

**Bochumer
Linguistische
Arbeitsberichte
15**



**Guidelines für die Normalisierung historischer deutscher Texte
Guidelines for Normalizing Historical German Texts**

**Julia Krasselt, Marcel Bollmann, Stefanie Dipper,
Florian Petran**

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 15 (April 2015)

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

Erscheinungsjahr 2015
ISSN **2190-0949**

**Julia Krasselt, Marcel Bollmann, Stefanie Dipper,
Florian Petran**

**Guidelines für die Normalisierung
historischer deutscher Texte**

**Guidelines for Normalizing Historical
German Texts**

2015

Bochumer Linguistische Arbeitsberichte

(BLA 15)

Inhaltsverzeichnis

I	Guidelines (DE)	4
1	Einleitung	4
2	Guidelines	5
2.1	Allgemeiner Aufbau	5
2.2	Spalte <i>Norm</i> : Normalisierung	5
2.3	Spalte <i>Mod</i> : Modernisierung	6
2.3.1	Abweichungen in der Bedeutung (<i>sem</i>)	7
2.3.2	Abweichungen in der Flexion (<i>infl</i>)	8
2.3.3	Extinkte Wortformen (<i>ext</i>)	12
2.4	Kontextabhängige Normalisierung	14
II	Guidelines (EN)	16
1	Introduction	16
2	Guidelines	17
2.1	General Description	17
2.2	The <i>Norm</i> Column: Normalization	17
2.3	The <i>Mod</i> Column: Modernization	18
2.3.1	Semantic Divergence (<i>sem</i>)	18
2.3.2	Inflectional Variation (<i>infl</i>)	19
2.4	Extinct Word Forms (<i>ext</i>)	24
2.5	Context-Dependent Normalization	25
	Literatur	26
	Appendizes: Beispiele – Appendices: Examples	27
A	Fälle ohne weitere Anpassung – Normal Cases	27
B	Semantik-Anpassungen – Semantic Adjustments	30
C	Flexionsanpassungen – Inflectional Adjustments	33
D	Extinkte Wortformen – Extinct Word Forms	36

Teil I

Guidelines (DE)

1 Einleitung

Historische Sprachdaten des Deutschen zeichnen sich durch einen hohen Grad an Variation aus, die insbesondere die Ebenen Phonologie und Graphematik betrifft, aber auch in den Bereichen Morphologie, Syntax und Lexik zu beobachten ist.¹ Diese Variation ergibt sich vor allem daraus, dass das Deutsche in den historisch belegten Sprachstufen Althochdeutsch (AHD), Mittelhochdeutsch (MHD) und Frühneuhochdeutsch (FNHD) nicht über eine allgemein verbindliche, übergeordnete Standardsprache verfügte, sondern von einer hohen dialektalen Diversivität geprägt war.

Insbesondere die Variation in Phonologie und Graphematik stellt ein enormes Hindernis bei der automatischen Verarbeitung solcher Daten dar. Existierende Tools, beispielsweise im Bereich der automatischen Wortart-Annotation, sind auf Daten des gegenwärtigen Standarddeutsch trainiert und können dementsprechend gar nicht oder nur bedingt auf Daten älterer Sprachstufen angewandt werden. Auch die Suche nach spezifischen Daten aus einem historischen Korpus wird durch den hohen Grad an Variabilität erschwert, da einzelne Wortformen nicht gezielt abgefragt werden können, ohne beispielsweise alle potentiell möglichen Schreibweisen zu kennen. Eine automatische Lemmatisierung aller Wortformen ist ebenfalls deutlich erschwert.

Um solche Daten dennoch automatisch analysierbar und abfragbar zu machen, ist es sinnvoll, die historischen Wortformen auf einheitliche Entsprechungen abzubilden, d.h. zu *normalisieren*. Als Referenz für diese Entsprechungen kann das moderne Standarddeutsch bzw. ein entsprechendes Wörterbuch dienen, z.B. der Duden (<http://www.duden.de/woerterbuch>). Eine Normalisierung auf das moderne Standarddeutsch ist insbesondere dann sinnvoll, wenn eine Anwendung bestehender Sprachverarbeitungstools geplant ist. Selbstverständlich ist aber beispielsweise auch eine Abbildung auf eine frühere Sprachstufe denkbar (bspw. bei einem mhd. Korpus, das auf ein vereinheitlichtes MHD abgebildet werden soll). Entscheidend ist, dass eine klar definierte Referenz zur Verfügung steht. Im Falle des MHD kann das beispielweise ein Referenzwörterbuch wie der Lexer (<http://woerterbuchnetz.de/Lexer>) sein.

Im folgenden Artikel werden Guidelines vorgestellt, mit dessen Hilfe historische deutsche Sprachdaten manuell normalisiert werden können. Dazu werden alle Wortformen auf ein modernes, standardsprachliches Äquivalent abgebildet. Die Guidelines sind im Rahmen des Projektes „St. Anselmi Fragen an Maria. (Digitale) Erschließung,

¹Wir danken Simone Schultz-Balluff, Malin Frey, Julika Nelken, Cornelia Johnen, Maurice Spengler, Katharina Bort, Nadine Lordick und Maria Anselm für die manuelle Normalisierung der Daten und ihre wertvollen Kommentare bei der Erstellung dieser Guidelines.

Auswertung und Edition der gesamten deutschsprachigen Überlieferung (14.–16. Jahrhundert)² entstanden (beschrieben bspw. in Schultz-Balluff and Dipper (2013)). Alle verwendeten Textbeispiele stammen aus dem Korpus, das in diesem Projekt erstellt wurde. Im Rahmen dieses Projektes wurden die manuell normalisierten Texte des Korpus verwendet, um ein eigens entwickeltes Tool zur automatischen Normalisierung zu trainieren und evaluieren (*Norma*, beschrieben bspw. in Bollmann et al. (2012)).

2 Guidelines

2.1 Allgemeiner Aufbau

Eine vollständig normalisierter Text besteht aus insgesamt vier Spalten (vgl. Tab. 1):

- 1. Spalte (*Orig*): das Originaltoken
- 2. Spalte (*Norm*): das Token in einer *normalisierten* Form
- 3. Spalte (*Mod*): das Token in einer *modernisierten* Form
- 4. Spalte (*Art der Anpassung*): eine der Abkürzungen *infl*, *sem* oder *ext*, die den Unterschied zwischen *normalisierter* und *modernisierter* Form näher erklärt (*infl* = Flexion, *sem* = Semantik, *ext* = extinkte Wortform)

Orig	Norm	Mod	Art der Anpassung
alz	als		
mein	mein		
chind	Kind		
di	die		
mandat	mandate	Fußwaschung	ext
begangen	begangen	vollzogen	sem
het	hat	hatte	infl

Tabelle 1: Tabellarischer Aufbau einer Normalisierung

2.2 Spalte *Norm*: Normalisierung

Diese Spalte ist obligatorisch auszufüllen. Sie enthält eine im modernen Standarddeutsch vorhandene³ Entsprechung des Originaltokens (Spalte 1). Diese Entsprechung darf **ausschließlich durch phonologische und/oder graphematische Entsprechungsoperationen** entstehen. Die Wortart des Originaltokens wird bei der Normalisierung beibehalten.

²Dieses Projekt wird finanziert durch die Deutsche Forschungsgemeinschaft, Projekt DI 1558/4.

³Bei Unsicherheiten sollte der Duden (<http://www.duden.de/hilfe/rechtschreibung>) benutzt werden; ist das Token hier enthalten, kann es verwendet werden.

- phonologische Operationen, bspw.:
 - Diphthongierung: frnhd. /u/ > nhd. /au/ (*hus* – *Haus*, *mus* – *Maus*); frnhd. /i/ > nhd. /ei/ (*min* – *mein*, *schriben* – *schreiben*)
 - qualitativer Diphthongwandel: frnhd. /ou/ > nhd. /au/ (*boum* – *Baum*, *vrouwe* – *Frau*)
 - Vokalsenkung: frnhd. /ii/ > nhd. /ö/ (*küneg* – *König*); frnhd. /u/ > nhd. /o/ (*sunne* – *Sonne*)
 - Vokalrundung: frnhd. /e/ > nhd. /ö/ (*zwelf* – *zwölf*); frnhd. /i/ > nhd. /ü/ (*flistern* – *flüstern*)
 - Palatalisierung von /s/ (frnhd. *slange* – nhd. *Schlange*)
 - Epenthese (frnhd. *eigenlich* – nhd. *eigentlich*, frnhd. *namenlich* – nhd. *namentlich*)
 - Kontraktionen (frnhd. *er seit* – nhd. *er sagte*)
- graphematische Operationen: $t = d$ (*kint* – *Kind*); $cz = z$ (*czu* – *zu*); $ai = ei$ (*gewonhait* – *Gewohnheit*); $i = j$ (*iare* – *Jahre*)

Die so entstandene Wortform wird in die Spalte *Norm* eingetragen (vgl. Tab. 2). Darüber hinaus müssen mit dieser Wortform zwei weitere Tests durchgeführt werden, die in den folgenden Abschnitten erläutert werden.

Orig	Norm	Mod	Art der Anpassung
lybe	liebe		
frouwe	Frau		
sage	sage		
myr	mir		
dynes	deines		
kyndes	Kindes		
marter	Marter		

Tabelle 2: Normalisierung (Bsp. B1, 262–269)

2.3 Spalte *Mod*: Modernisierung

Während die Spalte *Norm* obligatorisch auszufüllen ist, werden die Spalten *Mod* und *Art der Anpassung* nur unter bestimmten Bedingungen verwendet. Sie sind für diejenigen Fälle reserviert, bei denen die normalisierten Formen in *Norm* keinen sinnvollen modernen Satz ergeben, z.B. aufgrund von Abweichungen in Flexion und Bedeutung. Gibt es keine Abweichungen dieser Art, bleiben die Spalten *Mod* und *Art der Anpassung* leer (vgl. Tab. 2).

2.3.1 Abweichungen in der Bedeutung (*sem*)

In einigen Fällen existiert das in der Spalte *Norm* durch phonologische und graphematische Entsprechungsoperationen entstandene Token zwar nach wie vor im Standarddeutschen, jedoch mit einer vom Originaltoken stark abweichenden Bedeutung (bspw. mhd./fnhd. *Urlaub* i.S.v. *Erlaubnis*). In diesen Fällen wird in die Spalte *Mod* ein semantisch passenderes Äquivalent eingetragen; Spalte 4 wird mit dem Kürzel *sem* ausgefüllt (dadurch wird angezeigt, dass es sich in der Spalte *Mod* um eine Modernisierung auf der Ebene der Semantik handelt).

Achtung: Es geht bei diesem Schritt nicht darum, den Originaltext bspw. nach den Anforderungen einer Edition o.ä. zu bearbeiten ('zu übersetzen'). Anpassungen auf Ebene der Semantik sollten nur dann vorgenommen werden, wenn Originalform und moderne Form sehr stark voneinander abweichen.

Orig	Norm	Mod	Art der Anpassung
petrus	petrus		
behilt	behielt	versteckte	sem
sich	sich		
vnder	unter	hinter	sem
eyme	einem		
steyne	stein		

Tabelle 3: Modernisierung: Abweichungen in der Semantik (Bsp. B1, 1411–1418)

Tab. 3 zeigt zwei Beispiele für typische Anpassungen auf Ebene der Semantik: *sich unter einem Stein behalten* ist heute nicht mehr verständlich, so dass eine Anpassung zu *sich hinter einem Stein verstecken* angebracht ist.

Bedeutungsanpassungen können zur Folge haben, dass ein benachbartes Token ebenfalls angepasst werden muss – beispielsweise immer dann, wenn sich das Originaltoken und das Token mit der angepassten Bedeutung im Genus unterscheiden (vgl. Tab. 12).

Besondere Regelungen:

- Sehr veraltete Wörter:

Ist ein Wort zwar noch mit derselben Bedeutung im Duden vorhanden, gilt aber als sehr veraltet und wird heutzutage kaum noch benutzt, sodass es unverständlich erscheint (bspw. *Muhme* für *Tante*, *Schächer* für *Mörder*, *dünken* für *denken*), sollte ein moderneres Wort in die Spalte *Mod* eingetragen werden. Die Art der Anpassung wird mit dem Kürzel *sem* markiert, also als Anpassung auf der Ebene der Semantik.

2.3.2 Abweichungen in der Flexion (*infl*)

Es ist ein generelles Ziel der Normalisierung, morphologisch und syntaktisch korrekte Sätze zu erzeugen (mit der Einschränkung, die Wortstellung nicht zu verändern). In vielen Fällen bedeutet das, dass die in der Spalte *Norm* durch phonologische und/oder graphematische Entsprechungsoperationen entstandene Wortform nicht in der Flexionsform erscheint, in der sie nach der heutigen Standard-Grammatik stehen müsste. Die Gründe dafür können vielfältig sein (bspw. abweichende Flexionsklassen in der historischen Sprachstufe, abweichende Kasusrektion, graphematische Verkürzungen)⁴. Deswegen muss in solchen Fällen eine morphologisch und syntaktisch angepasste Form zusätzlich in Spalte *Mod* eingetragen werden; Spalte 4 wird mit dem Kürzel *infl* ausgefüllt (dadurch wird angezeigt, dass es sich in Spalte *Mod* um eine Modernisierung auf der Ebene von Morphologie/Syntax handelt).

Orig	Norm	Mod	Art der Anpassung
Dese	diese		
wort	wort	worte	infl
spricht	spricht		
vnser	unser		
liber	lieber		
here	herr		
ihesus	jesus		
cristus	christus		

Tabelle 4: Modernisierung durch abweichende Pluralmarkierung (Bsp. B1, 10–17)

Tab. 4 gibt ein Beispiel für eine Modernisierung. Das Originaltoken *wort* existiert in dieser *Oberflächenform* im modernen Standarddeutsch (einzig die Großschreibung weicht ab). Es handelt sich beim Originaltoken jedoch um eine Pluralform (erkennbar beispielsweise am Determinierer *dese*). Diese Art der Pluralbildung ist für das Lemma *Wort* heute nicht mehr möglich. In Spalte *Mod* muss daher zusätzlich die heutige Pluralform eingetragen werden.

⁴Es ist somit in der Spalte *Norm* nicht erlaubt, Flexionsendungen anzuhängen. Nur graphematische und/oder phonologische Operationen sind erlaubt, um die Form in Spalte *Norm* zu erzeugen.

Orig	Norm	Mod	Art der Anpassung
ich	ich		
wil	will		
dich	dich	dir	infl
berichten	berichten		

Tabelle 5: Modernisierung durch abweichende Kasusreaktion (Bsp. N1, 589–592)

Tab. 5 zeigt ein Beispiel für eine Modernisierung, die durch eine Abweichung in der Kasusreaktion des Verbs *berichten* entstanden ist. Es regiert im Text ein Pronomem im Akkusativ (*dich*). Standarddeutsch muss das Pronomen jedoch im Dativ (*dir*) stehen.

Besondere Regelungen:

- *ward* (3.P.Sg.Prät. von *sein*):

Diese Wortform wird zwar von vielen Sprechern des Deutschen noch verstanden, ist aber im Standarddeutschen nicht mehr üblich. Daher wird sie als abweichende Flexion klassifiziert und mit *wurde* modernisiert (vgl. Tab. 6).

Orig	Norm	Mod	Art der Anpassung
do	da	als	sem
er	er		
gefvuert	geführt		
wart	ward	wurde	infl

Tabelle 6: Sonderfall *ward* für die 3.P.Sg.Prät. von *sein* (Bsp. T1, 9040–9043)

- Ekthlipsis:

Hierbei handelt es sich um den Ausfall eines Konsonaten mit nachfolgender Kontraktion; die Präteritalform *betete* wird im Frühneuhochdeutschen bspw. oftmals zu *bete* verkürzt. Da *bete* in dieser Form auch im Standarddeutschen existiert (bspw. in *ich bete*), wird diese zunächst in Spalte *Norm* eingetragen. Wenn aus dem Kontext deutlich wird, dass es sich aber eigentlich um eine Präteritalform handelt (also eine Ekthlipsis vorliegt), wird *betete* in Spalte *Mod* eingetragen: *bete – bete – betete – infl*.

- *ge*-Präfix bei Infinitiven:

Manche Infinitive im Frühneuhochdeutschen werden mit einem *ge*-Präfix markiert. Es kann also beispielsweise *gesein* statt *sein* heißen (vgl. 7). Wenn, wie in Tab. 7, die präfigierte Infinitivform kein existierendes neuhochdeutsches Wort ist, wird der unpräfigierte (‘normale’) Infinitiv in Spalte *Norm* eingetragen. Alle restlichen Spalten

bleiben leer. Stimmt der mit *ge-* präfigierte Infinitiv jedoch mit einer existierenden neuhochdeutschen Form überein (bspw. *gehalten* – oberflächliche Übereinstimmung mit dem Partizip von *halten*), wird wie im Beispiel in Tab. 8 vorgegangen.

Orig	Norm	Mod	Art der Anpassung
daz	dass		
es	es		
anders	anders		
niht	nicht		
gesein	sein		
mag	mag	kann	sem

Tabelle 7: Sonderfall *ge-*präfigierte Infinitive (und keine oberflächliche Übereinstimmung mit Part. Prät.) (Bsp. W1, 7323–7328)

Orig	Norm	Mod	Art der Anpassung
daz	dass		
er	er		
daz	das		
Cruetz	Kreuz		
nit	nicht		
moht	mochte	konnte	sem
getragen	getragen	tragen	infl

Tabelle 8: Sonderfall *ge-*präfigierte Infinitive (und oberflächliche Übereinstimmung mit Part. Prät.) (Bsp. N4, 4157–4163)

- Dativ-*e*:

Im Frühneuhochdeutschen wird noch oft das Dativ-*e* verwendet (*dem Manne* statt *dem Mann*.) Im Neuhochdeutschen ist das Dativ-*e* als veraltet gekennzeichnet. Es wird bei der Normalisierung nicht berücksichtigt (s. Tab. 9), es sei denn, es entspricht der oberflächlichen Form eines existierenden Wortes (beispielsweise beim Plural-*e*, für ein Beispiel s. Tab. 10). In diesem Fall wird die entsprechende Form in die Spalte *Mod* eingetragen und als Anpassung der Flektion (*infl*) markiert.

Orig	Norm	Mod	Art der Anpassung
in	in		
dem	dem		
puche	Buch		

Tabelle 9: Sonderfall Dativ-e (Bsp. N1, 33–35)

Orig	Norm	Mod	Art der Anpassung
an	an		
dem	dem		
tage	tage	tag	infl

Tabelle 10: Sonderfall Dativ-e und Plural-e (Bsp. N1, 86–88)

- Ambiguität bei Nasalstrich:
Wird ein Nasalstrich verwendet, kann man oft nicht entscheiden, ob der Schreiber damit ein *n* oder *m* gekürzt hat (bspw. kann *einē* zu *einem* oder *einen* aufgelöst werden). In diesem Fall wird ‘zu unseren Gunsten’ normalisiert, wenn also *einem* vom Kasus her passt, wird das direkt in der Spalte *Norm* angesetzt.
- Ehemals starke Verben:
Es gibt einige Verben, die im Mittelhochdeutschen/Frühneuhochdeutschen noch stark flektieren, heute jedoch einer schwachen Flexion folgen. Entspricht die stark flektierte Form oberflächlich einer anderen Form aus dem Flexionsparadigma des Wortes, wird dieses in die Spalte *Norm* eingetragen und die angepasste Form in die Spalte *Mod* (gekennzeichnet als Flektionsfall (*infl*), bspw. *erscheine* vs. *erschien* in Tab. 11). Ist die stark flektierte Form neben der heute üblicheren schwach flektierten Form noch im Duden enthalten, wird diese in die Spalte *Norm* eingetragen und modernisiert (s. bspw. *ward* oder *koren* vs. *kürte* in Tab. 11). Ist die stark flektierte Form nicht mehr im Duden enthalten, jedoch im Lexer, wird in der Spalte *Norm* die Lexerform übernommen und der Fall wie bei extinkten Wörtern als *ext*-Fall markiert (s. *phlag* vs. *pflgte* in Tab. 11, für eine Erklärung der extinkten Formen siehe Abschnitt 2.3.3).

Orig	Norm	Mod	Art der Anpassung
da	da	als	sem
er	er		
mir	mir		
also	also	so	sem
geweldecliche	gewaltecliche	gewaltig	ext
erscheyn	erscheine	erschien	infl
vnd	und		
koren	koren	kürten	infl
eynen	einen		
weyzen	weisen		
meister	meister		
vnd	und		
phlag	phlag	pflachte	ext
sein	sein	ihn	infl
dez	des	am	sem
anderen	anderen		
tagiz	tages	tag	infl

Tabelle 11: Sonderfall ehemals starke Verben (Bsp. B2, 8812–8822, N1, 2495–2499 und N4, 6020–6024)

2.3.3 Extinkte Wortformen (*ext*)

Nicht alle Token können mit dem bisher beschriebenen Verfahren normalisiert bzw. modernisiert werden. Es gibt Wortformen, die nur bis zum Frühneuhochdeutschen gebraucht wurden und heute extinkt (= ausgestorben) sind. In all diesen Fällen wird in Spalte *Norm* das Lemma angesetzt, das im Mittelhochdeutschen Wörterbuch von Lexer (Onlineversion) zu finden ist⁵. Lexer verzeichnet manchmal mehrere alternative Schreibweisen für ein Lemma – für die Normalisierung ist stets die erste angesetzte Schreibweise zu wählen. In Spalte *Mod* wird eine passende neuhochdeutsche Übersetzung eingetragen, in Spalte 4 wird das Kürzel *ext* eingetragen. Es kann der Fall sein (wie bereits in Kap. 2.3.1 beschrieben), dass das unmittelbar vorausgehende Token ebenfalls angepasst werden muss, beispielsweise wenn Lexerform und Übersetzung im Genus voneinander abweichen. Tab. 12 gibt zwei Beispiele für extinkte Wortformen.

⁵Wenn die Wortform im Lexer nicht verzeichnet ist, kann alternativ noch im Deutschen Wörterbuch von Grimm nachgeschlagen werden (ebenfalls online verfügbar).

www.woerterbuchnetz.de/lexer
www.woerterbuchnetz.de/DWB

Orig	Norm	Mod	Art der Anpassung
zehant	zehant	sogleich	ext
do	da		
gieng	ging	kam	sem
Iudas	Judas		
daz	dass		
sie	sie		
daz	das	den	infl
hor	hor	Dreck	ext
vnd	und		
die	die		
stain	stein	steine	infl
an	an	auf	sem
mich	mich		
werfend	werfend	werfen	infl

Tabelle 12: extinkte Wortformen (Bsp. N4, 917–920 und N4, 4319–4328)

Besondere Regelungen:

• Negationspartikel *en*:

Die Negationspartikel *en* kann entweder der einzige negierende Ausdruck im Satz sein oder gemeinsam mit *nicht/kein* etc. stehen. In jedem Fall wird wie in Tab 13 vorgegangen.

Orig	Norm	Mod	Art der Anpassung
ich	ich		
en	en	nicht	ext
vynde	finde		
keyne	keine		
sach	sache		

Tabelle 13: Negationspartikel *en* (Bsp. B1, 2219–2223)

• Suffigierte Adjektive/Adverben:

Adjektive und Adverben werden im Frühneuhochdeutsch oftmals zusätzlich suffigiert (bspw. *völliglich*, statt *völlig*). Da die suffigierte Form so nicht mehr im modernen Deutsch existiert, werden diese Token wie extinkte Wortformen behan-

delt (vgl. Tab. 14).

Orig	Norm	Mod	Art der Anpassung
daz	das		
mag	mag	kann	sem
kein	kein		
mensche	Mensch		
volkomenlich	volkomenlich	vollkommen	ext
gesagin	sagen		

Tabelle 14: Suffigierte Adjektive/Adverben (Bsp. N1, 540–545)

2.4 Kontextabhängige Normalisierung

Während sich bei der Normalisierung einige Fälle immer auf die gleiche Weise lösen lassen (z.B. die Negationspartikel *en* zu *nicht*, *ward* zu *wurde*), gibt es für einige Fälle keine pauschale Lösung, mit der diese immer wieder gleich normalisiert werden. Hier kommt es immer auf den Kontext an. Für einen Vergleich siehe die folgenden drei Beispiele:

Orig	Norm	Mod	Art der Anpassung
zuo	zu		
dem	dem		
gemainen	gemeinen		
volk	Volk		
do	da		
schrieren	schrien		
arm	arm		
vnd	und		
reich	reich		
alle	alle		
gemein	gemein	gemeinsam	sem
do	da		
giench	ging		
pylatus	Pilatus		
tzv	zu		
der	der		
gemain	gemeine	Gemeinschaft	ext

Tabelle 15: Kontextabhängige Normalisierung, N4, 2957–2960; T1, 2899–2905; T1, 3225–3239

Teil II

Guidelines (EN)

1 Introduction

Historical language data in German are characterized by a high degree of linguistic variation, especially with respect to phonology and graphematics, but also morphology, syntax and lexicon.⁶ This variation is above all attributable to the lack of a universal standardized language in the historically attested linguistic stages Old High German (OHG), Middle High German (MHG) and Early New High German (ENHG), such that these stages of the German language exhibit a high degree of dialectal diversity.

The phonological and graphematical variation, especially, represent an enormous obstacle to the automatic processing of such data. Existing tools, for example in the domain of part-of-speech annotation, have been trained on modern Standard German data and as a result either can't be used at all for earlier linguistic stages or they are only useful to a limited extent. Searching for specific data in a historical corpus is also hampered by the high degree of variability, since certain word forms can't be specifically queried without knowing all of the potential spellings beforehand. Automatic lemmatization of all word forms is likewise significantly more complicated.

In order to make such data available for analysis and querying, it is reasonable to map the historical word forms onto uniform, equivalent forms, that is to *normalize* them. Modern Standard German or a corresponding dictionary, e.g., Duden (<http://www.duden.de/woerterbuch>), can function as a reference for these equivalent forms. Normalization onto modern Standard German forms especially makes sense when one plans to use existing language processing tools. Of course a mapping onto an earlier linguistic stage, for instance, is also conceivable (say, a Middle High German Corpus, which might be mapped onto a unified and simplified version of Middle High German). It is crucial, however, that a clearly defined reference is available. In the case of Middle High German a reference dictionary such as Lexer (<http://woerterbuchnetz.de/Lexer>) can serve this purpose.

In the following article, guidelines will be introduced, by means of which historical German language data may be manually normalized. For this purpose all word forms are to be mapped onto modern, standard-language equivalents. These guidelines are a part of the project “St. Anselm’s Questions for Mary: (Digital) Interpretation, Evaluation and Edition of the Entire German Textual Tradition (14th–16th Centuries)”⁷ (e.g., described in Schultz-Balluff and Dipper (2013)). All of the examples come from the corpus that was compiled in the course of this project. As part of this project manually normalized

⁶The authors thank Simone Schultz-Balluff, Malin Frey, Julika Nelken, Cornelia Johnen, Maurice Spengler, Katharina Bort, Nadine Lordick and Maria Anselm for their help in normalizing the data and their valuable comments on the guidelines, as well as Adam Roussel for the translation of the guidelines into English.

⁷This project is financed by the Deutsche Forschungsgemeinschaft (German Research Foundation), Project DI 1558/4.

texts from the corpus were used in order to train and evaluate a specially-developed utility for automatic normalization (*Norma*, e.g., described in Bollmann et al. (2012)).

2 Guidelines

2.1 General Description

A completely normalized text will contain four columns in all (cf. Table 1):

- 1st column (*Orig*): the original token
- 2nd column (*Norm*): the token in its *normalized* form
- 3rd column (*Mod*): the *modernized* token
- 4th column (*Type*): one of the abbreviations *infl*, *sem* or *ext*, specifying the difference between the *normalized* and *modernized* forms (*infl* = inflection, *sem* = semantics, *ext* = extinct word form)

Orig	Norm	Mod	Type
alz	als		
mein	mein		
chind	Kind		
di	die		
mandat	mandate	Fußwaschung	ext
begangen	begangen	vollzogen	sem
het	hat	hatte	infl

Table 1: Tabular representation of a normalization

2.2 The *Norm* Column: Normalization

This column is mandatory. It contains an equivalent of the original token which is present in modern Standard German (column 2)⁸. Equivalence may be determined **exclusively via phonological and/or graphematic equivalence operations**. The part of speech of the original token is retained during normalization.

- phonological operations, such as:
 - diphthongization: ENHG /u/ > NHG /au/ (*hus* – *Haus*, *mus* – *Maus*); ENHG /i/ > NHG /ei/ (*min* – *mein*, *schriben* – *schreiben*)

⁸Uncertain cases can be resolved using the Duden dictionary (<http://www.duden.de/hilfe/rechtschreibung>); when the token is present there, it may be used.

- qualitative diphthong change: ENHG /ou/ > NHG /au/ (*boum* – *Baum*, *vrouwe* – *Frau*)
- vowel lowering: ENHG /ii/ > NHG /ö/ (*küneg* – *König*); ENHG /u/ > NHG /o/ (*sunne* – *Sonne*)
- vowel rounding: ENHG /e/ > NHG /ö/ (*zwelf* – *zwölf*); ENHG /i/ > NHG /ü/ (*flistern* – *flüstern*)
- palatalization of /s/ (ENHG *slange* – NHG *Schlange*)
- epenthesis (ENHG *eigenlich* – NHG *eigentlich*, ENHG *namenlich* – NHG *namentlich*)
- contractions (*er seit* – *er sagte*)
- graphematic operations: *t = d* (*kint* – *Kind*); *cz = z* (*czu* – *zu*); *ai = ei* (*gewonhait* – *Gewohnheit*); *i = j* (*iare* – *Jahre*)

The word form produced by such operations is entered in the *Norm* column (cf. Table 2). This word form is then to be used for two further tests, which will be explained in the following sections.

Orig	Norm	Mod	Type
lybe	liebe		
frouwe	Frau		
sage	sage		
myr	mir		
dynes	deines		
kyndes	Kindes		
marter	Marter		

Table 2: Normalization (Ex. B1, 262–269)

2.3 The *Mod* Column: Modernization

Whereas the *Norm* column is mandatory, the columns *Mod* and *Type* are only used in certain particular circumstances. They are reserved for those cases in which the normalized forms in *Norm* do not result in a meaningful modern sentence, e.g., due to inflectional or semantic discrepancies. If there are no discrepancies of this sort, then the *Mod* and *Type* columns should remain empty (cf. Table 2).

2.3.1 Semantic Divergence (*sem*)

In some cases, the token in the *Norm* column, which was arrived at via phonological and graphematical equivalence operations, exists in standard German as it ever has, but

now possesses a meaning quite different from that of the original token (for instance MHG/ENHG *Urlaub* instead of *Erlaubnis*). In these cases, a semantically more appropriate alternative should be entered in the *Mod* column; the abbreviation *sem* should then be entered in column 4 (thereby making clear that the modernization in the *Mod* column has been carried out at the semantic level).

Caution: In this step one should not attempt to meet the demands of a published edition of the text, i.e. one should not “translate” the original text. Semantic adjustments should only be made when the original form and modern form diverge significantly from one another.

Orig	Norm	Mod	Type
petrus	petrus		
behilt	behielt	versteckte	sem
sich	sich		
vnder	unter	hinter	sem
eyme	einem		
steyne	stein		

Table 3: Modernization: Semantic divergence (Ex. B1, 1411–1418)

Table 3 illustrates two examples of typical semantic adjustments: *sich unter einem Stein behalten* would no longer be understood today, so an alteration to *sich hinter einem Stein verstecken* is allowable.

Semantic adjustments could result in neighboring tokens needing to be adjusted also – for instance, whenever the original token and the altered token have different genders (cf. Table 12).

Special cases:

- **Obsolete words:**

If a word with the same meaning is present in Duden, but is considered obsolete and rarely used nowadays, such that it would be difficult to understand (e.g., *Muhme* used for *Tante*, *Schächer* for *Mörder*, *dünken* for *denken*), a modern word should be entered in the *Mod* column. The type of the adjustment should be noted with the abbreviation *sem*, that is as a semantic adjustment.

2.3.2 Inflectional Variation (*infl*)

Normalization has the overarching goal of producing morphologically and syntactically correct sentences (apart from the restriction of not altering word order). In many cases this means that the word form in the *Norm* column, which was determined via phonological and/or graphematical equivalence operations, does not appear with

inflection it ought to according to the standard grammar of today. The reasons for this can be manifold (e.g., diverging inflection classes in the historical language variant, one form governs a different case, graphematical abbreviations)⁹. For this reason, in such cases a morphologically and syntactically adjusted form must additionally be entered in the *Mod* column; column 4 should then contain the abbreviation *infl* (specifying that the word form in *Mod* is a modernization at the level of morphology or syntax).

Orig	Norm	Mod	Type
Dese	diese		
wort	wort	worte	infl
spricht	spricht		
vnser	unser		
liber	lieber		
here	herr		
iheesus	jesus		
cristus	christus		

Table 4: Modernization due to divergent plural marking (Ex. B1, 10–17)

Table 4 presents an example of a modernization instance. The original token *wort* exists in this *surface* form in modern Standard German (apart from its capitalization). However, the original token is a plural form (recognizable, for instance, due to its determiner *dese*). This type of plural construction is no longer possible for the lemma *Wort* today. Therefore the modern plural form must be entered in the *Mod* column as well.

Orig	Norm	Mod	Type
ich	ich		
wil	will		
dich	dich	dir	infl
berichten	berichten		

Table 5: Modernization due to divergent case government (Ex. N1, 589–592)

Table 5 shows a modernization instance that is the result of the verb *berichten* governing different cases at different times. In the text it governs an accusative pronoun (*dich*). In Standard German the pronoun would have to be in the dative case (*dir*).

Special cases:

⁹Therefore it is **not** allowed to append inflectional endings in the *Norm* column. Only graphematical and/or phonological operations are allowed for the production of the form in the *Norm* column.

- *ward* (3.P.Sg.Pret. of *sein*):

This word form is still understood by many German speakers, however it is no longer usual in Standard German. For this reason it is classified as a inflectional variant and modernized as *wurde* (cf. Table 6).

Orig	Norm	Mod	Type
do	da	als	sem
er	er		
gefvuert	geführt		
wart	ward	wurde	infl

Table 6: Special case of *ward* for the 3.P.Sg.Pret. of *sein* (Ex. T1, 9040–9043)

- Ecthlipsis:

This case involves the suppression of a consonant and a subsequent contraction; the preterite form *betete* is often shortened to *bete* in Early New High German. Since the word form *bete* also exists in Standard German (for instance in *ich bete*), this form is first entered in the *Norm* column. If the context clearly indicates that a preterite form is intended (i.e. it represents a case of ecthlipsis), one should enter *betete* in the *Mod* column: *bete* – *bete* – *betete* – *infl*.

- *ge-* prefix for infinitives:

Some infinitives in Early New High German are marked with a *ge-* prefix. One may find *gesein*, for example, instead of *sein* (cf. Table 7). If, as in Table 7, the prefixed infinitive form is not an extant word in Modern German, one should enter the unprefixd ('normal') infinitive in the *Norm* column. All of the other columns should be left empty. However, if the infinitive prefixed with *ge-* matches some extant Modern German word form (e.g., *gehalten* – superficial match with the participle of *halten*), then this case should be handled as in the example in Table 8.

Orig	Norm	Mod	Type
daz	dass		
es	es		
anders	anders		
niht	nicht		
gesein	sein		
mag	mag	kann	sem

Table 7: Special case of infinitives prefixed with *ge-* (no superficial match with past participle) (Ex. W1, 7323–7328)

Orig	Norm	Mod	Type
daz	dass		
er	er		
daz	das		
Cruetz	Kreuz		
nit	nicht		
moht	mochte	konnte	sem
getragen	getragen	tragen	infl

Table 8: Special case of infinitive with *ge-* prefix (superficial match with past participle) (Ex. N4, 4157–4163)

- Dative *e*:

In Early New High German the dative *e* is used often (*dem Manne* instead of *dem Mann*.) In Modern German the dative *e* is marked as obsolete. It is not considered during normalization (see Table 9), unless it matches the surface form of an existing word (for example those with an *-e* plural, see Table 10). In this case the appropriate form should be entered in the *Mod* column and marked as an inflectional adjustment (*infl*).

Orig	Norm	Mod	Type
in	in		
dem	dem		
puche	Buch		

Table 9: Special case of dative *-e* (Ex. N1, 33–35)

Orig	Norm	Mod	Type
an	an		
dem	dem		
tage	tage	tag	infl

Table 10: Special case of dative *-e* and plural *-e* (Ex. N1, 86–88)

- Ambiguity of macron:

When a macron is used, it is often hard to determine whether the scribe in this instance is omitting an *n* or an *m* (e.g., *einē* can mean *einem* or *einen*). In such cases, one should normalize “for our benefit”, so if *einem* would match case-wise, it is entered directly into the *Norm* column.

- Formerly strong verbs:

There are some verbs in Middle High German and Early New High German that were inflected as strong verbs, but which now are inflected as weak verbs. If the strongly inflected form superficially matches another form of the word's inflectional paradigm, then this form should be entered in the *Norm* column and the adjusted form in the *Mod* column (marked as relating to inflection (*infl*), e.g., *erscheine* vs. *erschien* in Table 11). If the strongly inflected form is still present in Duden along with the more common weakly inflected form, then this form should be entered in the *Norm* column and modernized (see for example *ward* or *koren* vs. *kürte* in Table 11). If the strongly inflected form is no longer in Duden, but still in Lexer, then the Lexer form should be entered in the *Norm* column and the case should be marked with *ext*, as with extinct words (see *phlag* vs. *pflęte* in Table 11, for an explanation of extinct word forms see Section 2.4).

Orig	Norm	Mod	Type
da	da	als	sem
er	er		
mir	mir		
also	also	so	sem
geweldecliche	gewaltecliche	gewaltig	ext
erscheyn	erscheine	erschien	infl
vnd	und		
koren	koren	kürten	infl
eynen	einen		
weyzen	weisen		
meister	meister		
vnd	und		
phlag	phlag	pflęte	ext
sein	sein	ihn	infl
dez	des	am	sem
anderen	anderen		
tagiz	tages	tag	infl

Table 11: Special case of formerly strong verbs (Ex. B2, 8812–8822, N1, 2495–2499 and N4, 6020–6024)

2.4 Extinct Word Forms (*ext*)

Not all tokens can be normalized or modernized according to the procedure described above. There are word forms that were only used up until Early New High German and which are now extinct. In all of these cases the lemma which is found in the Middle High German dictionary of Lexer (online version) should be entered in the *Norm* column¹⁰. Lexer sometimes records multiple alternative spellings for a lemma – the first spelling is always to be used for normalization. In the *Mod* column an appropriate Modern German translation should be entered, and column 4 should contain the abbreviation *ext*. It could be the case (as described in Section 2.3.1) that the previous neighboring token also needs to be adjusted, for example when the Lexer form and translation vary in gender. Table 12 contains two examples of extinct word forms.

Orig	Norm	Mod	Type
zehant	zehant	sogleich	ext
do	da		
gieng	ging	kam	sem
Iudas	Judas		
daz	dass		
sie	sie		
daz	das	den	infl
hor	hor	Dreck	ext
vnd	und		
die	die		
stain	stein	steine	infl
an	an	auf	sem
mich	mich		
werfend	werfend	werfen	infl

Table 12: Extinct word forms (Ex. N4, 917–920 and N4, 4319–4328)

Special cases:

- Negative particle *en*:

The negative particle *en* can either be the sole negating expression in a sentence or occur in conjunction with *nicht/kein*. In both cases one should proceed as in Table 13.

¹⁰If the word form is not present in Lexer, one may also look it up in the Grimms' Deutsches Wörterbuch (also available online).

www.woerterbuchnetz.de/lexer
www.woerterbuchnetz.de/DWB

Orig	Norm	Mod	Type
ich	ich		
en	en	nicht	ext
vynde	finde		
keyne	keine		
sach	sache		

Table 13: Negative particle *en* (Bsp. B1, 2219–2223)

- Suffixed adjectives/adverbs:

Adjectives and adverbs are often further suffixed in Early New High German (e.g., *völliglich*, instead of *völlig*). Since the suffixed form no longer exists in Modern German, these tokens are to be handled as extinct word forms (cf. Table 14).

Orig	Norm	Mod	Type
daz	das		
mag	mag	kann	sem
kein	kein		
mensche	Mensch		
vollkomenlich	vollkomenlich	vollkommen	ext
gesagin	sagen		

Table 14: Suffixed adjectives/adverbs (Ex. N1, 540–545)

2.5 Context-Dependent Normalization

Though some normalization cases can always be handled in the same way (for instance, the negative particle *en* always becomes *nicht*, *ward* becomes *wurde*), other cases won't have a blanket solution with which they can always be normalized. These cases depend on their context. For a comparison, see the following three examples:

Orig	Norm	Mod	Type
zuo	zu		
dem	dem		
gemainen	gemeinen		
volk	Volk		
do	da		
schrieren	schrien		
arm	arm		
vnd	und		
reich	reich		
alle	alle		
gemein	gemein	gemeinsam	sem
do	da		
giench	ging		
pylatus	Pilatus		
tzv	zu		
der	der		
gemain	gemeine	Gemeinschaft	ext

Table 15: Context-dependent normalization, N4, 2957–2960; T1, 2899–2905; T1, 3225–3239

Literatur

- Marcel Bollmann, Stefanie Dipper, Julia Krasselt, and Florian Petran. Manual and Semi-automatic Normalization of Historical Spelling – Case Studies from Early New High German. In *Proceedings of the First International Workshop on Language Technology for Historical Text(s) (LThist2012)*, Wien, 2012.
- Simone Schultz-Balluff and Stefanie Dipper. St. Anselmi Fragen an Maria – Schritte zu einer (digitalen) Erschließung, Auswertung und Edition der gesamten deutschsprachigen Überlieferung (14.-16. Jh.). In Anne Bohnenkamp-Renken, editor, *Medienwandel/Medienwechsel in der Editionswissenschaft*, number 35 in editio, pages 173–191. De Gruyter, Berlin/New York, 2013.

Appendizes: Beispiele – Appendices: Examples

A Fälle ohne weitere Anpassung – Normal Cases

Sigle Source	Originaltoken Orig	Normalisierung Norm	Modernisierung Mod	Art der Anpassung Type
H1, Token 161ff.	do sprach sancte anshelm sage mir liebe mutter maria wie wasz der anfang desz liedensz dienisz lieben kindsz	da sprach sankt anselm sage mir liebe mutter maria wie war der anfang des leidens deines lieben kinds		
H1, Token 259ff.	sancte anshelmus sprach sage mir liebe maria welcherlei waren die pfennige	sankt anselm sprach sage mir liebe maria welcherlei waren die pfennige		
H1, Token 399ff.	sage mir liebe maria warest ü ouch dobie do sprach die mutter gottesz nein	sage mir liebe maria warst du auch dabei da sprach die mutter gottes nein		
W1, Token 197ff.	da gieng Iudas scarioth zw den	da ging judas iscariot zu den		

	fursten der Iuden	fürsten der juden
W1, Token 319ff.	sy warn tewr und chospar und die Iuden warn da cz yeherusalem die daz gut hetten von der erschafft	sie waren teuer und kostbar und die juden waren da zu jerusalem die das gut hatten von der erbschaft
M9, Token 16ff.	EIn hoher lerer der hiez anshelmus	ein hoher lehrer der hieß anselm
M9, Token 316ff.	Do nam er drey besunder seiner iunger	da nahm er drei besondere seiner jünger
b1, Token 192ff.	Das hore mvter vnd sich vnd erbarme dich vber mich	das höre mutter und sieh und erbarme dich über mich
N4, Token 1104ff.	do viengen sie min kint do zoch sant peter sin swert vzz	da fingen sie mein kind da zog sankt petrus sein schwert aus

	der schaid	der scheide
D3, Token 4905ff.	Ich	ich
	weynite	weinte
	claite	klagte
	vnd	und
	sprach	sprach
	wee	weh
	mir	mir
	wee	weh
	myn	mein
	lybes	liebes
	kint	kind
	wy	wie
	gar	gar
	vorblichin	verblichen
	dyne	deine
	wangen	wangen
	sint	sind

B Semantik-Anpassungen – Semantic Adjustments

Sigle Source	Originaltoken Orig	Normalisierung Norm	Modernisierung Mod	Art der Anpassung Type
M2, Token 129ff.	vber alles himelische her und vber alle creature dye göt ie geschüeff	über alles himmlische heer und über alle kreatur die gott je schuf	jede	sem
M2, Token 332ff.	Vnd waren meinem chinde als veint wie lange si das güt heten gehabt	und waren meinem kind als feind wie lange sie das gut hatten gehabt	so	sem
M2, Token 358ff.	der waz also geytig do er dye phennig ansach	der war also geizig da er die pfennig ansah	so gierig als Pfennige	sem sem sem infl
M2, Token 441ff.	do gieng mein chind mit seyne *i*ungerner ab dem perg syon	da ging mein kind mit seinen jüngern ab dem berg sion	von	sem
M2, Token 511ff.	also verre als man mit ainem staine	also verre als man mit einem steine	so weit wie Stein	sem ext sem infl

	gewerfen mag vnd viel nieder an seine chnie	werfen mag und fiel nieder an seine knie	kann auf	sem sem
M6, Token 471ff.	vnd erhor mich schir wann ich pin petrubt	und erhöre mich schier wann ich bin betrübt	schnell denn	ext sem
N4, Token 1075ff.	do gieng Iudas dar	da ging Judas dar	dorthin	sem
N4, Token 4012ff.	do gesigten die Iuden an im pylatus sprach iur will sol an im ergan	da siegten die juden an ihm pilatus sprach ihr wille soll an ihm ergehen	geschehen	sem
T1, Token 1039ff.	Wen ich chusse der ist iz den habt vaste	wen ich küsse der ist es den habt feste	haltet fest	sem infl
D4, Token 5415ff.	bin ich schier aller meiner kreffft entsetzt worden	bin ich schier aller meiner kraft entsetzt worden	all beraubt	infl sem
N1, Token 639ff.	do machte lazaruz vnd	da machte lazarus und	bereiteten	sem

	marta vnd maria magdalena meynem libin kinde vnd seynen iungern eyne wirtschafft	martha und maria magdalena meinem lieben kind und seinen jünger eine wirtschafft	ein gastmahl	infl sem
N4, Token 936ff.	Ir sint zwen gar glich ainander Iesus vnd Iacobus siner muomen sun den man da haizzet sinen bruoder	ihr sind zwei gar gleich einander jesus und Jakobus seiner muhmen sohn den man da heißt seinen bruder	ihrer tante nennt	infl sem sem
Be1, Token 3870	Im luiff rich vnd arm nach wib vnd kint	ihm lief reich und arm nach weib und kind	liefen frauen kinder	infl sem infl

C Flexionsanpassungen – Inflectional Adjustments

Sigle Source	Originaltoken Orig	Normalisierung Norm	Modernisierung Mod	Art der Anpassung Type
N3, Token 7ff.	der	der		
	bat	bat		
	vnsr	unser	unsere	infl
	libe frawn	liebe frauen	Frau	infl
M4, Token 330ff.	die	die		
	waren	waren		
	meinen kint	meinen kind	meinem	infl
	also veint	also feind	so	sem
Stu1, Token 404ff.	vnd	und		
	inen	ihnen		
	die	die		
	fuessz	füße		
	het	hat	hatte	infl
	gewaeschen	gewaschen		
	vnd	und		
	in	ihn	ihnen	infl
	sinen	seinen		
	hailigen fronlichnam	heiligen fronleichnam		
het	hat	hatte	infl	
geben	geben	gegeben	infl	
Hk1, Token 340ff.	da	da	als	1
	er	er		
	die	die		
	pfennig	pfennig	pfennige	infl
	ansach	ansah		
	da	da		
	geuiellen	gefielen		
sy im	sie ihm			
b1, Token 30ff.	namen	nahmen		
	sie	sie		
	die	die		
	strengen	strängen	stränge	infl
	Vnd	und		
	zugen	zogen		
	den	den		
	andern	anderen		
arm sein	arm sein			
N4, Token 48ff.	vnd	und		
	do	da	als	sem
	er	er		
	des	des	das	infl
	lang	lang	lange	infl
	zit	zeit		
	vnd	und		
	lang stund	lang stunde	lange stunden	infl infl

	von allem sinem hertzen het begert	von allem seinem hertzen hat begehrt	all	infl
	· Do erschain im vnser fraw	· da erscheine ihm unser fraw	hatte erschien unsere	infl infl
B1, Token 2338ff.	ab her der were	ob er der wäre		
	· durch den syn vater so vil kynder lys toten	· durch den sein vater so viel kinder ließ töten	wegen dem viele	sem infl infl
B2, 6182ff.	isz ward also grosze vinsternisse über alles erteriche	es ward also große finsternisse über alles erdreiche	wurde so finsternis erdreich	infl sem infl infl
B2, 6389ff.	da stund eyn here bie dem kreuze der hiesze Centurio der sulde mynes kyndes huden	da stand ein herr bei dem kreuze der hieße centurio der sollte meines kundes hüten	kreuz hieß	infl infl
			mein kind	infl infl
T1, 8597	wan ich wil gern hie bei im meines	wann ich will gern hier bei ihm meines	denn mein	sem infl

	endes beiten	endes beiten	ende erwarten	infl ext
Me1, 2377ff.	nu hor anshelmus / waz ich dier iemerleicher ding sagen will	nun höre anselm / was ich dir jämmerlicher ding sagen will	jämmerliche dinge	infl infl
Me1, 3032ff.	fraw waz gedecht du do	frau was gedacht du da	dachttest	infl

D Extinkte Wortformen – Extinct Word Forms

Sigle Source	Originaltoken Orig	Normalisierung Norm	Modernisierung Mod	Art der Anpassung Type
M5, Token 49ff.	do kam zwhand Iudas In den garten	da kam zehant judas in den garten	sogleich	ext
M6, Token 87ff.	die saligen magt mit vil haissen zehern vnd mit vil gepetz	die seligen Magd mit viel heißen zeheren und mit viel gebet	selige jungfrau vielen tränen	infl sem infl ext
M2, Token 8391ff.	sach er mich als minnichleychen vnd als gutleichen an das ich	sah er mich als minneclichen und als guotlichen an dass ich	so lieblich so freundlich	sem ext sem ext
Be1, Token 236ff.	es von dem anfang ergieng vncz an das end	es von dem anfang erging unz an das ende	bis	ext
N1, Token 157ff.	do salt u ansteigen mit aller diner begerunge	da sollt du ansteigen mit aller deiner begerunge	sollst hinaufsteigen all deinem begehren	infl sem infl infl ext
N4, Token 572ff.	wann din erbermd ist an mir	wann dein erbermede ist an mir	denn Erbarmen	sem ext

	erfuelt	erfüllt		
B1, Token 2823ff.	vnd czogen ym eyn syden cleyt an czu smoheit	und zogen ihm ein seiden kleid an zu smaheit	seidenes zur schmach	infl infl ext
B1, Token 3197ff.	bys an dy stat	bis an die stat	stätte	ext
B1, Token 4192ff.	Do sach myn lybes kynt das itczunt dy schrift an ym irvullz was	da sah mein liebes kind dass iezuo die schrift an ihm erfüllst war	jetzt erfüllt	ext infl
N1, Token 10012ff.	von wannen bist u	von wannen bist du	woher	ext
n1, Token 4986ff.	vnd wart ein grosze vinster veber alles ertreich uncz an die newenten stund	und ward ein große vinster über alles erdreich unze an die neunten stunde	kam eine finsternis bis zu der	sem infl ext ext sem infl
N4, Token 37ff.	wie es von dem angeng ergiang hintz an daz end	wie es von dem aneganc erging hinze an das ende	Anfang bis	ext ext
N4, Token 141ff.	so en	so en	nicht	ext

	mag ich noch en sol nit mer wainen	mag ich noch en soll nicht mehr weinen	kann nicht	sem ext
N4, Token 1483ff.	die pfariseien die besamnoten sich all	die Pharisäer die besamneten sich alle	versammelten	ext
N4, Token 1762ff.	so hast du min drystund verlaeugent	so hast du mein dristunt verleugnet	mich dreimal	infl ext