführen), wurde der zweite Bestandteil als **AVD** annotiert.

Token	Lemma	Lemma-Tag	POS-Tag
um...	um	AP	APPR
... herum	herum	AVD	AVD

¹⁴ Mhd. Bsp.; Dipper u.a. (2013), 130.

V. Literatur

ANNIS, Version 3.4.4, Vaadin-Version: 7.6.6. Online unter: <http://corpus-tools.org/annis/>. Suchmaske ANNIS online abrufbar unter: <https://linguistics.rub.de/annis/annis3/REF/>.

Bollmann, Marcel/Petran, Florian/Dipper, Stefanie/Krasselt, Julia (2014): CorA: A web-based annotation tool for historical and other non-standard language data. In: Proceedings of the EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH). Gothenburg, Sweden, 86-90.

Bonner Frühneuhochdeutschkorpus = Bernhard Schröder/Klaus-Peter Wegera/Hans-Joachim Solms/Hans-Christian Schmitz/Bernhard Fisseni: Das Bonner Frühneuhochdeutschkorpus (FnhdC). Korpora.org. Online abrufbar unter: <http://korpora.zim.uni-due.de/Fnhd/>.

Capelli, Adriano (1928): Wörterbuch lateinischer und italienischer Abkürzungen, wie sie in Urkunden und Handschriften besonders des Mittelalters gebräuchlich sind; dargestellt in über 14000 Zeichen, nebst einer Abhandlung über die mittelalterliche Kurzschrift, einer Zusammenstellung epigraphischer Sigel, der alten römischen und arabischen Zählung und der Zeichen für Münzen, Masse und Gewichte. 2. Aufl. Leipzig.

Dipper, Stefanie/Donhauser, Karin/Klein, Thomas/Linde, Sonja/Müller, Stefan/Wegera, Klaus-Peter (2013): HiTS: ein Tagset für historische Sprachstufen des Deutschen. In: Journal of Language, Technology and Computational Linguistics 28 (1), 85-137.

DWB = Deutsches Wörterbuch von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden. Leipzig 1854-1961. Quellenverzeichnis Leipzig 1971. Digitalisierte Fassung im Wörterbuchnetz des Trier Center for Digital Humanities, Version 01/21. Online abrufbar unter: <https://www.woerterbuchnetz.de/DWB>.

Frnhd. Gr. = Frühneuhochdeutsche Grammatik. Hrsg. Oskar Reichmann und Klaus-Peter Wegera. Bearb. von Robert P. Ebert, Oskar Reichmann, Hans-Joachim Solms u. Klaus-Peter Wegera. Tübingen (Sammlung kurzer Grammatiken germanischer Dialekte. A. Hauptreihe Nr. 12).

Gr. d. Frnhd. = Grammatik des Frühneuhochdeutschen. Beiträge zur Laut- und Formenlehre hg. v. Hugo Moser u. Hugo Stopp [ab Bd. III: hg. v. Hugo Moser, Hugo Stopp u. Werner Besch]. I.1: Vokalismus der Nebensilben I, bearbeitet von K. O. Sauerbeck. Heidelberg 1970. I.2: Vokalismus der Nebensilben II (Die Entsprechungen von mhd. unbetontem *e*) unter Benutzung der Sammlung von K. O. Sauerbeck und weiteren Materials bearbeitet von H. Stopp. Heidelberg 1973; I.3: Vokalismus der Nebensilben III, bearbeitet von H. Stopp. Heidelberg 1978; III: Flexion der Substantive von K.-P. Wegera. Heidelberg 1987; IV: Flexion der starken und schwachen Verben von U. Dammers, W. Hoffmann, H.-J. Solms. Heidelberg 1988; VI: Flexion der Adjektive von H.-J. Solms, K.-P. Wegera. Heidelberg 1991; VII: Flexion der Pronomina und Numeralia von M. Walch, S. Häckel. Heidelberg 1988.

Herbers, Birgit (2016): „Referenzkorpus Deutsche Inschriften“ – Chancen und Grenzen der Auswertung. In: Kwekkeboom, Sarah/Waldenberger, Sandra (Hrsg.): PerspektivWechsel oder: Die Wiederentdeckung der Philologie. Bd. 1: Sprachdaten und Grundlagenforschung in der Historischen Linguistik. Berlin 2016, 27-41.

Lexen = Mittelhochdeutsches Handwörterbuch von Matthias Lexen, digitalisierte Fassung im Wörterbuchnetz des Trier Center for Digital Humanities, Version 01/21. Online abrufbar unter: <https://www.woerterbuchnetz.de/Lexen>.

Medek (*Gießler), André/Pöckelmann, Marcus/Bremer, Thomas/Solms, Hans-Joachim/Molitor, Paul/Ritter, Jörg (2015): Differenzanalyse komplexer Textvarianten – Diskussion und Werkzeuge. In: Datenbank-Spektrum 15, 25-31. Online abrufbar unter: <https://sada.uzi.uni-halle.de>, s. <http://dx.doi.org/10.1007/s13222-014-0173-y>.

VI. Anhang

VI.1 Eigennamen

Konventionen zur Annotation von Eigennamen in den Texten der Bochumer Arbeitsstelle sowie in ReDI:

VI.1.1 Vor- und Nachnamen

Bei Vornamen wurden im Morpho-Tag alle Informationen (Genus, Kasus, Numerus) angegeben.

Bei Nachnamen wurden im Morpho-Tag nur der Kasus und der Numerus (und nicht das Genus) angegeben. Zudem wurde ein neues Lemma angesetzt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
HEINRICH	heinrich	NE	NE	Mask.Nom.Sg.
WECKENESEL	weckenesel	NE	NE	*.Nom.Sg

Ausnahme: Bei ‚femininen‘ Nachnamen wurde i.d.R. das Genus mit angesetzt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Barbera	barbera	NE	NE	Fem.Nom.Sg.
Malšchitzin	malschitzin	NE	NE	Fem.Nom.Sg

VI.1.2 Fremdsprachliche Eigennamen

Lateinisch flektierte Namen (Personen-/Nach-/Ortsnamen) wurden insgesamt wie deutsche Namen behandelt und morphologisch annotiert. In REF wurden sie nicht mit *f markiert, während sie in REDI mit *f und als FM annotiert wurden.

Existiert eine deutsche Entsprechung zur belegten lateinischen Form, wurde ein deutschsprachiges Lemma angesetzt und es wurden die morphologischen Informationen annotiert.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Henricus	heinrich	NE	NE	Mask.Nom.Sg.
Henrico	heinrich	NE	NE	Mask.Dat.Sg.

Existiert kein deutsches Äquivalent zur lateinischen Form, wurde die lateinische Form als Lemma angesetzt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Quintus	Quintus	NE	NE	Mask.Nom.Sg.
Serapius	Serapius	NE	NE	Mask.Nom.Sg.

VI.1.3 Nomina sacra (Heiligennamen)

Nomina sacra wurden nach ihrer Belegform lemmatisiert, d.h. es wurde keine Anpassung an die deutsche Schreibung vorgenommen. Im Lemmaansatz wird daher zwischen ‚dem heiligen Petrus‘ und ‚dem heiligen Peter‘ differenziert (auch wenn es sich um dieselbe Person handelt).

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Petrus	Petrus	NE	NE	Mask.Nom.Sg.
Peter	Peter	NE	NE	Mask.Nom.Sg.
Paulus	Paulus	NE	NE	Mask.Nom.Sg.

VI.2 Heiligtage

Heiligtage werden als Komposita betrachtet und entsprechend in der Prädiction (durch #) zusammengezogen. Sie wurden mit NE getaggt.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Marien#tag	marientag	NE	NE	Mask.Nom.Sg.

VI.3 Ortsnamen

Bei Ortsnamen wurden wie bei Nachnamen nur die Numerus- und Kasusinformation annotiert. Sofern eine flektierte Form oder Verschmelzungsform (APPRDDART) Rückschlüsse auf das Genus des Ortsnamens zulässt, kann dieses angegeben werden.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Statt\$chreibers	stadtschreiber	NE	NA	Mask.Gen.Sg.
allhier	allhier	AVD	AVD	*
zur	zur	AP	APPRDDART	–
Eißen\$statt	Eisenstadt	NE	NE	Fem.Dat.Sg.

Bei (zusammengesetzten) Ortsnamen, die ohne Artikel stehen (z.B. *Hamburg*, *Wuppertal*, *Düsseldorf*), wurde das Genus hingegen nicht angesetzt, sondern wurde angesternt.

Das Lemma wird nach Editionsvorlage bzw. nach Beleg angesetzt; in einem Kommentar erfolgt der Hinweis auf einen Neuansatz.

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Virdensi	Virdensis	NE	NE	*.Gen. Sg.
Donauwörth	Donauwörth	NE	NE	*.Nom. Sg.

Ortsname als Apposition, z.B. *in dem Land Spanien*

Token	Lemma	Lemma-Tag	POS-Tag	Morpho-Tag
Spanien	Spanien	NE	NE	*.*.Sg.

VI.4 Annotation ‚Jahr‘ in Alters-/Zeitangaben in ReDI

mit Präposition		
Beispiel	Kasus/Numerus	Kommentar
ires alters bey 40 jarn	Dat.Pl.	
des 1645. jar	Gen.Sg.	
im jar 1631 ist ...	Dat.Sg.	
ist entslaffen im jar 1641	Dat.Sg.	
seines alters in 43. jar	Dat.Sg.	<i>in</i> als Fehler für <i>im</i> aufgefasst (Einzelfall: I003-075)
in den jaren unseres herren	Dat.Pl.	

ohne Präposition		
Beispiel	Kasus/Numerus	Kommentar
seines alters 53 jar	Akk.Pl.	Duden-Grammatik: §1245f.: reine Zeitangaben stehen grundsätzlich im Akk., ein adverbialer Akk. drückt ein Maß aus
seines alters ein jar	Akk.Sg.	
mit im gehaust 18 jar	Akk.Pl.	
aetatis 26 jar 5 monate	Akk.Pl.	
man schreibt 1552 jar	Akk.Pl.	
nach christi geburt 1284 jar gingen die kinder ...	Akk.Pl.	
als man zalt eintausent jar	Akk.Pl.	
haben si miteinander gelebt 36 jar	Akk.Pl.	
2 jar gewesener kirchendiener	Akk.Pl.	
1490 jar wart geslozzten x ...	Akk.Pl.	
ir alter ist 19 jar 5 monat	Nom.Pl.	
so sint es 80 jar und ...	Nom.Pl.	
hat machen lassen dises kreuz 1630 jar	.*.*	Kausus u. Numerus uneindeutig

pleonastische (?) oder formelhafte (?) Kombination mit <i>anno</i> (sehr häufig)		
Beispiel	Kasus Numerus	Kommentar
anno domini 1490 jar uf Montag	.*.*	Kasus unklar, denkbar ist ‚im xx Jahr‘ (Dat.Sg.), ‚nach xx Jahren‘ (Akk.Pl.) oder ‚das x Jahr‘ (Nom.Sg.); die Kombination mit vorangestelltem <i>anno</i> ist uneindeutig, daher in diesen Fällen Kasus und Numerus ausgesternt
gestorben anno 1639 jar	.*.*	
got hat dis kreuz machen lassen anno 1625 jar	.*.*	

VI.5 Header-Beispiele

VI.5.1 Referenzkorpus Frühneuhochdeutsch

+H

Korpus-Sigle: F165, III-Ca-V1, D

Titel: Historij des beleegs van Nuys

Verfasser: Christian Wierstrait (Wierstraat)

Textart: V1 (D)

Textsorte: CB

Zuordnungsqualität: Z2

Hoffmann/Wetter-Nr: HW 1455, ZK(W)

Bibliothek/Archiv: Ex. Cambridge/Mass., Houghton Libr. (ehemals
Brüssel), u. Düsseldorf, UB, D. Sp. G.?434 (defekt)

Signatur:

Auswahl Referenzkorpus: komplett (ohne handschriftlichem Vorwort)

Umfang in Wortformen (Referenzkorpus): 18.004

Datierung: 1476

Lokalisierung: Ca Ripuarisch

Druckort: Köln

Drucker: Arnold Terhoernen

verwendete Edition (für Sek.-Zählung): Reimchronik der Stadt Neuß aus
der Zeit der Belagerung durch Herzog Karl den Kühnen von Burgund. Nach
dem Originaldruck von 1476 unter Berücksichtigung der Ausgabe von 1497
und 1564 hrsg. v. Karl Meisen. Bonn, Leipzig 1926. VIII, 204 S.

Literatur: Vgl. Frieder Schanze, in: 2VL 10 (1999), Sp. 1055-1058 + 2VL
11 (2004), Sp. 1658.; Christian Wierstraet: Die Geschichte der
Belagerung

von Neuss. Faksimile der Erstausgabe bei Arnold ther Hoernen, Köln
1476.

Übertragung und Einleitung Herbert Kolb.

Sonstiges:

- * Sekundärzählung nach Meisen (Verse)
- * Majuskeln, Satzmajuskeln und überhöhte Minuskeln sind rubriziert.
- * Text weist eine eingetragene Lagenzählung auf.
- * \&1 = Kürzung nach \$, die mit \$ verschmolzen ist (zumeist 'er'
kürzend)

Text eingegeben nach Edition/Handschrift/Druck:

Datum: XXX

Bearbeiter: XXX

Text kollationiert (zum ersten Mal):

Datum: 02.05.2013 bis 14.05.2013

Bearbeiter: Div.

Text kollationiert (zum zweiten Mal):

Datum:

Bearbeiter:

Text präeditiert

Datum: XXX

Bearbeiter: XXX

Lat. Passage geprüft:

Datum: XXX

Bearbeiter: XXX

Annotiert:
Datum:
Bearbeiter:
KTX-Korrektur:
Datum:
Bearbeiter:
@H

VI.5.2 Referenzkorpus Deutsche Inschriften

VI.5.2.1 Header für den Gesamtband

+H
Korpus-Sigle: I029
Edition: Die Inschriften der Stadt Worms. Gesammelt und bearbeitet von
Rüdiger Fuchs. Wiesbaden 1991. (DI 29)

Besonderheiten:
1. Allg.:
2. Einzelne Inschriften:
3. Sonderzeichen:

Bearbeitung:
Text eingegeben nach Edition:
Datum: 13.07.2015 -|
Bearbeiter: Katharina Westphal
Text kollationiert (zum ersten Mal):
Datum:
Bearbeiter:
Text kollationiert (zum zweiten Mal):
Datum:
Bearbeiter:
Text präeditiert:
Datum: Januar 2016
Bearbeiter: B. Herbers
Lat. Passage geprüft:
Datum:
Bearbeiter:
Annotiert:
Datum:
Bearbeiter:
Annotations-Korrektur:
Datum:
Bearbeiter:
@H

VI.5.2.2 Header für jede einzelne Inschrift

ID:	<input type="text" value="I029-0525A"/>
Ort:	<input type="text" value="Stadtmuseum, vom lutherischen Friedhof"/>
(Überlieferungs-) Träger:	<input type="text" value="Grabstein der Otilia Demerten geb. Grun"/>
Beschreibstoff:	<input type="text" value="hellgelber Sandstein"/>
Textart:	<input type="text" value="P"/>
Textbereich:	<input type="text" value="Alltag"/>
Edition:	<input type="text" value="DI029"/>
Besonderheiten:	<input type="text"/>

hochdeutsch, niederdeutsch	<input type="text" value="hochdeutsch"/>
Lokalisierung:	<input type="text" value="Worms"/>
Zeit (Jahrhunderthälfte):	<input type="text" value="16,2"/>
Zeit (genaue Datierung):	<input type="text" value="1583"/>
Inskript abgeglichen mit:	<input type="text" value="Abbildung"/>
Verortung der Abbildung:	<input type="text" value="online (DIO)"/>
Kommentar zur Abbildung:	<input type="text"/>
DIO URN:	<input type="text" value="http://dnb.resolving.de/di029mz02k0052508"/>
DIO-URL:	<input type="text" value="http://www.inschriften.net/worms/inschrift/nr/di029-0525.html#content"/>

VI.6 Tagset

Lemma-Tag	Erklärung	POS-Tag	Erklärung	Beispiele
ADJ	Adjektiv	ADJA	attributiv, vorangestellt	<i>der höchst vnd beste artzt</i>
		ADJN	attributiv, nachgestellt	<i>Drey wörtlein klein</i>
		ADJD	prädikativ	<i>der ist vnsinnig vnnnd vndanckbar</i>
		ADJS	substituierend (stellvertretend, allein stehend, ohne Substantiv, kann substantiviert sein)	<i>mitt der alten; die weisen lerent; mag ze deutsch haïzzen</i>
		AVD	adverbial	<i>fleissig einander helffen</i>
AP	Adposition	APPR	Präposition	<i>auf den henden</i>
		APPO	Postposition	<i>der gebüre nach; des kaisers halben</i>
		APZR	Zirkumposition	<i><u>durich</u> deiner sünden willen</i>
		PAVAP	präpositionaler Teil des Pronominaladverbs in Distanzstellung	<i><u>da</u> sint sî gar girig nach</i>
		APPRDDART	Präposition + Artikel	<i>am, vom, zur</i>
AVD	Adverb	AVD	Adverb	<i>von hinden oder von vorn</i>
		AVNEG	negativ	<i>nie</i>
		AVREL	relativ	<i>durch die(,) So fîch der an nemen</i>

		PTKVZ	Partikelverbzusatz	<i>zeuch den zirkel auf (aufziehen)</i>
		PAVD	pronominaler Teil des Pronominaladverbs	<i>da sint sî gar girig <u>nach</u></i>
		PAVREL	relativischer Teil eines Pronominaladverbs in Distanzstellung	<i>wasser/(,) da sye <u>In</u> gefottn werden</i>
AVDAP	Präpositionaladverb/ Pronominaladverb, nicht in Distanzstellung	PAVDAP	Pronominaladverb (da(r), hie(r) + Präp.)	<i>dafür, hiemit</i>
		PAVRELAP	relativisch	<i>zaichen(.) dar#an man sih bar(.) dar#vber die Capellen gestandeen ist</i>
		PAVWAP	interrogativ	<i>warmit, warum, worin</i>
AVG	generalisierend; nur wenn s-Markierung vorhanden	AVG		<i>swie, swenne, swan</i>
		PAVG	pronominaler Teil des Pronominaladverbs in Distanzstellung, generalisierend	<i>swarumbe</i>
AVW	Interrogativadverb	AVW		<i>wie froh</i>
		PAVW	pronominaler Teil des Pronominaladverbs in Distanzstellung, interrogativ	<i>war umbe</i>
CARD	Kardinal- und Ordinalzah-	CARDA	attributiv, vorangestellt	<i>Die zehen gepot; Diu dritt chraft</i>

	len	CARDN	attributiv, nachgestellt	<i>der kloster drey</i>
		CARDD	prädikativ	<i>der warn drey</i>
		CARDS	substituierend (stellvertretend, alleine stehend, ohne Substantiv, kann substantiviert sein)	<i>Der erste haist; zum dritten</i>
DD	Determinativum, definit bzw. demonstrativ	DDART	bestimmter Artikel, vorangestellt	<i>der hals; die chraft, daz hirn</i>
		DDA	attributiv, vorangestellt	<i>diese Rede; das#selb brieflein</i>
		DDN	nachgestellt	<i>frawen diese; an im selber</i>
		DDS	substituierend (stellvertretend, alleine stehend, ohne Substantiv, kann substantiviert sein)	<i>Daz ist daz puoch; vnd dicz sehen sie</i>
		DRELS	relativisch, substituierend	<i>der Pfeil, den ich gezogen habe; schaden(,) der dir geschicht</i>
DG	Determinativum, generalisierend (nur bei s-Formen)	DGA	attributiv, vorangestellt	<i>vernehmt, swelch rat euch besser gefällt</i>
		DGS	substituierend (stellvertretend, alleine stehend, ohne Substantiv, kann substantiviert sein)	<i>vernehmt, swelch euch besser gefällt</i>
DI	Determinativum, indefinit (bei Indefinitpron. hier solche, die attributiv ver-	DIART	unbestimmter Artikel, vorangestellt	<i>aIn stuck; zuo einer zeügnüß</i>
		DIA	attributiv, vorangestellt	<i>alle Jahr; an ieder hant; einigerley vrsach</i>
		DIN	nachgestellt	<i>unter den Königen allen; vn jn beyden</i>

	wendet werden können: sonst PI!)	DIS	substituierend (stellvertretend, allein stehend, ohne Substantiv, kann substantiviert sein)	<i>daz ez allez von got jst; von dem andrn</i>
		DNEGA	negativ, attributiv, vorangestellt	<i>keine Angst; zw chainer zeit</i>
		DNEGN	negativ, nachgestellt	<i>er gab Antwort keine;</i>
		DNEGS	negativ, substituierend (stellvertretend, allein stehend, ohne Substantiv, kann substantiviert sein)	<i>kainer pis an ainen; vnnd kainem sein Traid schütten</i>
DPOS	Possessivum	DPOSA	attributiv, vorangestellt	<i>mein haus; sein har;</i>
		DPOSD	prädikativ	<i>der wil ist mein; das nit sein i\$st</i>
		DPOSN	nachgestellt	<i>dancht der marter sein; lieber herre mein</i>
		DPOSS	substituierend (stellvertretend, allein stehend, ohne Substantiv, kann substantiviert sein)	<i>daz du daz dein dar#zu dust; das haar: seines ist rot</i>
DW	Determinativum, interrogativ	DWA	attributiv, vorangestellt	<i>in welher sprach; welche swester</i>
		DWS	substituierend (stellvertretend, allein stehend, ohne Substantiv, kann substantiviert sein)	<i>welches gleicher#weise gewonne; welch?</i>
FM	Fremdsprachliches Material	FM		<i>*fItem</i>

ITJ	Interjektion	ITJ	Interjektion	<i>oh, eya</i>
KO	Junktion	KON	nebenordnende Konjunktion	<i>vnd, oder, denn/dann, sonder(n), wann 'denn'</i>
		KOUS	unterordnende Subjunktion	<i>daz er beleiben muoz; weil mann sonsten... hat</i>
		KOUI	unterordnende Konjunktion mit Infinitiv	<i>um ze leben</i>
		KO*	neben- oder unterordnend (nicht entscheidbar)	<i>wann si ist in dem#selben czaichen (konditional und V2)</i>
		KOKOM	vergleichende Konjunktion	<i>wie ein grimiger leo; mit worten als da; nit pesser den dein herr; Mer dann drevvndreizzig iar</i>
NA	Substantiv, Nomen appellativum	NA	Nomen appellativum	<i>mensch, arbeit, blut</i>
		PI	bei Indefinitpronomen <i>man</i> (abgeleitet aus: Substantiv <i>Mann</i>)	<i>man</i>
NE	Eigename	NE	Eigename	<i>paul, regenspurg, ze rome</i>
PG	Pronomen generalisierend	PG	generalisierend	<i>swer (swas, swes)</i>
PI	Pronomen, indefinit (PI, wenn nicht attributiv verwendbar)	PI	indefinit	<i>so sev iemand straft; daz es yederman sech; ab ich icht 'irgendetwas' guttes funde</i>
		PTKNEG	Negationspartikel <i>nicht</i> (entstanden aus mhd. <i>ne + iht</i>)	<i>nicht</i> u.a. mit Formen: <i>ni(c)ht, nit(t), nüt(t), neüt;</i>

				<i>daz sie dez gesanges iht ('nicht') hörn; niht schouwen</i>
	Pronomen, indefinit, negativ	PNEG	negativ	<i>das er nichts wolt, niemand acht dein;</i>
PPER	Personalpronomen	PPER	Personalpronomen	<i>ich, du, er (+ sie, es), wir, ihr, sie</i>
		PRF	reflexiv gebrauchtes Personalpronomen	<i>solt du dich got genczlichen geben; bezevgt er im 'sich' selber; ist si vnuerhauwen an ir Selben einander</i>
PRF	Reflexivpronomen	PRF	Reflexivpronomen	<i>dabei got sich let erchennen</i>
PTK	Partikel	PTKA	Partikel bei Adjektiv oder Adverb	<i>zu lang, czu viel, ze mittelst</i>
		PTKANT	Antwortpartikel	<i>nein, ja, doch, jawohl in Antworten</i>
		PTKINT	Fragepartikel	-
		PTKNEG	Negationspartikel (nur Satznegation!)	<i>daz en vindet man niht (ne, un)</i>
		PTKREL	Relativpartikel	-
		PTKZU	bei Infinitiv mit zu	<i>fur genomen hab zu gedencken</i>
PW	Pronomen, interrogativ	PW	interrogativ	<i>wer (wes(sen), wem, wen, was(?))</i>
		PI	indefinit	<i>irgendwer (irgendwas, irgendwem)</i>

SPELL	Buchstabierfolge	SPELL		<i>mei\$ter schreybt man so: MEISTER</i>
SYM	Symbol	SYM		z.B. Kreuz, Paragraphenzeichen, *C
TRUNC	Platzhalter	TRUNC	z.B. unvollständiges Erstglied bei ADJ und NA, z.T. mit Bindestrich	<i>ein vnd au\$ganck, er- vnd tugent\$am</i>
UNK	nicht mehr erschießbare oder nachvolziehbare Wörter	UNK	Wörter oder Abkürzungen	<i>snatra snatra, M. (für Meister oder Magister)</i>
VA	Verb, auxiliar	VAFIN	finite Verbform	<i>si habent einen vlizzygen aneseher;daz <u>ist</u> ein zaichen; derselb kuing wardt getauft;</i>
		VAIMP	Imperativ	<i>habt guetten fleyss; Seyt mänlich; biß gelobt;</i>
		VAINF	Infinitiv	<i>haben, sein, wesen, werden (auch wenn Kopula)</i>
		VAPP	Partizip Präteritum im Verbalkomplex	<i>das er gehabt hat; sie ist gewesen; gestolen ist worden;</i>
		VAPS	Partizip Präsens im Verbalkomplex	-
VM	Verb, modal	VMFIN	finite Verbform	<i>von ir geferttign chunde; mag ain iuncfraw haizzent; Nit allein muoft ir auch leiden; daz sie daz behüten wellent</i>
		VMINF	Infinitiv	<i>dürfen, können, mögen, müssen, sollen, wollen</i>
		VMPP	Partizip Präteritum im Verbalkomplex	<i>hiet er gewolt;</i>

			plex	
		VMPS	Partizip Präsens im Verbalkomplex	-
VV	Vollverb	VVFIN	finite Verbform	<i>als aristotiles Spricht; so stürb der mensch</i>
		VVIMP	Imperativ	<i>mach daz prunnwazzer chalt; bittet vor mich;</i>
		VVINFINF	Infinitiv	<i>blicken, gehen, trincken</i>
		VVPP	Partizip Präteritum im Verbalkomplex	<i>geschwollen ist; sol es auch gehalten werden</i>
		VVPS	Partizip Präsens im Verbalkomplex	<i>dy speyse betreffende</i>
VAINF	Verb, infinitiv, auxiliar	NA	substantivierte Formen	<i>in bey sein dessen, der ...;</i>
VMINF	Verb, infinitiv, modal	NA	substantivierte Formen	<i>das Müssen und Sollen; das wöllen vnd thuon</i>
VVINFINF	Vollverb, infinitiv	NA	substantivierte Formen	<i>mit Seufzen und mit Weinen; zw dem fechten</i>
VVPP	Partizip Präteritum im Verbalkomplex, Vollverb	ADJA	adjektivisch, vorangestellt	<i>mit eynem gesaltzem wasser; dy ye genanten fürstn;</i>
		ADJN	adjektivisch, nachgestellt	<i>lilienwa\$\$er geprant</i>
		ADJD	prädikativ	<i>dass es bekannt werde; so leit doch <u>verporgen</u> di warhait;</i>
		ADJS	substituierend	<i>zeichen miner vz erwelten; die gefangnen</i>
VVPS (VAPS, VMPS)	Partizip Präsens im Verbalkomplex, Vollverb	ADJA	adjektivisch, vorangestellt	<i>vz dem lebenden brunnen; nach folgenter vrfachen halben</i>
		ADJN	adjektivisch, nachgestellt	-

		ADJD	prädikativ	<i>die do nit sehen(,) sehndt werden; vnd fant fy schlaf-fent</i>
		ADJS	substituierend	<i>die erst tragenden; von den essenden; wenn yr Synt nit all glich mugende (hier VMPS > ADJS)</i>

Sonstige Tags in der Prädeditio

XY	<i>Alle Formen von Textausfall/Auslassungen im Satz, restl. Felder bleiben leer</i>	<i>[...], ver[...]</i>	XY
\$_	<i>originale Interpunktion [.]</i>	<i>[. : , / ? !]</i>	\$_
\$(<i>sonstige Satzzeichen [(.)]</i>	<i>[() ' “ ”]</i>	\$(
\$QL	<i>Durch Präditor eingefügte Anführungszeichen links (eröffnend)</i>	<i>Der Herr sprach: [(“)]gloubet mir.(“)</i>	\$(
\$QR	<i>Durch Präditor eingefügte Anführungszeichen rechts (schließend)</i>	<i>Der Herr sprach: (“)gloubet mir.[(“)]</i>	\$(