

Korpushandbuch
– Texterfassung, Lemmatisierung, Indizierung –
von
Tobias Anselm Kemper und Thomas Klein
Bonn, im Januar 2006

Seitherige Änderungen oder Ergänzungen:
Müller, Joschko, Klein, Wojtinek

Inhalt

1	Das Projekt “Mittelhochdeutsche Grammatik”	5
2	Das Textkorpus.....	5
2.1	Das Hauptkorpus.....	7
2.2	Das Erweiterungskorpus.....	7
2.3	Das Kernkorpus.....	7
2.4	Das reduzierte Kernkorpus.....	7
3	Daten- und Dateorganisation.....	8
3.1	Textsiglen.....	8
3.2	Dateiarten.....	8
3.2.1	Einheit aus Text- und Indexdateien.....	8
3.2.2	Textdateien.....	10
3.2.3	Vorlemmatisierungsdateien und Indices.....	10
3.3	Verzeichnisse.....	11
3.4	Datensicherung.....	12
4	Texterfassung.....	13
4.1	Arbeitsgrundlage.....	13
4.2	Dateiformat.....	13
4.3	Aufbau der Textdateien.....	13
4.4	Der Ordnungsteil.....	13
4.4.1	Der Aufbau des Ordnungsteils insgesamt.....	14
4.4.2	Der handschriftenorientierte Ordnungsteil.....	15
4.4.3	Der editionsorientierte Ordnungsteil.....	15
4.4.4	Die computerunterstützte Erstellung des Ordnungsteils.....	16
4.5	Transkription und Codierung.....	17
4.5.1	Zeichenbestand im Korpus (mit Unicode-Dezimalzahl).....	17
4.5.2	Literatur zur Paläographie und Kodikologie.....	17
4.5.3	Transkription.....	18
4.5.4	Besondere Buchstabenformen.....	18
4.5.5	Abkürzungen, Super- und Subskripte.....	18
4.5.5.1	Zur Codierung von Abkürzungen, Super- und Subskripten.....	18
4.5.5.2	Die Codierung einzelner Abkürzungen.....	18
4.5.5.3	Auflösung von Abkürzungen.....	20
4.5.6	Codierung von Initialen, Initialmajuskeln, Versalien, Schriftgraden, Rubriken.....	20
4.5.7	Handschriftliche Interpunktion.....	20
4.5.8	Schwer lesbare, nicht lesbare und verlorene Buchstaben und Wörter.....	21
4.5.9	Kommentierung und Dokumentation diverser handschriftlicher Befunde.....	22
4.5.9.1	Zur Platzierung und Gestaltung der K-Klammern.....	22
4.5.9.2	Empfohlene Abkürzungen für häufige Bemerkungen in K-Klammern.....	22
4.5.10	Emendationen und Kommentare.....	23
4.5.11	Tilgungen und andere eindeutige Fehler des handschriftlichen Textes.....	23
4.5.12	Nachträge, Verweiszeichen, Korrekturen.....	24
4.6	Die Schlußüberprüfung.....	24
5	Präeditierung.....	24
5.1	Worteinheiten.....	25

5.1.1	Trennung von in der Handschrift zusammengeschriebenen Wörtern.....	25
5.1.2	Krasis.....	25
5.1.3	Zusammenfügung von in der Handschrift (fehlerhaft) getrennten Wörtern.....	26
5.1.4	Komposita.....	26
5.1.5	Worttrennung am Zeilenende.....	26
5.1.6	Partikelverben (trennbare Präfixverben).....	26
5.2	Mehrgliedrige Konjunktionen.....	27
5.3	Fremdsprachliches.....	28
5.4	Interpunktion.....	29
5.4.1	Moderne Interpunktionszeichen und Interpunktionsregeln.....	29
5.4.2	Doppelkomma, einfaches Komma, kein Komma.....	29
5.4.2.1	Das Doppelkomma.....	29
5.4.2.2	Das einfache Komma.....	30
5.4.2.3	Kein Komma.....	30
5.4.3	Unterbrochene Elementarsätze.....	30
5.4.4	Direkte Rede.....	31
6	Segmentierung von Belegen im fertigen Index.....	31
7	Datensatzgestalt. Lemma, Kategorie, Bestimmung.....	32
7.1	Unveränderlicher und veränderlicher Ansatz. [!], [!!] und *.....	32
7.2	Der Lemmaansatz.....	34
7.3	Die Kategorien.....	37
7.4	Grammatische Formenbestimmung.....	39
7.4.1	Abkürzungen.....	39
7.4.2	Die Reihenfolge innerhalb der grammatischen Bestimmung.....	40
7.4.3	Nomina.....	40
7.4.4	Verben.....	40
7.4.5	Adjektive und Partizipien.....	41
7.4.5.1	Zusatzanzeichnungen bei Adjektiven.....	43
7.4.6	Pronomina und Artikel.....	44
7.4.6.1	Zusatzanzeichnungen bei Pronomina.....	44
7.4.6.2	Anzeichnung von <i>all</i> und <i>sëlb</i>	44
7.4.7	Zahlwörter.....	46
7.4.8	Übungsbeispiele zur Anzeichnung und Zusatzanzeichnung von Adj, Pron, Num.	47
7.4.9	Indeclinabilia.....	48
7.4.10	Eigennamen.....	48
7.4.11	“Kleine Wörter” und einzelne Wendungen.....	51
8	Die automatische Vorlemmatisierung.....	55
8.1	Dateikonvertierung von Word- in ASCII-Text.....	56
8.2	(Zur Dateikonvertierung zwischen Excel- und ASCII-Text).....	56
8.3	Die Erstellung des WOK-Indexes.....	57
8.4	Die automatische Vorlemmatisierung.....	58
8.5	Verbesserung der automatischen Vorlemmatisierung.....	60
8.6	Das “Löcherstopfen”.....	61
8.7	Die Erstellung des VK-Indexes.....	62
9	Die Vorlemmatisierungskorrektur.....	63
10	Von der Vorlemmatisierung zum Normalformindex.....	67
10.1	Auswertung der Korrekturinträge.....	67
10.2	Markierung der Pronominaladverbien.....	68
10.3	Markierung der Partikelverben.....	68
10.4	Formale Überprüfung.....	69
10.5	Überprüfung der Zählung.....	69
10.6	Erzeugung des Normalformindexes.....	69
10.7	Überprüfung des Normalformindexes.....	70
10.8	Abgleich des Normalformindexes mit der Textdatei.....	71
11	Die Normalform.....	72
11.1	Synkope, Apokope, Elision und Kontraktion.....	72

11.2	Unorganische Vokale und Konsonanten.....	73
11.3	Auslautverhärtung.....	73
11.4	Volle Nebensilbenvokale im Beleg.....	73
11.5	Vereinfachung von Geminaten im Auslaut und vor Konsonant.....	73
11.6	Besonderheiten bei den schwachen Präterita.....	73
11.7	Auflösung des Nasalstriches.....	74
11.8	Einzelfälle.....	74
11.8.1	-ære, -er (Nomina agentis, <i>rîtære</i>).....	74
11.8.2	<i>bi-gihte, bihte</i>	74
11.8.3	<i>ëht, oht, êckert, ockert</i>	74
11.8.4	<i>ent-</i> (Verbalpräfix).....	75
11.8.5	<i>heiden, kristen</i> und ähnliche Leute.....	75
11.8.6	-hèit (Substantivsuffix).....	75
11.8.7	-inne bzw. -în (movierte Feminina).....	75
11.8.8	<i>iuwer</i>	75
11.8.9	<i>komen</i>	76
11.8.10	<i>lich</i> und <i>lîch-</i>	76
11.8.11	<i>lâzen, liè</i> (nicht <i>liè<z></i>) bei <i>\lie</i>	76
11.8.12	<i>oder</i> oder <i>eder</i>	76
11.8.13	<i>sîneme</i>	76
11.8.14	<i>sol(e)n</i>	76
11.8.15	<i>unde</i> (nicht <i>un<de></i>) bei <i>un\</i>	76
11.8.16	<i>vrouwe</i> (Nebenformen).....	76
11.8.17	<i>vröuwen</i>	76
11.8.18	<i>w</i> für <i>wu</i>	76
11.8.19	<i>wa-</i> und <i>wô-</i> Stämme.....	76
11.8.20	<i>zer-</i> (Verbalpräfix).....	77
12	Anmerkungsfeld.....	77
12.1	Anzeichnung von Dittographien.....	77
13	Der Kontrolltext und seine Korrektur.....	77
13.1	KTX-Korrekturausdruck.....	77
13.2	KTX-Korrektur am Bildschirm.....	78
14	Die Lemmaänderung.....	79
14.1	Änderungen in der Lemmaliste und Einarbeitung in die Indizes.....	79
14.2	Große Änderung.....	79
14.2.1	Einarbeitung der angesammelten Änderungsvorschläge in die Lemmaliste.....	79
14.2.2	Anpassung der Ansätze des Korpus an die neue Lemmaliste.....	80
14.3	Änderungen mit <i>lem-aend</i>	81
15	Der Grapho-Index.....	82
15.1	Zur Gestaltung des Graphofeldes.....	83
15.1.1	Nasalstrich über <i>m</i> und <i>n</i>	83
15.1.2	Hsl. Schreibung < <i>w</i> > für normalmittelhochdeutsch < <i>wu</i> >, < <i>wuo</i> >.....	83
15.1.3	Wortformen mit rundgeklammerten (<i>en</i>) und (<i>ni</i>) in der Nform:.....	83
15.1.4	Hsl. Wortformen <i>\menscheit</i> etc., Nform <i>~mèn(ni)sch-hèit</i> etc.....	83
15.1.5	Klammern im grapho-Feld.....	84
15.1.6	Sonderformen.....	84
15.1.7	Fragliches.....	84
16	Das Reimregister.....	85
16.1	Die Isolation der Reimwörter.....	85
16.2	Die Reimanalyse.....	86
16.3	Das Anlegen der drei Reimindices.....	86
17	Bearbeitung der Lemmatisierungswörterbücher.....	87
18	Die Projektprogramme.....	89
18.1	Grundsätzliches.....	89
18.2	A-ZU-H.....	90
18.3	BASIS-S.....	90

18.3.1	Voraussetzungen.....	90
18.3.2	Aktualisierung dieser Dateien.....	91
18.3.3	Arbeitsweise von BASIS-S.....	91
18.3.4	Die Ergebnisdatei von BASIS-S.....	93
18.4	GRAPHO.....	93
18.5	IND-NORM.....	94
18.6	KORP-LST.....	94
18.7	KORP-STA.....	94
18.8	LEM.....	94
18.9	LEM-AEND.....	94
18.10	MORPHO.....	94
18.11	NF-KTX.....	94
18.12	NFO-WOK.....	94
18.13	NFORM.....	94
18.14	NFORM-TX.....	94
18.15	NONS.....	95
18.16	PRUEF-NF.....	95
18.17	SORTA.....	95
18.18	SUCH-NF.....	95
18.18.1	Beispiele.....	97
18.18.1.1	Einfache Suche.....	97
18.18.1.2	Kombinierte Suchbegriffe.....	98
18.18.1.3	Zweizeilige Suchbegriffe.....	98
18.18.1.4	Mehrzeilige und komplexe Suchbegriffe.....	99
18.18.1.5	Suche mit Schleifen.....	99
18.19	TX-PRUEF.....	100
18.20	WBK-TEIL.....	100
18.21	ZAHL.....	100
18.21.1	Die Erstellung / Ergänzung einer (fortlaufenden) Numerierung.....	100
18.21.2	Die Überprüfung einer (fortlaufenden) Numerierung.....	101
18.21.3	Vers- und Wortzahl mit Nullen auffüllen.....	101
19	Der WordPerfect-Editor.....	101
19.1	Grundlegende Funktionen.....	101
19.2	Tastaturlayouts.....	102
19.2.1	Tastaturlayout "Vorlemmatisierungskorrektur".....	102
19.2.2	Tastaturlayout "Ergänzung des Lemmatisierungswörterbuchs".....	103
19.2.3	Tastaturlayout "KTX-Korrektur am Bildschirm".....	103
19.2.4	Tastaturlayout "Indexkorrektur und -auswertung".....	103
19.2.5	Texteingabe.....	104
19.3	Personalisierte Makros.....	104
19.4	Übersicht über die Editor-Makros.....	104
20	Literatur.....	107
21	Anhang.....	109
21.1	Die Projektarbeitsstellen.....	109
21.1.1	Bochum.....	109
21.1.2	Bonn.....	109
21.1.3	Halle.....	109
21.1.4	Jena.....	109
21.1.5	Oslo.....	110
21.2	Das Korpus.....	110
21.2.1	Das Hauptkorpus.....	110
21.2.2	Das Erweiterungskorpus.....	113
21.3	Header.....	116
21.3.1	Der Header für Textdateien.....	116
21.3.2	Der Header für Indices.....	117
21.4	Annotierungsfahrplan: Von der Textdatei zum N-Index.....	118

1 Das Projekt “Mittelhochdeutsche Grammatik”

Das seit 1997 von der DFG geförderte Projekt “Mittelhochdeutsche Grammatik” ist aus den seit einigen Jahren artikulierten Überlegungen zur Erarbeitung einer neuen wissenschaftlichen Grammatik des Mittelhochdeutschen hervorgegangen. Gegen die vorhandenen Grammatiken des Mittelhochdeutschen richtet sich – teilweise bereits seit dem 19. Jahrhundert – die Kritik, daß sie wegen der Konzentration auf die höfische Literatur als Textbasis und wegen des Konstruktes des “Normalmittelhochdeutschen” nicht die sprachliche Realität der mittelhochdeutschen Zeit abbilden.

Mittlerweile hat die sprachhistorische Forschung der letzten Jahrzehnte – v. a. im Zusammenhang mit den Forschungen zur Grammatik des Frühneuhochdeutschen – die (korpus-) theoretischen, methodologischen und technischen Möglichkeiten zur Erfassung, Analyse und Beschreibung nicht-normalisierter Sprachstufen entwickelt und erprobt. Ziel des Projektes “Mittelhochdeutsche Grammatik” ist nun die Erarbeitung einer neuen umfassenden wissenschaftlichen Grammatik des Mittelhochdeutschen auf der Basis eines strukturierten, möglichst handschriftennahen Textkorpus', das computergestützt lemmatisiert, indiziert und ausgewertet wird.

Erarbeitet wird diese Grammatik an drei Arbeitsstellen unter der Leitung von Prof. Dr. Thomas Klein in Bonn, Prof. Dr. Klaus-Peter Wegera in Bochum und Prof. Dr. Hans-Joachim Solms in Halle; im Bereich der mhd. Wortbildung besteht eine Kooperation mit Dr. Heinz Prell in Oslo.¹

Über die (unzureichenden) Quellengrundlagen und Methoden der bisherigen Grammatiken des Mittelhochdeutschen und die Grundlagen der geplanten neuen Grammatik (Quellen und Korpusstruktur) informiert grundlegend (mit Hinweisen auf weitere Literatur):

Klaus-Peter Wegera: Grundlagenprobleme einer neuen mittelhochdeutschen Grammatik. In: Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Hg. von Werner Besch u. a. Bd. 2, 2. vollst. neu bearb. u. erw. Aufl. Berlin, New York 2000 (Handbücher zur Sprach- und Kommunikationswissenschaft 2), S. 1304-1320.

Das Verfahren der computergestützten Lemmatisierung und grammatischen Auswertung der Sprachquellen für die geplante Grammatik des Mittelhochdeutschen beschreibt grundsätzlich und überblickshaft:

Thomas Klein: Zur Frage der Korpusbildung und zur computerunterstützten grammatischen Auswertung mittelhochdeutscher Quellen. In: Klaus-Peter Wegera (Hg.): Mittelhochdeutsche Grammatik als Aufgabe. ZfdPh 110 (1991), Sonderheft, S. 3–23.

Das vorliegende “Projekthandbuch” dokumentiert die im Grammatikprojekt geltenden Konventionen, beschreibt alle Arbeitsschritte von der Texterfassung und Präeditierung über die Lemmatisierung bis hin zur Erstellung verschiedener grammatischer Indices und informiert über die dabei anzuwendenden Computerprogramme.

2 Das Textkorpus

Aufgrund der umfangreichen Überlieferung mittelhochdeutscher Texte und damit Sprachquellen, die insbesondere seit der zweiten Hälfte des 13. Jahrhunderts stark zunimmt, kann einer Grammatik des Mittelhochdeutschen nicht die Gesamtheit des Überlieferten zugrundegelegt werden, sondern es muß eine begründete exemplarische Auswahl getroffen werden.

Bei der Zusammensetzung eines Korpus' für eine Grammatik des Mittelhochdeutschen kommen vor allem drei Prinzipien zur Anwendung, nämlich das **Prinzip der Handschriftentreue**, das **Prinzip des strukturierten Korpus'** und das **Prinzip des dynamischen Korpus'**.

¹ Die Adressen und Mitarbeiter der verschiedenen Arbeitsstellen siehe im Anhang (Kap. 21.1).

Der **Rückgriff auf die Handschriften** selbst (oder Faksimiles) – und nicht die Editionen – ist unerlässlich angesichts der Editionspraxis, aus mehreren Parallelüberlieferungen einen “kritischen Text” zu erstellen und den Sprachstand zu normalisieren. Die Eingriffe in den Text erfolgen zwar in unterschiedlichem Maße – religiöse und juristische Prosatexte sowie Urkunden sind in der Regel weniger betroffen –, der Rückgriff auf die Handschriften ist aber in jedem Falle unerlässlich. Die Berücksichtigung der jeweiligen Editionen oder – im Idealfall – von diplomatischen Abdrucken insbesondere bei Zweifelsfällen ist damit natürlich keineswegs ausgeschlossen. Die Auswahl von Handschriften für das Korpus der mittelhochdeutschen Grammatik ist nach der Maßgabe erfolgt, daß die Handschriften nicht mehr als fünfzig Jahre von der Entstehung des Textes selbst entfernt sind, im gleichen oder wenigstens benachbarten Sprachraum geschrieben wurden und von einer einzigen Schreibhand stammen.

Das **Korpus** muß **strukturiert** sein, um die Gesamtheit des Überlieferten wenigstens ansatzweise exemplarisch zu vertreten und um der Variabilität und der Diachronie der Sprache in einem Zeitraum von drei Jahrhunderten Rechnung zu tragen. Die grundlegenden Kriterien zur Strukturierung des Korpus¹ sind die diachrone und die diatopische Binnengliederung des Mittelhochdeutschen sowie die grundlegende Unterscheidung von verschiedenen Textsorten.

Nach den Erfahrungen mit der Arbeit an der “Grammatik des Frühneuhochdeutschen” und angesichts der oft nur ungefähren Datierung der mittelalterlichen Handschriften wird der Gesamtzeitraum von 300 Jahren zwischen 1050 und 1350 in **Abschnitte von je 50 Jahren Dauer** unterteilt. Aufgrund der geringen Überlieferungsdichte werden allerdings die ersten beiden Abschnitte zusammengefaßt. Insgesamt ergibt sich somit eine Unterteilung in fünf Abschnitte: 1050–1150 (I), 1150–1200 (II), 1200–1250 (III), 1250–1300 (IV), 1300–1350 (V).

Diatopisch wird das Korpus untergliedert in – je nach Zeitabschnitt variierend – bis zu **sieben Sprachräume**, drei oberdeutsche und vier mitteldeutsche: bairisch (1), bairisch-alemannisch (2), alemannisch (3), mittelfränkisch (4a), rheinfränkisch-hessisch (4b), ostmitteldeutsch (5), ostfränkisch (6). Durch die Kombination der fünf Zeiträume mit bis zu sieben möglichen Dialekträumen ergeben sich 25 Korpussektoren.²

Die **Unterscheidung verschiedener Textsorten** erfolgt zunächst nur rudimentär zwischen Vers- und Prosatexten, da Verstepte im Gegensatz zu Prosatexten den Zwängen von Metrum und Reim unterliegen. Von den Prosatexten werden die Urkunden als eigene Form unterschieden, da sie einerseits vorgegebenen Formen und Formeln entsprechen mußten, andererseits regionale Besonderheiten in der Sprache der in aller Regel sehr genau datierten und lokalisierten Urkunden besonders gut beobachtet werden können.

Über diese grundlegende Unterscheidung von Vers- und Prosatexten sowie Urkunden hinaus werden alle Texte – soweit möglich – nach formalen und inhaltlichen Kriterien verschiedenen **“Texttypen”**³ zugeordnet, um damit u. U. texttypenabhängige Besonderheiten erfassen zu können: Bibelübersetzung / Bibelerzählung (Bi), Bibelkommentar (BK), Theologica (Th), Gebetstexte (G), Mystica (My), Legenden (L), Predigten (Pr), Epik (E), Historica (H), Sachtexte (S), Rechtstexte (R), Urkunden (U). Die Zuordnung der Texte zu diesen Texttypen ist den tabellarischen Übersichten in Anhang 21.2.1 zu entnehmen.

Schließlich ist das **Korpus dynamisch**, um der Tatsache Rechnung zu tragen, daß zur Auswertung hochfrequenter sprachlicher Phänomene weniger Text notwendig ist als zur Auswertung seltener oder äußerst rarer Erscheinungen. Unterschieden wird daher zwischen (a) dem **Hauptkorpus** als der wesentlichen Bezugsgröße, das möglichst pro Korpussektor aus je zwei Vers- und Prosatexten sowie einer Urkundenstrecke im Umfange von jeweils idealerweise 12.000 bis 12.500 Wortformen besteht (insgesamt 102 Sprachquellen, siehe 2.1), (b) dem **Erweiterungskorpus**, das zusätzlich etwa 120 weitere Texte im Umfang von möglichst 3.000 Wortformen umfaßt (siehe 2.2), (c) dem **Kernkorpus**, einer Reduktionsstufe des Hauptkorpus auf 66 Texte (siehe

² Theoretisch wäre eine größere Zahl von Korpussektoren möglich; die geringere Zahl erklärt sich dadurch, daß aufgrund der Überlieferungslage manche Sektoren nicht besetzt werden können.

³ Der problematische Begriff “Textsorte” wird hier ausdrücklich gemieden.

he 2.3), und schließlich (d) dem **reduzierten Kernkorpus**, bei dem die 66 Texte des Kernkorpus' auf 7.000 Wortformen gekürzt sind (siehe 2.4).

2.1 Das Hauptkorpus

Das Hauptkorpus umfaßt im Idealfall pro Korpussektor zwei Vers- und zwei Prosatexte bzw. Textausschnitte sowie eine Urkundenstrecke im Umfang von jeweils 12.000 bis 12.500 Wortformen. Tatsächlich sind zahlreiche Sektoren aber nur teilweise besetzt, wenn aus dem entsprechenden Zeit- und Sprachraum keine oder nicht ausreichend viele Handschriften überliefert sind, die den an sie gestellten Anforderungen gerecht werden. Urkundenstrecken etwa gibt es grundsätzlich erst ab der zweiten Hälfte des 13. Jahrhunderts.

Über die Zusammensetzung des Hauptkorpus', das bereits vollständig erfaßt worden ist und auch vollständig lemmatisiert und ausgewertet werden soll, informieren die tabellarischen Übersichten in Anhang 21.2.1. Durch Kursivdruck sind alle die Texte gekennzeichnet, die nicht zugleich auch Bestandteil des Kernkorpus sind (dazu siehe 2.3). Kleinere Änderungen in der Zusammensetzung des Hauptkorpus sind nach Abschluß der aktuellen Projektphase (2001–2003) möglich.

NB: Aufgrund geänderter Datierungen und Lokalisierungen müssen mehrere Handschriften im Hauptkorpus ausgetauscht werden [Stand: Juli 2004]!

2.2 Das Erweiterungskorpus

Das Erweiterungskorpus umfaßt über das Hauptkorpus hinaus etwa 120 zusätzliche Texte bzw. Textausschnitte im Umfang von jeweils 3.000 Wortformen. Der geringere Umfang ist zum einen arbeitspragmatisch begründet, zum anderen dadurch, daß das Erweiterungskorpus vor allem für Fragen der Lautlehre ausgewertet werden soll, um in diesem Bereich die diachrone und diatopische Variabilität und Entwicklung besser in den Blick nehmen zu können.

Über die Zusammensetzung des Erweiterungskorpus' orientieren die tabellarischen Übersichten in Anhang 21.2.2. Texte, über deren Aufnahme ins Erweiterungskorpus noch nicht endgültig entschieden ist, sind durch Kursivierung gekennzeichnet. Zu beachten ist ferner, daß nicht alle Texte des Erweiterungskorpus vollständig eingelesen und lemmatisiert werden sollen; manche Texte sollen als "Fußnotentexte" nur herangezogen werden, um einzelne Phänomene zu belegen. Vor der Arbeit mit Texten des Erweiterungskorpus ist daher auf jeden Fall Rücksprache notwendig.

2.3 Das Kernkorpus

Gemäß dem Prinzip des dynamischen Korpus' wird der Untersuchung frequenter Phänomene – etwa im Bereich der Lautlehre, aber u. U. auch im Bereich der Wortbildung – aus arbeitspragmatischen Gründen nicht das gesamte Hauptkorpus zugrundegelegt. Das Kernkorpus ist gegenüber dem Hauptkorpus reduziert auf in der Regel je einen Vers- und Prosatext pro Korpussektor. In der tabellarischen Übersicht über das Hauptkorpus (siehe 21.2.1) sind die Texte, die auch zum Kernkorpus gehören, recte gedruckt, diejenigen, die nicht zum Kernkorpus gehören, kursiviert.

2.4 Das reduzierte Kernkorpus

Für hochfrequente Phänomene kann u. U. noch nicht einmal das Kernkorpus komplett ausgewertet werden; phänomenabhängig kann daher das Kernkorpus im Umfang noch weiter verkleinert werden. Die Texte des reduzierten Kernkorpus' werden in diesem Falle auf 7.000 Wortformen gekürzt, und zwar auf die ersten und die letzten 3.500 Wortformen. Da bislang nur die Hallenser Arbeitsstelle mit dem reduzierten Kernkorpus gearbeitet hat, sind dort die Texte und Indices auch bereits entsprechend gekürzt worden. Aus diesem Grunde sollte vor einer er-

neuten Reduzierung des Kernkorpus mit der Hallenser Arbeitsstelle Rücksprache genommen werden.

3 Daten- und Dateiorganisation

3.1 Textsiglen

Jeder Korpus text wird durch eine Sigle bezeichnet, die sich zusammensetzt aus

- einer römischen Zahl, die den Zeitabschnitt bezeichnet, dem die Handschrift zugeordnet wird (I bis V)⁴,
- einer arabischen Zahl (ggf. ergänzt durch a oder b), die die dialektale Einordnung des Textes angibt⁵,
- den Buchstaben P, V oder U, mit der die Textsorte angegeben wird (Prosa, Verse, Urkunde),
- einem möglichst sprechend gewählten Buchstabenkürzel.

Die einzelnen Bestandteile dieser Sigle werden getrennt durch einen mittelhohen Punkt · [ALT+250]; zwischen dem Buchstaben, mit dem die Textsorte angegeben wird, und dem Buchstabenkürzel steht stattdessen der Unterstrich _.

Beispiel: III·3·P_Luci = Zeitraum III (1200–1250), Dialekt 3 (Alemannisch), Prosatext, Titel “Lucidarius”

3.2 Dateiarten

3.2.1 Einheit aus Text- und Indexdateien

Im Zuge eines Abgleiches von Text- und Indexdateien sind die Belegformen und Zeilungen der Indexdateien verbessert und zusätzliche Inhalte der Textdateien, die bisher nicht in die Indexdateien übergegangen waren, in die Indexdatei übernommen worden. Die so entstandenen neuen Dateien ersetzen das nach der Anzeignungsarbeit entstandene Doppel aus Text- und Indexdatei.

Das Abgleichsverfahren ist wie folgt bewandt:

- (Automatisch:) Der Textdateiheader wird unverändert dem Indexheader vorangestellt. Das erste @H wird getilgt. Der neue Header enthält also zwischen +H und @H die beiden alten Header, aber auch noch ein weiteres +H, das zwischen dem ersten und dem zweiten Teil steht. (Bei Gelegenheit wären manche Header wohl etwas auszulichten.)
- (Automatisch:) Die Felder **Wort**, **Wortart**, **Bestimmung**, **Normalform** enthalten das, was im Index stand, oder, wenn die Form im Index fehlte, die Feldeingangsmarke und ein [!].
- (**Händisch:**) In den Fällen mit [!] liegt entweder eine Verwerfung aufgrund geänderter Wortformengrenzen vor und ist, falls der Index hier falsch ist, zu verbessern; (automatisch:) oder es handelt sich um eine Wortform, die im Index gefehlt hat: meist eine Form aus einer ausgesternten Textdateizeile. In diesem Fall ist nichts zu ändern; die Zeile mit der fehlenden Wortform und den [!] wird übernommen.
- (Automatisch:) Feld **hsl. Wortform**: Vergleich zwischen Text- und Indexdatei, allerdings ohne die Interpunktion zu vergleichen. Abweichungen werden markiert.
 - **Lateinklammern** sind in einzeln als fremdsprachlich gekannte Einheiten aufgelöst worden.⁶
 - **Rubrizierungsklammern** sind so aufgelöst, dass jede geklammerte Wortform ein /R ins **Kommentarfeld** bekommt.
 - **E- und K-Klammern** werden in der Form /E[...] oder /K[...] ins **Kommentarfeld** derjenigen Wortform, hinter der sie erscheinen, geschrieben.

⁴ Dabei gilt I = 1050–1150; II = 1150–1200; III = 1200–1250; IV = 1250–1300; V = 1300–1350 (siehe Kap. 1).

⁵ Dabei gilt 0 = oberdeutsch (ohne weitere Differenzierung; nur in Zeiträumen I und III); 1 = bairisch; 2 = bair.-alemannisch (schwäbisch); 3 = alemannisch; 4 = westmitteldeutsch (in I, II und III); 4a = mittelfränkisch (nur in IV und V); 4b = rheinfränkisch-hessisch (nur in IV und V); 5 = ostmitteldeutsch; 6 = ostfränkisch (nur in V) (siehe Kap. 1).

⁶ In Übereinstimmung mit der Regel, die bisher für die Wiedergabe von Lateinklammern in den Indizes galt. Zur neueren Regel vgl. 5.3.

- **Schweifklammern** (mit zusätzlicher Stellenangabe) werden dem **Ordnungsteil** hinzugefügt. Die in einigen Texten schon zugesetzten eckig geklammerten Stellenangaben müssen nicht dieselben wie die geschweiften geklammerten sein.
- **Formen mit (=) oder =** sind zusammengezogen, **Formen mit /** getrennt worden.
- **(Händisch:) Belegfeld:** Prüfung der Abweichungen: Ist die Index- oder die Textdateilesart richtig? Wenn die Textdateilesart richtig ist, wird die Indexdateilesart verbessert.
- (Automatisch:) **Ordnungsteil:** Der Ordnungsteil der Indexdatei wird an den der Textdatei angeglichen, nur eine Abweichung im Stehen oder Fehlen der @-Marke wird bewahrt und vermerkt.
- **(Händisch:) Ordnungsteil:** Abweichungen im Stehen oder Fehlen der @-Marke prüfen: Wo liegt der Zeilenumbruch tatsächlich? Gegebenenfalls Verbesserung der Indexdatei, was die @-Marke oder auch die Zeilen- oder Blattzählung betrifft. (Die Wortformzählung wird nach dem Vergleich anhand der @-Marken automatisch erneuert; im Vergleichsblatt steht überall "01".)
- (Automatisch:) Das **Statusfeld** wird aus der Indexdatei übernommen, wird aber im Stehen oder Fehlen von * an der Textdatei ausgerichtet. Fehlt die Zeile im Index, wird als Status 00 angesetzt.
- (Automatisch:) **Kommentarfeld:** wird aus der Indexdatei übernommen, allerdings gegebenenfalls um /R, /E[...], /K[...] (siehe oben) aus der Textdatei ergänzt. Das geschieht aber nur dann automatisch, wenn es bei den Belegen keine Abweichung zwischen Index- und Textdatei gibt.
- **(Händisch:) Kommentarfeld:** Ist aus der Textdatei /R, /E[...], /K[...] (siehe oben) zu ergänzen, weicht aber der Beleg der Indexdatei von dem der Textdatei ab, dann muss die aus der Textdatei stammende Folge (/R, /E[...], /K[...]) händisch der richtigen Wortform zugeordnet werden, also in das richtige Kommentarfeld des Indexes gesetzt werden. /E- und /K-Vermerke stehen im Kommentarfeld der letzten Form, auf die sie sich beziehen.
- (Automatisch:) Das **Graphofeld** wird, wenn vorhanden, aus dem Index übernommen.
- Der Vergleich wird automatisch vorbereitet
 - **a_p_a_MhdVersenkrechungTextdatei.py** auf P anwenden, ergibt P_2,
 - **a_p_a_MhdZeilenzahlUndZählung.py** auf P_2 anwenden, ergibt P_2_2,
 - Inhalt von P_2_2 in P kopieren,
 - **müle008ea_VergleichTextUndIndex** auf P anwenden
 und händisch nachbearbeitet:
 1. Sich als erster oder zweiter Bearbeiter für einen bestimmten Text (z. B. Williram) in die Bearbeitungsliste eintragen. (Dieser Eintrag bleibt auch nach Rückgabe des Textes bestehen!)
 2. Den bestimmten Text – und zwar den ganzen Ordner – von Artus\C:\PP\P auf eigenen Rechner in ein Verzeichnis X kopieren.
 3. Vergleichsdatei .xls (im Mutterordner von I) durcharbeiten: Abgleich vornehmen. Gegebenenfalls einem anderen Bearbeiter Bescheid sagen.
 4. Abgleichsergebnis mit dem Makro **müle008ab_AusfuhrImAnschlussAnVergleichVonNfOderGr** aus Excel in .xxt-Datei umwandeln. Zu beachten ist, dass aus der Vergleichsdatei die jeweils linken Spalten (die aus der Indexdatei stammen) übernommen werden.
 5. .xxt-Datei mit Initiale versehen (z. B. ...-D.xxt für Diane).
 6. .xxt-Datei in Ordner Artus\C:\PP\P-Ablage sammeln.
 7. Wieder bei 1. einfangen, bis alle Texte im Korpus zweimal abgearbeitet sind.
 8. Wichtig für die Texte, von denen keine Abbildung vorlag: Vor dem nächsten Schritt muss sichergestellt sein, dass alle Zeilen mit einem noch an der Handschrift zu prüfenden Beleg zwei Gatter ## im Statusfeld haben.
 9. Die Paare aus Artus\C:\PP\P-Ablage in Artus\C:\PP\P neu verschieben und **a_p_a_MhdZeilenzahlUndZählung.py** auf diesen Ordner anwenden; den Ergebnisordner öffnen, alle Ordner markieren, ausschneiden und in den Ausgangsordner Artus\C:\PP\P neu einfügen.
 10. Jede .xxt-Datei aus dem Ordner Artus\C:\PP\P-Ablage mit der entsprechenden, von einem anderen Bearbeiter erstellten .xxt-Datei abgleichen (mit Makro **müle008aa_VergleichVonNfOderGrIndizes**); Vergleichsdatei ohne Initiale nennen. Wichtig ist, zuerst die Datei aus dem ersten Durchgang, dann die aus dem zweiten Durchgang anwählen (damit dann die geänderten Header, die ja zu den Dateien des ersten Durchganges gehören, übernommen werden, nicht die ungeänderten).
 11. Vergleichsdatei mit dem Makro **müle008ab_AusfuhrImAnschlussAnVergleichVonNfOderGr** aus Excel in .xxt-Datei umwandeln, und zwar aus den jeweils linken Spalten (was auch die vom Makro zuerst angebotene Aktion ist). Dies sind die Spalten der erstgewählten Datei, und dazu gehört dann auch der gegebenenfalls geänderte Header.
 12. Außer dieser .xxt-Datei die anderen Dateien im selben Unterordner von Artus\C:\PP\P neu löschen. Die .xxt-Datei in [Name].k oder (bei Grapho-Indizes) [Name].gr umbenennen, wobei [Name] für den bei der jeweiligen Datei gebräuchlichen Kurztitel steht.
 13. Artus\C:\PP\P neu in Artus\C:\PP\P umbenennen.

3.2.2 Textdateien

Die Benennung der Textdateien erfolgt im Prinzip durch die jeweiligen Textsiglen. Da aus computer-technischen Gründen die Dateinamen aus maximal acht Zeichen aus dem ASCII-Bereich bis ASCII 126 und einer dreistelligen Dateinamenextension bestehen sollten, sind im einzelnen folgende zusätzlichen Regelungen zu beachten:

- Der Punctus elevatus · [ALT+250] wird ersetzt durch den Bindestrich -.
- An die Stelle des Buchstabenkürzels tritt eine je eigene Zählung der Prosatexte, Verstexte und Urkundenstrecken eines jeden Sektors in arabischen Zahlen; die Ziffern werden direkt an die Buchstaben P, V oder U angefügt.
- Die Dateinamenextension der Textdateien zeigt den Bearbeitungsstand der jeweiligen Datei an; dabei bezeichnet

op eine Textdatei ohne präeditierung, die ggf. für die Präeditierung benutzt werden soll, **txt** eine präeditierte und interpungierte Textdatei, die für die Vorlemmatisierung benutzt werden soll bzw. worden ist,

hop eine Textdatei ohne präeditierung mit handschriftenorientiertem Ordnungsteil, mit der i. d. R. nicht weitergearbeitet wird,

txh eine präeditierte und interpungierte Textdatei mit handschriftenorientiertem Ordnungsteil, mit der i. d. R. nicht weitergearbeitet wird,

x, **opx**, **txx** eine Datei, deren Texterfassung noch nicht abgeschlossen ist und die für weitere Arbeitsschritte nicht ohne Rücksprache verwendet werden darf.

Beispiele: III-3-p2.op = unpräeditierte Textdatei des "Lucidarius" (III·3·P_Luci); III-3-p2.txt = präeditierte, zur Vorlemmatisierung freigegebene Textdatei des "Lucidarius"

3.2.3 Vorlemmatisierungsdateien und Indices

Vorlemmatisierungsdateien und Indices erhalten (zumindest in der Bonner Arbeitsstelle) als Dateinamen nicht die vollständige Sigle, sondern allein das Buchstabenkürzel; aus computer-technischen Gründen sollen dabei eventuell vorkommende Umlaute durch die jeweiligen unumgelauteten Vokale ersetzt werden. Mit Bindestrich - wird dem Buchstabenkürzel ein weiteres Kürzel zur Bezeichnung der jeweiligen Vorlemmatisierungsdatei bzw. des jeweiligen Indexes angefügt, und zwar

- wok** ("Wortform ohne kontext") für die Ausgabedatei von LEM, Option 1,
- vl.1** für die Ausgabedatei von LEM-2, Option 2,
- vl.1s** für die alphabetisch sortierte Datei xxx-vl.1,
- vl.2** für die Ausgabedatei von LEM-3 (Datei zum "Löcherstopfen"),
- vl.2z** für die in der Textreihenfolge sortierte Datei xxx-vl.2,
- vl** für die Ausgabedatei von LEM-4,
- vl.k** für die Ausgabedatei von LEM, Option 5, in der Vorlemmatisierungskorrektur erfolgt,
- lem.0** für die Ausgabedatei von LEM, Option 6,
- lem.d** für die Ausgabedatei von LEM, Option 7,
- lem** für die Ausgabedatei von LEM, Option 8,
- nf.0** für die (unkorrigierte) Ausgabedatei von NFORM,
- nf.k** für die Datei xxx-nf.0 nach Abschluß der KTX-Korrektur,
- ktx** für die Ausgabedatei von LEM, Option 9 (Korrekturtextdatei).

3.3 Verzeichnisse

Die zentralen Verzeichnisse für Projektdateien befinden sich sämtlich auf dem Rechner \\Artus. Artus wird als Zentralrechner benutzt, auf den über das Netzwerk zugegriffen werden kann. Dateien werden jedesmal vor Arbeitsbeginn von Artus herunter auf den jeweils benutzten Rechner **verschoben** (nicht kopiert!) und jedesmal nach Arbeitsende wieder auf Artus **kopiert** (nicht verschoben!), und zwar nicht in ein persönliches Verzeichnis C:\Name o. Ä. (was

es auf Artus nicht gibt), sondern in das sachlich zuzuordnende Verzeichnis. KTX-Ordner aber werden in die Ablage kopiert, nicht in den Korpusordner P selbst! Folgende für das Projekt wichtige Oberverzeichnisse finden sich auf Artus und sind dort für den Netzwerkzugriff freigegeben:

Zum einen:

- **C:\Editor** mit den Dateien (auch Makros enthaltend) zum Betrieb des WP-Editors.
- **C:\lem-wb** mit einer Kopie von LRG (Lemmaliste, das Original unter C:\PP\LRG), LRG-bst (Liste mit Endungen und zugehörigen Kategorien zur Bestimmung von Wortformen) und Lemmatisierungswörterbüchern.
- **C:\lem-wb_alt** mit etwelchen alten Dateien aus \lem-wb.
- **C:\Lexer** mit den Lexer-Dateien.
- **C:\spx** mit den Dateien zum Betrieb der Spitbol-Programme.

Kopien dieser Verzeichnisse befinden sich auf allen Rechnern, an denen auf den Editor, auf die Lemmaliste, den Lexer oder ein Spitbol-Programm zugegriffen werden muss. Ausgenommen sind die Lemmatisierungswörterbücher in C:\lem-wb und das Verzeichnis C:\lem-wb_alt.

Zum anderen

- **C:\PP.**

In **C:\PP** wiederum finden sich folgende Ordner:

C:\PP\Ausw_... enthält für ein bestimmtes Sachgebiet Dateien, die bei der Korpusauswertung, also beim Erstellen der Grammatik, entstanden sind und bearbeitet werden.

C:\PP\Editor-Makros neu enthält einige geänderte Editor-Makros; Kopien von ihnen liegen unter C:\Editor.

C:\PP\LRG enthält das aktuelle LRG (Lemmaregister = Lemmaliste).

C:\PP\P enthält das Mittelhochdeutschkorpus selbst, untergliedert in einen Ordner je Text, der die entsprechende Indexdatei mit sämtlichem Quell- und Annotationstext enthält. Das ist der Zustand nach der in 3.2.1 beschriebenen Änderung.⁷

⁷ Zum älteren Zustand: C:\P ist das "Archiv" des Projekts und beinhaltet sämtliche aktuelle Textdateien und Indices. Um Datenverluste zu vermeiden und sicherzustellen, daß sich in diesem Verzeichnis alle, aber auch nur die aktuellen Dateien an der jeweils richtigen Stelle befinden, sollen Änderungen an diesem Verzeichnis nur durch die wissenschaftlichen Projektmitarbeiter bzw. zusammen mit ihnen vorgenommen werden. Dateien, die in das Verzeichnis \P eingebracht werden sollen, werden daher ggf. zuerst nur in das Verzeichnis \P_Ablage kopiert und bleiben dort, bis sie in das Zielverzeichnis verschoben werden.

Das Verzeichnis \P ist untergliedert in einzelne Verzeichnisse für einen jeden Text des Kernkorpus. Diese Einzelverzeichnisse sind genauso benannt wie die jeweiligen Textdateien; als Verzeichnisnamenextension werden zur leichten Identifizierung die ersten drei Buchstaben des Buchstabenkürzels hinzugefügt (Beispiel: III-3-p2.Luc).

Diese Einzelverzeichnisse beinhalten in aller Regel nur eine einzige Datei, nämlich die jeweils aktuelle Textdatei (entweder mit der Extension op oder txt) sowie drei Unterverzeichnisse \A, \I und \T. Im Unterverzeichnis \T befinden sich ggf. weitere Textdateien mit den Extensionen op, hop und txh, im Unterverzeichnis \A Dateien mit Auswertungen, die sich auf den jeweiligen Text beziehen, im Unterverzeichnis \I die jeweils benötigten und aktuellen Indices. Dabei sind folgende Fälle zu unterscheiden:

- Während der Vorlemmatisierung können sich im Unterverzeichnis \I die Vorlemmatisierungsdatei sowie alle Vorstufen befinden, um bei eventuellen Programmfehlern Rückschlüsse auf die Genese des Fehlers ziehen zu können (d. h. xxx-wok, xxx-vl.1, xxx-vl.1s, xxx-vl.2, xxx-vl.2z, xxx.vl, xxx-vl.k). Es ist allerdings ausreichend, wenn diese Dateien während der Vorlemmatisierung nur in den Arbeitsverzeichnissen des jeweiligen Bearbeiters gespeichert sind.

- Nach Abschluß der Vorlemmatisierung befinden sich im Unterverzeichnis \I die Vorlemmatisierungsdatei, der Normalformindex und ggf. die Kontrolltextdatei (xxx-vl.k, xxx-nf.0, xxx.ktx). Alle anderen Dateien werden in das entsprechende Unterverzeichnis von \P_alt verschoben.

- Nach Abschluß der Kontrolltextkorrektur befindet sich im Unterverzeichnis \I nur noch der fertige Normalformindex (xxx-nf.0 oder xxx-nf.k). Alle anderen Dateien werden in das entsprechende Unterverzeichnis von \P_alt verschoben.

C:\PP\P_Ablage (vorhanden oder nicht je nach Bedarf) nimmt sämtliche Dateien auf, die in das Verzeichnis \P eingefügt werden sollen. Das Verschieben von Dateien von \P_Ablage nach \P soll in der Regel nur durch die wissenschaftlichen Projektmitarbeiter bzw. in Rücksprache mit ihnen erfolgen.

C:\PP\P_alt enthält alle älteren und mittlerweile überholten Dateiversionen, die für die aktuelle Arbeit nicht mehr benötigt werden, aber aus Sicherheitsgründen nicht gelöscht, sondern dorthin verschoben werden (vgl. oben).

C:\PP\P_E, **C:\PP\P_E_Ablage** und **C:\PP\P_E_alt** sind für das Erweiterungskorpus eingerichtet und entsprechen den oben genannten drei Hauptkorpusverzeichnissen.

C:\PP\P_Konventionen enthält Projekthandbücher, Protokolle, Listen mit verabredeten Abkürzungen u.Ä.

C:\PP\QuellenUndWeitereCorpora enthält

- **\AbbildungenAusgabenMetadaten** mit
 - **\Abbildungen**: Mikrofilme und Faksimilia
 - **\Ausgaben**: (Roths Alexanderausgabe:) musste die Abbildung vertreten, da Hs. verloren.
 - sowie der Projektdatenbank, der Abbildungsbestandsliste und weiteren, für den schnelleren Zugriff gefilterten Bestandslisten in einem Unterordner.
- **\CorporaReliqua** mit
 - **\P_MD**: enthält Textdateien und Indices zu verschiedenen mitteldeutschen Texten, die nicht zum Grammatikkorpus gehören, aber u. U. bei der Auswertung mit berücksichtigt werden können / sollen.
 - Dateien von Texten, die aus dem P-Korpus entfernt wurden, und bei der Sichtung des P_E_Korpus angefallene fragliche Dateien.
- weitere Corpora
- sprachwissenschaftliche Darstellungen und Bibliographien.

C:\PP\Sammlung ... enthält Dateien, die bei der Arbeit der im Ordnernamen genannten Mitarbeiter zeitweilig anfallen oder dauerhaft von ihnen hinterlassen worden sind.

C:\PP\Seminarangelegenheiten enthält Dateien mit Vorgaben für Hausarbeiten, Formulare und Ähnliches.

3.4 Datensicherung

- Ordner aus Artus\\C:\PP werden regelmäßig auf Triuwe\\C:\PP\PP_SICHERUNG gesichert. Selten oder gar nicht werden Ordner, deren Name “_alt” enthält, sowie \Zusatzcorpora und \Sammlung ... gesichert.

Triuwe\\C:\PP ist eine eigene Abteilung auf der Festplatte von Triuwe – größer als C: und nur hinsichtlich des Pfades der Abteilung C: untergeordnet.

- \Editor, \lem-wb, \Lexer und \spx sind durch die auf den anderen im Projekt genutzten Rechnern vorhandenen Kopien gesichert.
- Gerade in Arbeit befindliche Dateien werden nach dem täglichen Arbeitsende auf Artus kopiert, nicht verschoben; so verbleibt eine Kopie auf dem jeweils benutzten Rechner.

4 Texterfassung

4.1 Arbeitsgrundlage

Voraussetzung für die angestrebte möglichst genaue Transkription der Texte sind Arbeitsgrundlagen möglichst guter Qualität. Eine Arbeit mit den Handschriften oder Fragmenten selbst wird aus arbeitspraktischen und finanziellen Gründen nur im Ausnahmefall in Betracht kommen, wenn es nicht möglich ist, den Text von Reproduktionen zu transkribieren⁸; stattdessen sind im

⁸ Bislang sind nur die Würzburger Fragmente der “Mitteldeutschen Predigten” (III·5·p4) am Original kolationiert worden; weiterhin scheint eine Arbeit mit der Handschrift notwendig im Falle der “Margarethenlegende” Hartwigs von dem Hage (V·2·v2), weil die Handschrift so eng gebunden ist, daß auf dem

Idealfall ein Faksimile, ansonsten Photographien, Fotokopien, Microfilmaufnahmen und davon erstellte Rückvergrößerungen zu benutzen.

Zur Klärung der Frage, ob zu einer bestimmten Handschrift ein Faksimile existiert, ist neben dem "Verfasserlexikon" die "Bibliographie faksimilierter Handschriften" von H. und H. Zotter unter <http://www.kfunigraz.ac.at/ub/sosa/faksbib/faksbib-frame.html> zu konsultieren.

Sofern die Transkription und Kollation zuerst anhand von Rückvergrößerungen erfolgt, sind die Stellen, die dort nur schlecht oder gar nicht lesbar sind, zusätzlich am Microfilm zu überprüfen.

4.2 Dateiformat

Die Texterfassung kann entweder mit dem WordPerfect-Editor oder mit WordPerfect erfolgen. Zu empfehlen ist die Benutzung des WordPerfect-Editors, weil dann die Speicherung automatisch im richtigen Dateiformat erfolgt (ASCII). Wenn zur Texterfassung WordPerfect benutzt wird, so kann die Datei während der Arbeit durchaus im WordPerfect-Format gespeichert werden; nach Abschluß der Texterfassung muß die Datei auf jeden Fall im **Format "ASCII ohne Codes (DOS)"** gespeichert werden. Die Verwendung von MS Word ist wegen Schwierigkeiten bei der Konvertierung in das und aus dem ASCII-Format weniger empfehlenswert.

Bei allen folgenden Arbeitsschritten ist in aller Regel nur der WordPerfect-Editor zu benutzen; Ausnahmen werden im folgenden ausdrücklich erwähnt.

4.3 Aufbau der Textdateien

Am Anfang einer jeden Textdatei befindet sich der sogenannte Header, der mit "+H" als erster Zeile beginnt und mit "@H" als letzter Zeile endet (vgl. das Muster im Anhang unter 21.1). Der Header soll alle wichtigen Informationen über den Text, die Handschrift und über den aktuellen Bearbeitungsstand der jeweiligen Datei enthalten.

An den Header schließt sich, getrennt durch eine oder mehrere Leerzeilen, der Text an. Jede Textzeile beginnt mit dem Ordnungsteil (dazu vgl. 4.4); nach einem [TAB] und einem Leerzeichen folgt dann der Text. Jede Zeile endet mit einem festen Zeilenumbruch; schematisch hat eine Textzeile also folgenden Aufbau:

Ordnungsteil[TAB][BLANK]*Text*[RETURN]

4.4 Der Ordnungsteil

Der auf den ersten Blick unverhältnismäßig lange Ordnungsteil ist notwendig, damit auch bei der späteren Auswertung, z. B. in Gesamtindices, jede einzelne Wortform mit allen wesentlichen Informationen (Zeitraum, Dialekt, Vers- oder Prosatext, genaue Stelle in der Handschrift oder in der Edition) verbunden ist.

4.4.1 Der Aufbau des Ordnungsteils insgesamt

Der Ordnungsteil der Textdateien hat folgende Bestandteile:

1. die Korpussigle mit den Informationen zu Zeitraum (I bis V), Dialekt (0 bis 5) und Textart (P, V, U), getrennt jeweils durch den halbhohen Punkt · [ALT+250]⁹;
2. ein Textkürzel zur Bezeichnung des jeweiligen Textes, an die Sigle angeschlossen mit dem Unterstrich _;
3. der eigentliche Ordnungsteil, angeschlossen mit Bindestrich, und zwar entweder eine Zeilen-zählung nach der Handschrift ("**handschriftenorientierter Ordnungsteil**"¹⁰) oder eine Vers- bzw. Seiten- und Zeilen-zählung nach der / einer maßgeblichen Edition ("**editionsorientierter**

Microfilm das jeweils innere Drittel der Seiten nicht sichtbar ist.

⁹ Vgl. die Ausführungen zur Korpussigle in Kap. 3.1 und 3.2.

¹⁰ Dazu siehe 4.4.2.

Ordnungsteil¹¹). Bei der Entscheidung zwischen einem handschriften- und editionsorientierten Ordnungsteil gilt – von wohlbegründeten Ausnahmen abgesehen¹² – die Regel, daß sich die Zählung der Prosatexte primär (d. h. im Ordnungsteil) nach der Handschrift, die der Verstexte nach der Edition richtet.

Parallel zu der primären Zählung im Ordnungsteil wird zusätzlich die jeweils andere Zählung als sekundäre Zählung im laufenden Text in geschweiften Klammern {...} geführt. Bei Prosatexten, die im Ordnungsteil nach der Handschrift gezählt werden, sollte im laufenden Text wenigstens jeweils der Beginn einer neuen Seite oder eines neuen Kapitels angegeben werden, um einzelne Textstellen auch in der Edition auffinden zu können.

Beispiel:

II·1·P_Phys-129v,03 {1,1} *{D*}itze bu\och
II·1·P_Phys-129v,04 redenot unde zellet
II·1·P_Phys-129v,05 michilen wi\$tu\om. uon tieren un(=
II·1·P_Phys-129v,06 de uon fogilen.(;) aller#eri\$ t uon
II·1·P_Phys-129v,07 dem lewen.,, wie \$iniu dinch gele(=
II·1·P_Phys-129v,08 gen \$int.(.) {1,2} *(D*)o der alte iacob. \$inen
II·1·P_Phys-129v,09 \$un ge\$egenote unde gewihte.,, do
II·1·P_Phys-129v,10 chod er.(;) “welf de\$ lewen.,, bi\$ t|u iuda.(?)
II·1·P_Phys-129v,11 wer \$col irwekchen uon dineme
II·1·P_Phys-130r,01 ge\$lahte einen man.,, wer. ane got.(?)”
II·1·P_Phys-130r,02 {1,3} *(V*)on dem lewen zellent diu bu\och.,,

Umgekehrt muß bei Verstexten die Seiten-/Folio- und Zeilenzählung der Handschrift ebenfalls in geschweiften Klammern {...} eingefügt werden. Da in der Mehrzahl der mittelalterlichen Handschriften von Versdichtungen die Verse abgesetzt geschrieben sind, decken sich meist die handschriftlichen Zeilen mit den Versen, so daß sich eine i. w. additive Zählung ergibt, bei der jede Zeile der Textdatei mit dem editionsorientierten Ordnungsteil beginnt, der nach [TAB] und [BLANK] in geschweiften Klammern die Zählung der Handschrift folgt. Bei den wenigen, zumeist frühen Handschriften mit nicht abgesetzt geschriebenen Versen steht die handschriftliche Zählung jeweils dort, wo in der Handschrift ein Zeilenumbruch erfolgt ist, d. h. in der Regel mitten im Vers.

Bei Minusversen der Handschrift gegenüber der Edition müssen dann natürlich ggf. die entsprechenden Verszahlen in der Editionsählung übersprungen werden; bei Plusversen der Handschrift werden die Plusverse mit der Verszahl des unmittelbar vorangehenden Verses der Edition, ergänzt mit den Indices a, b, c, ..., weitergezählt.

Beispiel:

III·0·V_Iw-8151 {158v,23} div het mit ir \$inne.
III·0·V_Iw-8152 {158v,24} ir beider vnminne.
III·0·V_Iw-8153 {158v,25} Braht z|allem gv\ote.,,
III·0·V_Iw-8154 {158v,26} al\$ \$i in ir mv\ote.
III·0·V_Iw-8155 {159r,01} Lange het gegert.(.)
III·0·V_Iw-8156 {159r,02} ir dïen\$t wa\$ wol lon\$ wert.(.)
III·0·V_Iw-8157 {159r,03} ôvch wa_ene \$i|\$ al\$o genoz.,,
III·0·V_Iw-8158a {159r,04} daz \$i de\$ kv\omber\$ niht verdroz.(.)
III·0·V_Iw-8158b {159r,05} er londe ir nach frôvn \$a_elden bet.(.)
III·0·V_Iw-8158c {159r,06} Bvrge, lant, riche \$tet.
III·0·V_Iw-8158d {159r,07} machet er ir vndertan.(.)
III·0·V_Iw-8158e {159r,08} vn_- al\$ ich vernomen han.,,
III·0·V_Iw-8158f {159r,09} Si ne wart mit hirat niht bet%%ogen.

¹¹ Dazu siehe 4.4.3.

¹² Eine Ausnahme ist insbesondere dann zu machen, wenn für einen Verstext keine Edition vorliegt, wie etwa im Falle der “Rheinfränkischen Reimpredigten” (V·4b·v3).

4.4.2 Der handschriftenorientierte Ordnungsteil

Der handschriftenorientierte Ordnungsteil hat außer der Textsigle folgende Bestandteile:

1. die Foliozahl und unmittelbar daran angeschlossen die Buchstaben r oder v für “recto” und verso” oder der Seitenzahl (abhängig von der jeweiligen Handschrift und der im konkreten Fall üblichen Zählung);
2. ggf. unmittelbar angeschlossene Buchstaben a, b, c zur Unterscheidung mehrerer Kolumnen;
3. die Zeilenzahl, getrennt von der Folio- oder Seitenzählung durch ein Komma

Beispiele:

reine Folienzählung:

III·0·V_lw-**001r,14** {0014} Si iehent,, er lebe noch hîvte.(.)

III·0·V_lw-**001r,15** {0015} er hat den lop erworben.(.)

III·0·V_lw-**001r,16** {0016} išt im der lip er\$torben.,,

III·0·V_lw-**001r,17** {0017} So lebt doch ĩemer \$in nam.(.)

Folien- und Kolumnenzählung:

V·2·V_Türh-**062ra,08** {00069} Verloru mu\ezen \$ie ymmer we\$en

V·2·V_Türh-**062ra,09** {00070} Da kunde div gotheit gene\$en

V·2·V_Türh-**062ra,10** {00071} Da man die men\$heit begru\op

V·2·V_Türh-**062ra,11** {00072} Wol vns was \$ich vra_euden hu\op

Seitenzählung:

IV·4b·V_PrRei-**012a,07** *Cnv\o \$al eyn ieglich men\$che gvt

IV·4b·V_PrRei-**012a,08** bedrahten lazen \$inen mvt,

IV·4b·V_PrRei-**012a,09** wie vn\$er lieber h're quam

IV·4b·V_PrRei-**012a,10** zv\o dem%e er\$ten von Bethaniam.(.)

4.4.3 Der editionsorientierte Ordnungsteil

Der editionsorientierte Ordnungsteil folgt in der Zählung der jeweils maßgeblichen Ausgabe.¹³ Bei mehreren konkurrierenden Zählungen ist im Header anzugeben, nach welcher Ausgabe sich die Zählung richtet. Bei einer reinen **Verszählung** besteht die Zählung nur aus der Verszahl. Bei einer **Strophengliederung** (z. B. Nibelungenlied) oder einer anderen **mehrgliedrigen Zählung** (z. B. Lachmanns “Dreißiger” in der Zählung von Wolframs “Parzival”) werden die Strophenzahl / Einheitenzählung und die Verszählung , durch Kommata getrennt, angeführt. Bei **Prosatexten** besteht die Zählung aus der Seiten- und der Zeilenzahl der jeweils maßgeblichen Ausgabe.

Beispiele:

reine Verszählung:

III·0·V_lw-**0008** {001r,08} er hat bi \$inen zîten.

III·0·V_lw-**0009** {001r,09} Gelebt al\$o \$chône.,,

III·0·V_lw-**0010** {001r,10} daz er der êren krône.

III·0·V_lw-**0011** {001r,11} do trv\oc. vn\ noch \$in nam trêit.(.)

Strophenzählung:

III·0·V_Nib-**0033,1** Si liefen,, da \$i funden. ge\$atelt manic march.(.)

III·0·V_Nib-**0033,2** in|hofe {02r,25} Sigemundes. d' bvhurt wart \$o \$tarch.,,

III·0·V_Nib-**0033,3** daz man erdiezen horte. Palas vn\ Sal.(.)

III·0·V_Nib-**0033,4** {02r,26} die hochgemvten degene. heten vrolichen \$cal.(.)

Einheitenzählung:

III·0·V_Parz-**019,17** {010a,18} *{S*}waz da wa\$ volche\$ inne.,,

III·0·V_Parz-**019,18** {010a,19} mo\ere vn\ mo\erinne.

III·0·V_Parz-**019,19** {010a,20} wa\$ beidiv wip vn\ man.(.)

III·0·V_Parz-**019,20** {010a,21} der herre \$cho\vwen began.

¹³ Hierzu ist immer das “Verfasserlexikon” zu konsultieren!

In seltenen Fällen kann es vorkommen, daß eine Handschrift Verse in einer anderen Reihenfolge anordnet als die Edition des jeweiligen Textes. In diesen Fällen wird die Textreihenfolge der Edition hergestellt, aber die Textreihenfolge der Handschrift durch eine Doppelzählung angezeigt, bei der die Verszahl, die sich aus der Handschrift ergibt, der Verszahl der Edition in eckigen Klammern direkt angefügt wird, z. B. 1873[1874].

4.4.4 Die computerunterstützte Erstellung des Ordnungsteils

Bei der Texterfassung ist es nicht notwendig, in jeder Zeile den vollständigen Ordnungsteil manuell einzugeben; es ist ausreichend – und so sollte aus arbeitsökonomischen Gründen auch unbedingt verfahren werden! –, jede Zeile mit einem [TAB] und einem [BLANK] zu beginnen und ansonsten die Vers- bzw. die Seiten- und Zeilenzahlen in regelmäßigen Abstand anzugeben. Bei Verstexten hat es sich bewährt, alle 50 oder 100 Verse die Verszahl anzugeben; bei einer Seiten- und Zeilenzählung muß jeweils der Beginn der ersten Zeile einer neuen Seite markiert werden. Dabei ist darauf zu achten, daß die Zählung direkt am Zeilenanfang steht. Nach Abschluß der Texterfassung wird dann die Zählung durch das Programm ZAHL (siehe 18.21) vervollständigt. Die Korpusigle und das Textkürzel kann mit Hilfe des WordPerfect-Editors automatisch jeweils am Zeilenanfang ergänzt werden. Dazu ist folgender Befehl zu verwenden:

1. [Alt+F2] (= suchen und ersetzen)
2. "Mit Bestätigung?" "n"
3. Suche: "[FNZ]"
4. Ersetzen durch: "[FNZ]Ordnungsteil_Sigle"
5. [F2]

Auch die jeweils sekundäre Zählung sollte nicht komplett manuell eingegeben werden. Bei Prosatexten, die primär nach der Handschrift gezählt werden, ist es ausreichend, jeweils nur den Beginn einer neuen Seite durch die Seitenzahl in geschweiften Klammern anzuzeigen. Die Angabe des Zeilenumbruchs in der Edition nicht unbedingt notwendig.

Sofern sich die primäre Zählung nach der Edition richtet – so v. a. bei Verstexten –, muß in der gleichen Weise jeweils der Beginn einer neuen Seite der Handschrift durch die Seitenzahl in geschweiften Klammern angezeigt werden. Zusätzlich ist jeder Zeilenwechsel der Handschrift durch eine öffnende und schließende geschweifte Klammer {} zu markieren; nach der Umwandlung der Datei mit Editions-zählung in eine Datei mit Handschriftenzählung durch das Programm A-ZU-H (siehe 18.2) kann dann die Zählung ebenfalls durch das Programm ZAHL automatisch ergänzt werden.

4.5 Transkription und Codierung

4.5.1 Zeichenbestand im Korpus (mit Unicode-Dezimalzahl)

!"#\$%&'()*+,-./0123456789;<=>?@ABCDEFGHIJKLMN O PQRSTU VWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~\$`«°·»ÄÅÆÉÍÖÖÜßàáâäæèéëïíîðóôöùúû |

(Â, Ê, Î, Ô, Û, Ý werden als A\^, E\^, I\^, O\^, U\^, y\^ verschlüsselt.)

(Tab.)	9	*	42	6	54	B	66	N	78	Z	90
(Enter)	10	+	43	7	55	C	67	O	79	[91
(Leerz.)	32	,	44	8	56	D	68	P	80	\	92
!	33	-	45	9	57	E	69	Q	81]	93
"	34	.	46	:	58	F	70	R	82	^	94
#	35	/	47	;	59	G	71	S	83	_	95
\$	36	0	48	<	60	H	72	T	84	`	96
%	37	1	49	=	61	I	73	U	85	a	97
&	38	2	50	>	62	J	74	V	86	b	98
'	39	3	51	?	63	K	75	W	87	c	99
(40	4	52	@	64	L	76	X	88	d	100
)	41	5	53	A	65	M	77	Y	89	e	101

f	102	q	113		124	Ä	196	â	226	ó	243
g	103	r	114	}	125	Å	197	ä	228	ô	244
h	104	s	115	~	126	Æ	198	æ	230	ö	246
i	105	t	116	§	167	É	201	è	232	ù	249
j	106	u	117	“	168	Í	205	é	233	ú	250
k	107	v	118	«	171	Ó	211	ê	234	û	251
l	108	w	119	®	174	Ö	214	ë	235	ü	252
m	109	x	120	°	176	Ü	220	ì	236		9474
n	110	y	121	´	180	ß	223	í	237		
o	111	z	122	·	183	à	224	î	238		
p	112	{	123	»	187	á	225	ò	242		

4.5.2 Literatur zur Paläographie und Kodikologie

Die Transkription der in Handschriften des 12. bis 14. Jahrhunderts überlieferten Texte setzt grundlegende Kenntnisse der Paläographie und der Kodikologie voraus. Sowohl zur grundlegenden Orientierung als auch zur Klärung von Spezialfragen sind insbesondere – aber natürlich nicht unbedingt ausschließlich! – folgende Bücher heranzuziehen:

Kodikologie und Handschriftenbeschreibung:

Christine Jakobi-Mirwald: Buchmalerei. Ihre Terminologie in der Kunstgeschichte. Vollständig überarbeitete und erweiterte Neuauflage. Berlin 1997 [insbesondere Kap. 4 und 6].

Paläographie allgemein:

Bernhard Bischoff: Paläographie des römischen Altertums und des abendländischen Mittelalters. 2., überarb. Aufl. Berlin 1986 (Grundlagen der Germanistik 24).

Walter Heinemeyer: Studien zur Geschichte der gotischen Urkundenschrift. 2. Aufl., Köln 1982 (AfD Beiheft 4).

Karin Schneider: Paläographie und Handschriftenkunde für Germanisten. Eine Einführung. [Tübingen](#) 1999 (Sammlung kurzer Grammatiken germanischer Dialekte B, Ergänzungsreihe 8).

Paläographie deutscher Handschriften des Mittelalters:

Karin Schneider: Gotische Schriften in deutscher Sprache. Bd. 1: Vom späten 12. Jahrhundert bis um 1300. 1,1: Textband. 1,2: Tafelband. Wiesbaden, 1987

Abkürzungen:

Adriano Cappelli: Dizionario di abbreviature latine ed italiane. Usate nelle carte e codici specialmente del medio-evo riprodotte con oltre 14000 segni incisi con l'aggiunta di uno studio sulla brachigrafia medioevale, un prontuario di Sigle Epigrafiche, l'antica numerazione romana ed arabica ed i segni indicanti monete, pesi, misure, etc. Milano 1994

Jürgen Römer: Geschichte der Kürzungen. Abkürzungen in deutschsprachigen Texten des Mittelalters und der frühen Neuzeit. Göppingen 1997, ²1999 (GAG 645; zugl. Phil. Diss. Marburg u. d. T.: Die Verwendung von Kürzungen in deutschsprachigen Texten des Mittelalters und der frühen Neuzeit).

Weitere grundlegende Literatur zur Paläographie und Kodikologie ist im Internet unter folgenden Adressen zu finden:

<http://artsci.wustl.edu/~smcarey/Paleographywemsk.html>

<http://artsci.wustl.edu/~smcarey/codicologywemsk.html>

4.5.3 Transkription

Grundsätzlich müssen die Texte so handschriftengetreu wie möglich erfaßt und dabei alle linguistisch relevanten Merkmale eindeutig abgebildet werden. Der Ordnungsteil, die Kodierung von Sonderzeichen und die Beschreibung der kodikologischen Besonderheiten müssen so gehalten sein, daß sich die handschriftennahe Textwiedergabe auch aus späteren Bearbeitungs- und Auswertungszuständen der Quelle uneingeschränkt wiederherstellen läßt.

Um Kompatibilitätsprobleme zu vermeiden, werden für die Transkription des Textes und die Codierung von Sonderzeichen fast ausschließlich ASCII-Zeichen aus dem Bereich zwischen [AS-

CII 32] und [ASCII 126] und keine an eine bestimmte Software gebundenen Sonderzeichen verwendet; ausgenommen sind einzig aus praktischen Gründen die Umlaut-Graphien <ä>, <ö>, <ü> und <æ>.

4.5.4 Besondere Buchstabenformen

Nach den Buchstabenformen werden jeweils **Majuskeln** und **Minuskeln** unterschieden, nach Größe und Auszeichnung auch **Initialen**, **Initialmajuskeln** und **Versalien**. Nicht unterschieden wird zwischen normalem und überhöhtem Minuskel-<z> sowie zwischen normalem und rundem <r>, wohl aber wird differenziert zwischen <s> und <ſ>; letzteres wird in der Transkription durch das Dollarzeichen <\$> wiedergegeben.

4.5.5 Abkürzungen, Super- und Subskripte

Zur Information über die in (deutschsprachigen) Handschriften des Mittelalters verwendeten Abkürzungen sollte ggf. Jürgen Römers “Geschichte der Kürzungen” (siehe oben unter 4.5.1) herangezogen werden.

4.5.5.1 Zur Codierung von Abkürzungen, Super- und Subskripten

Alle Abkürzungen, Super- und Subskripte werden so genau wie möglich transkribiert. Für Akzente auf Vokalen werden die üblichen Tastaturkombinationen verwendet (<é, è, ê, æ> etc.); in allen anderen Fällen werden zur Darstellung entweder einzelne Zeichen oder Zeichenkombinationen benutzt, und zwar meistens durch eine Kombination aus dem jeweiligen Grundzeichen, gefolgt von dem Back-Slash \ oder dem Unterstrich _ oder dem Prozentzeichen % und schließlich dem Modifikator. Grundsätzlich ist die Zeichenfolge mit Backslash <x\y> zu lesen als “y über oder unter x”, die Zeichenfolge <x_y> als “x in Ligatur mit y”, die Zeichenfolge mit Prozentzeichen <x%y> als “x, danach hochgestellt y”.

4.5.5.2 Die Codierung einzelner Abkürzungen

Im einzelnen gelten für die Transkription von Abkürzungen folgende Codierungen:

æ	Ligatur <æ> (früher als <i>a_e</i> wiedergegeben).
d_e	Ligatur aus <d> und an der Oberlänge angebundenem <e>.
'	Kürzung von <er>, <re>, <ir>, <re>, <ri>, unabhängig von der graphischen Realisation in der Handschrift als <'>, < ^s >, < ^r >. Das Zeichen <'> wird auch dann verwendet, wenn das Zeichen in der Handschrift übergeschrieben ist; es wird dort eingefügt, wohin es sinnvollerweise gehört.
y\'	<y> mit Superskript (Punkt, Akzent, Trema), sofern dieses Superskript transkribiert werden soll. ¹⁴
\-	Nasalstrich und sämtliche Varianten mit dieser Funktion (vgl. dazu Römer).
\x	Buchstabe als Superskript. (Zu beachten ist, daß nur Vokalmodifikatoren als Superskripte transkribiert werden – etwa <i\e>, <u\o> –, alle anderen superskribierten Buchstaben als hochgestellt aufgefaßt und mit % angezeichnet werden.)
e\	für <i>ę</i> o. <i>Ä</i> . (<i>e</i> caudata).
∕	Schrägstrich durch die oder Modifikation der Unterlänge, v. a. bei <q> und <p> (<i>q</i> ∕ = <i>q</i> für <i>que</i> oder <i>qu</i> ; <i>p</i> ∕ = <i>p</i> für <i>pro</i>).

¹⁴ Superskripte über <y> können unberücksichtigt bleiben, wenn in einem Text jedes <y> mit einem Superskript versehen ist; in diesem Falle ist es ausreichend, auf die durchgehende Existenz des Superskriptes hinzuweisen und seine Form zu beschreiben (Akzent, Punkt, zwei Punkte). Wenn nur vereinzelt Superskripte über <y> geschrieben sind, könnte damit eine Funktion verbunden sein, und deshalb sind sie dann auch zu transkribieren.

- _ waagerechter Strich durch die Unterlänge, v. a. bei <p> (p_ = p für *per*, *prae*).
(Der Schrägstrich und der waagerechte Strich durch die Unterlänge von <p> sind graphisch oft nicht sicher zu unterscheiden, so daß die Differenzierung hier nach der Funktion erfolgen muß.)
- %x Hochstellung eines Buchstabens <x>; zu beachten ist, daß der Buchstabe – sofern irgend vertretbar – an der Stelle im Wort eingefügt wird, an die er sinnvollerweise gehört, auch wenn er in der Handschrift leicht nach rechts oder links versetzt ist.
- %%a Kürzung von <r> durch Superskription des Folgevokals, meist <a> (z. B. \$p%%ach = *sprach* = *sprach*).
- %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>.
- 3 Kürzung von auslautend <et> durch <3> (z. B. gewill3 = gewill3 für *gewillet*).
- 4 Kürzung von <rum>.
- \&1 variables Zeichen für eine Kürzung, für die keine verbindliche Codierung vereinbart ist. Im Dateikopf muß das Zeichen beschrieben und nach Möglichkeit in seiner Funktion bestimmt werden (möglichst unter Bezugnahme auf Römer). Gibt es in einer Handschrift mehrere außergewöhnliche Abkürzungen, müssen entsprechend mehrere Kürzungszeichen <\&1>, <\&2> usw. definiert werden.

Superskripte über <i> in Form eines Punktes oder Striches sollten zumindest dann berücksichtigt werden, wenn in einer Handschrift <i> mit und ohne Superskript vorkommen. Im Header einer Textdatei ist auf jeden Fall anzugeben, ob Superskripte über <i> berücksichtigt sind oder nicht; ggf. wird in der Datei dann zwischen <i> für den Buchstaben ohne Superskript und <i> für den Buchstaben mit i-Strich unterschieden. Wenn in einer Handschrift aber konstant nur eine Form, mit oder ohne Superskript vorkommt, kann einfach <i> geschrieben werden, vorausgesetzt, im Header wird die Buchstabenform beschrieben.¹⁵ Entsprechendes gilt für Superskripte über <j>.

4.5.5.3 Auflösung von Abkürzungen

Abkürzungen in deutschen Texten und Wörtern dürfen nicht aufgelöst werden, sondern müssen so genau wie möglich durch die entsprechende Codierung wiedergegeben werden. Innerhalb lateinischer Textabschnitte dürfen Abkürzungen in runden Klammern aufgelöst werden, sofern über die Art der Auflösung kein Zweifel besteht; sie müssen nicht aufgelöst werden. Bei der Auflösung von Abkürzungen sind ggf. die entsprechenden Nachschlagewerke (v. a. Bischoff, Cappelli; siehe 4.5.1) und natürlich die maßgebliche(n) Edition(en) des jeweiligen Textes zu konsultieren. Bei der Benutzung der Edition ist zu beachten, daß der Text der konkreten Handschrift vom Wortlaut der Edition abweichen kann; der Kontext muß immer berücksichtigt werden.

Alle aus der Auflösung der Abkürzungen gewonnenen Buchstaben müssen in runde Klammern gesetzt werden; alle ausgeschriebenen Buchstaben – auch solche, die Teil einer Ligatur sind oder durch Streichungen o. ä. modifiziert sind – bleiben entsprechend dem Usus der Paläographie außerhalb der Klammern (z. B. ¶ = p(ro); gl̄a = gl(ori)a).¹⁶

¹⁵ Dies gilt nicht für die Zeitabschnitte IV und V. Hier durfte auf die Transskription der i-Superskripte unabhängig von ihrem Auftreten verzichtet werden, da sie nur zur Unterscheidung des i von u-, m-, n-Strichen dienen können.

¹⁶ In älteren Dateien ist teilweise anders verfahren worden, indem auch die modifizierten Buchstaben in die Klammerung einbezogen worden sind (z. B. ¶ = (pro); gl̄a = g(lori)a). Diese Vorgehensweise ist aber in der Paläographie nicht üblich und soll daher nicht fortgeführt werden.

Die oft stark abgekürzten *nomina sacra* <ihc> und <xpc> sind in manchen Dateien in Rundklammern aufgelöst; die Flexionsendung steht dabei außerhalb der Rundklammern.¹⁷ Diese Auflösungen sollten im Interesse einer möglichst genauen Wiedergabe des Textes in Zukunft nicht vorgenommen werden, zumal auch bislang nicht einheitlich verfahren worden ist.

4.5.6 Codierung von Initialen, Initialmajuskeln, Versalien, Schriftgraden, Rubriken

Durch Größe und / oder Rubrizierung gegenüber der Textschrift hervorgehobene Buchstaben werden durch geschweifte oder runde Klammern, denen jeweils der Asterisk vorausgeht, codiert: *{X*}* oder *(X*)*. Dabei werden durch *{X*}* **Initialen**, durch *(X*)* **Initialmajuskeln**, Versalien und Lombarden gekennzeichnet.¹⁸ Unter Umständen sind nur durch Rubrizierung gekennzeichnete Versalien und Lombarden auf Rückvergrößerungen und Microfilmen nicht sicher zu erkennen; in diesen Fällen kann die Codierung entfallen, wenn im Header darauf hingewiesen wird.

Rubriken (üblicherweise in roter Tinte und oft in Auszeichnungsschrift ausgeführte Überschriften oder Quellenangaben)¹⁹ werden durch “+R” und “@R” umschlossen und damit markiert. Oftmals sind Rubriken in den Handschriften – aus ästhetischen Gründen? – nicht exakt an der Stelle geschrieben, an die sie von der sinnvollen Lesereihenfolge her gehören; beispielsweise stehen die letzten Wörter eines Kapitels am Ende einer Zeile, während am Anfang derselben Zeile bereits die Rubrik zur Einleitung des folgenden Kapitels steht. In diesen Fällen ist der Text in der sinnvollen Textreihenfolge zu erfassen; die Wortfolge der Handschrift muß in einem Kommentar (“K-Klammer”; siehe 4.5.9) erläutert werden.

4.5.7 Handschriftliche Interpunktion

Interpunktionszeichen der Handschrift werden so genau wie möglich wiedergegeben und folgendermaßen codiert:

- . tiefer Punkt
- % halbohoher Punkt
- .\^ punctus elevatus
- .\' punctus elevatus
- .\7 punctus elevatus
- = handschriftliches Trennzeichen, mit dem Worttrennung am Zeilenende markiert ist (unabhängig von der Form des Trennzeichens in der Handschrift als waagerechter einfacher oder doppelter Strich)

Verwendet eine Handschrift ausschließlich den halbhohen Punkt, kann eine Transkription auch durch den normalen Punkt erfolgen; in der Metadatenbank ist dann ein entsprechender Hinweis zu geben.

Die Interpunktionszeichen werden ohne [BLANK] direkt an das Wort angeschlossen, auf das sie in der Handschrift folgen. Eine Ausnahme sind einzig die Punkte, die Wörter umschließen, die aus nur einem Buchstaben bestehen (v. a. Abkürzungen wie <.f.> für *sanctus*, *sancte* und das Wort *ê* <.e.> “Gesetz”); hier darf der Punkt, der dem Wort vorangeht, nicht an das vorige Wort angeschlossen werden, sondern muß ohne [BLANK] dem betreffenden Wort vorangestellt werden (Beispiel: *diu .e.*, nicht *diu. e.*; *do \$prach .\$. Martinus*, nicht *do \$prach. \$. Martinus*). In jedem Falle gilt, daß Interpunktionszeichen niemals isoliert ohne direkten Anschluß an ein Wort stehen dürfen. Insbesondere ist darauf zu achten, daß bei der Einfügung einer Shift-Klammer Wort und Interpunktionszeichen nicht (versehentlich) getrennt werden.

¹⁷ Im einzelnen ist wie folgt verfahren worden: ih\~s = (Jesu)s; ih\~c = (Jesus); ih\~u = (Jes)u; ih\~m = (Jesu)m; xp\~c = (Christus); xp\~s = (Christu)s; xp\~o = (Christ)o; xp\~i = (Christ)i; xp\~m = (Christu)m. Außerdem: isrl\~ = (Israel); irlm\~ = (Jerusalem).

¹⁸ Zur Terminologie vgl. Jakobi-Mirwald; siehe 4.5.1.

¹⁹ Zur Terminologie vgl. Jakobi-Mirwald; siehe 4.5.1.

Moderne Interpunktionszeichen werden ebenfalls immer direkt an ein Wort angeschlossen; folgen auf ein Wort sowohl ein handschriftliches als auch ein modernes Interpunktionszeichen, so steht in jedem Fall das moderne hinter dem handschriftlichen (zur modernen Interpunktion siehe ansonsten 5.4).

4.5.8 Schwer lesbare, nicht lesbare und verlorene Buchstaben und Wörter

Einzelne Buchstaben oder ganze Wörter, die in der Handschrift (bzw. auf dem Mikrofilm oder den Rückvergrößerungen) nur teilweise oder schwer lesbar sind (z. B. Verlust von Ober- oder Unterlängen durch Beschnitt, Beschädigung von Buchstaben durch Abrieb, Löcher im Pergament oder Flecken), aber doch noch mit hinreichender Sicherheit identifiziert werden können, werden in Spitzwinkelklammern <...> gefaßt.

Buchstaben oder Wörter, die gar nicht mehr lesbar sind, aber entweder aus dem Kontext erschlossen werden können oder aus einer früheren Edition bekannt sind, werden in einfache eckige Klammern [...] gesetzt. Sofern eine Ergänzung nicht möglich ist, werden entsprechend der (vermuteten) Anzahl nicht lesbarer Buchstaben Punkte gesetzt. Sind u. U. ganze Zeilen nicht lesbar, die nicht ergänzt werden können, reicht es aus, pro Zeile drei Punkte in eckigen Klammern zu setzen; im Header sollte aber auf diese Vorgehensweise explizit hingewiesen werden.²⁰

Buchstaben oder Wörter, die durch Beschnitt verloren gegangen sind, aber aus dem Kontext erschlossen oder aus einer Edition übernommen werden können, werden in doppelten eckigen Klammern [[...]] ergänzt. Ist eine Ergänzung nicht möglich, werden die verlorenen Buchstaben oder Wörter durch Punkte vertreten. Ebenfalls in doppelten eckigen Klammern [[...]] werden Wörter oder Wortteile ergänzt, die in der Handschrift (versehentlich) ausgelassen worden sind, aber für die grammatikalische Auswertung unabdingbar sind; dies betrifft v. a. fehlende Reimwörter, syntaktisch unbedingt notwendige Konstituenten (Subjekt, Prädikat, Präpositionen, etwa *do [[er]] diz wort gesprach*) sowie fehlerhaft unvollständige Wortformen, die andernfalls gar nicht lemmatisierbar wären (z. B. *na[[m]]en*). Derartige Ergänzungen sind aber äußerst zurückhaltend vorzunehmen und ggf. mit den wissenschaftlichen Projektmitarbeitern abzusprechen. In einem Kommentar (E-Klammer, siehe 4.5.10) wird ggf. erläutert, woher die jeweilige Ergänzung stammt (vgl. dazu 4.5.10).

4.5.9 Kommentierung und Dokumentation diverser handschriftlicher Befunde

Sämtliche handschriftlichen Befunde, die möglicherweise von Bedeutung sind, aber nicht durch die Transkription dargestellt werden können oder aber zusätzlich zur Transkription der Erläuterung bedürfen, werden im laufenden Text dokumentiert und / oder kommentiert. Dieser Kommentar wird zur Unterscheidung von der Transkription gekennzeichnet, indem er durch "+K" eröffnet und mit "@K" geschlossen wird. Formal entspricht eine solche "K-Klammer" der Markierung der Rubriken mit dem Unterschied, daß Text, der durch "+R" und "@R" eingeschlossen wird, der weiteren Auswertung zur Verfügung steht, während die in "+K" und "@K" gefaßten Kommentare von den automatisierten Auswertungsprozeduren ignoriert werden.

K-Klammern beinhalten Kommentare zu objektiven kodikologischen und paläographischen Gegebenheiten der Handschrift, so etwa Erläuterungen zu Streichungen oder Ergänzungen, Textverlust durch Beschnitt oder sonstige Beschädigung, Bemerkungen zu etwaigem Schreiberwechsel.

²⁰ In manchen älteren Textdateien ist die Codierung nicht lesbarer Buchstaben und Wörter teilweise stärker differenziert. Text, der von früheren Herausgebern noch gelesen werden konnte, gegenwärtig aber nicht mehr lesbar ist, erscheint in eckigen und Spitzklammern [<...>]. Text, der auf dem Mikrofilm oder den Rückvergrößerungen nicht zu sehen ist, aber in der Handschrift vermutlich noch vorhanden und lesbar ist, ist in eckige und runde Klammern [(...)] gesetzt.

4.5.9.1 Zur Platzierung und Gestaltung der K-Klammern

Die K-Klammer sollte sich, wenn irgend möglich, auf das unmittelbar vorangehende Wort beziehen. Dieses Bezugswort muß zu Beginn der Klammer nach “+K” und davon durch einen [BLANK] getrennt wiederholt werden. Die Wiederholung des Bezugsworts soll sicherstellen, daß die Klammer auch wirklich an der richtigen Stelle steht; sie erleichtert zudem eine ggf. später notwendige automatische oder makrogestützte Einfügung der Klammern in einen aus dem NF-Index rückerzeugten Fließtext. Die eigentlichen Erläuterungen folgen dann nach einem Doppelpunkt.

Beispiel:

swer an rechte güe[te] +K güete: teilweise nicht lesbar wegen Tintenfleck @K

Bezieht sich die Klammer nicht auf das Bezugswort selbst, sondern auf einen diesem unmittelbar folgenden Handschriftenbefund, wird dies durch den Unterstrich <_> angezeigt.

Beispiel:

swer an rechte +K rechte: _ Loch im Pergament @K güete

Wird ausnahmsweise auf einen Befund vor dem Bezugswort verwiesen (dies sollte nur der Fall sein, wenn das Bezugswort das erste Wort der Zeile ist), so wird darauf durch das Wort “davor” hingewiesen.

Beispiel:

swer +K swer: davor {an} gestr. @K an rechte güete

Mit Ausnahme des Bezugswortes, das ausreichend dadurch gekennzeichnet ist, daß es unmittelbar auf “+K” folgt und von dem eigentlichen Kommentar durch einen Doppelpunkt getrennt ist, sollten mhd. Wörter oder auch einzelne Buchstaben in der Klammer zur besseren Unterscheidung in geschweifte Klammern {...} gesetzt werden.

Beispiele:

swer an rechte +K rechte: {h} üdZ @K güete

swer an rechte +K rechte: auf rad. {edele} @K güete

4.5.9.2 Empfohlene Abkürzungen für häufige Bemerkungen in K-Klammern

Die Informationen in den K-Klammern sollen zwar knapp, aber zugleich auch eindeutig und für alle späteren Bearbeiter verständlich formuliert sein. Für häufig zu beschreibende Befunde sollten daher folgende standardisierten Abkürzungen verwendet werden:

üdZ	über der Zeile ergänzt
üdZ*	über der Zeile ergänzt mit Verweiszeichen
untZ	unter der Zeile ergänzt
untZ*	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

Beispiele:

swer an rechte +K rechte: üdZ @K güete

swer an rechte +K rechte: < {rechten}, {n} rad. @K güete

swer an rechte +K rechte: aR* @K güete

swer an rechte +K rechte: < {rechse} @K güete

4.5.10 Emendationen und Kommentare

Für das Textverständnis (möglicherweise) wichtige Zusätze, Streichungen, Konjekturen oder Kommentare der / einer Edition (oder auch eines Projekt-Mitarbeiters) werden ebenfalls in eine "Shift-Klammer" gefaßt, für die die gleichen Regelungen gelten wie für die K-Klammer mit dem einzigen Unterschied, daß sie mit "+E" eröffnet und mit "@E" geschlossen wird. Zusätze zu dem handschriftlich überlieferten Text erscheinen in der Regel nur in der E-Klammer, es sei denn, diese Zusätze wären für das Textverständnis oder die grammatikalische oder syntaktische Auswertung unbedingt erforderlich; in diesem Fall müssen sie in doppelten eckigen Klammern [[...]] in den transkribierten Text inseriert werden (vgl. dazu 4.5.8).

In E-Klammern sind folgende Abkürzungen häufiger gebraucht:

str.	streichen!
str. NN	streicht NN
l.	lies
l. NN	liest NN

Beispiele:

ein guoter +E guoter: str. Lachmann @E ritter so geleret was

ein rtter +E rtter: l. {ritter} Lachmann @E so geleret was

Einzelne Editionen und Handschriften sollen nach Möglichkeit durch Siglen bezeichnet werden, die im Header der betreffenden Datei aufzulösen sind. Bei abweichenden Lesungen oder Konjekturen einer Edition kann auf "l." in der Regel verzichtet werden; meist reicht es aus, die abweichende Lesung und direkt dahinter die Sigle der Edition oder der Handschrift anzuführen.

Beispiel:

swer an rehte êre +E êre: {güete} Lachmann @E

4.5.11 Tilgungen und andere eindeutige Fehler des handschriftlichen Textes

In der Handschrift durch Rasur, Unterpungierung, Unterstreichung (weit eher als Durchstreichung²¹) getilgte Wörter und Buchstaben werden – sofern noch lesbar – transkribiert und von eckigen Klammern, denen jeweils ein Asterisk vorangeht, eingefasst *[X*], gegebenenfalls gefolgt von den sie ersetzenden Wörtern oder Buchstaben. Der genaue paläographische Befund wird in einer unmittelbar folgenden K-Klammer erläutert.

Beispiel:

spr*[u*]ach +K spr*[u*]ach: {u} unterpung. @K

Wörter, die in der Handschrift überliefert sind, aber dem Sinn oder der Syntax nach überflüssig sind – in der Regel handelt es sich um Doppelschreibungen –, werden übernommen; in einer E-Klammer wird aber darauf hingewiesen, daß das betreffende Wort überflüssig ist.

Beispiel:

swer an rehte rehte +E rehte: doppelt; überflüssig @E güete

4.5.12 Nachträge, Verweiszeichen, Korrekturen

Nachträgliche Ergänzungen in der Handschrift *in margine* oder interlinear werden in den laufenden Text inseriert, sofern die fraglichen Ergänzungen von der Hand stammen, die auch den Haupttext geschrieben hat. In einer K-Klammer ist jeweils die tatsächliche Anordnung des Textes zu dokumentieren. Stammen Ergänzungen oder Nachträge von einer anderen oder gar späteren Hand, so werden sie nur in K-Klammern dokumentiert, aber nicht in den laufenden Text integriert.

²¹ Durchstreichung ist in mittelalterlichen Handschriften in aller Regel nicht Markierung einer Tilgung, sondern Mittel der Hervorhebung.

4.6 Die Schlußüberprüfung

Nach Abschluß der Texterfassung und der Kollationierung – in Bonn erst nach der Präeditierung (siehe 5) werden die Textdateien durch den Leiter der jeweiligen Arbeitsstelle oder ggf. durch einen der wissenschaftlichen Mitarbeiter daraufhin überprüft, ob die Texterfassung so genau und korrekt wie möglich erfolgt ist und ob alle relevanten Projektkonventionen beachtet und umgesetzt worden sind. Es hat sich bewährt, stichprobenartig etwa 10 Prozent der betreffenden Datei zu überprüfen. Die zu überprüfenden Abschnitte sind dabei gleichmäßig aus der gesamten Datei auszuwählen.

5 Präeditierung

An die Texterfassung schließt sich als nächster Arbeitsschritt die Präeditierung an²², d. h. die Aufbereitung der Transkription für die dann folgende computerunterstützte Lemmatisierung und grammatikalische Bestimmung. Zur Präeditierung gehören die Überprüfung und ggf. Kennzeichnung der Worteinheiten, die Markierung von Partikelverben, mehrgliedrigen Konjunktionen und fremdsprachigen Wörtern sowie die Interpunktion des Textes. Die nachstehende Übersicht orientiert über die möglichen Präeditierungszeichen, die unten näher erläutert werden:

- | Trennung von in der Handschrift zusammengeschriebenen Wörtern²³
- # Zusammenfügung von in der Handschrift getrennt geschriebenen Wortteilen oder Wörtern
- (=) fehlende Markierung der Worttrennung am Zeilenende
- =| überflüssiges handschriftliches Trennzeichen am Zeilenende
- *1 Adverb als Teil eines Partikelverbs
- *2 Verb als Teil eines Partikelverbs
- ~ mehrgliedrige Konjunktion
- ~| mehrgliedrige Konjunktion, in der Handschrift zusammengeschrieben
- ~(=) mehrgliedrige Konjunktion, Bestandteile durch Zeilenumbruch getrennt
- *f fremdsprachliches Wort (in mhd. Satz eingebaut); dieses *f erscheint in den Indizes als dem Lemmaansatz hinzugefügter *.²⁴ Daneben +L ... @L. Vgl. 5.3.

5.1 Worteinheiten

Die Auffassung der mittelalterlichen Schreiber davon, was als Worteinheit zu betrachten ist und dementsprechend durch Spatien von anderen Wörtern abgetrennt wird, geht zwar in aller Regel mit unseren heutigen Auffassungen konform, aber nicht immer. Unterschiede bestehen – je nach Handschrift in unterschiedlichem Grade – insbesondere bei komplexeren Wortbildungen und Komposita, präfigierten Bildungen (etwa *ver lâzen* statt *verlâzen*), Präpositionen und nachfolgenden Nomina (etwa *indaz hus* statt *in daz hûs*) sowie bei flektierten Verbformen und nachgestellten Personalpronomina (etwa *sagestu* statt *sagest du*).

Voraussetzung für die Lemmatisierung ist daher als erstes die Überprüfung der handschriftlichen Worttrennung und Zusammenschreibung, die ggf. so gekennzeichnet werden müssen, daß einerseits der handschriftliche Befund noch immer klar erkennbar ist, aber andererseits auch die Wortgrenzen nach moderner Auffassung bzw. entsprechend den Projektkonventionen markiert sind. Als Richtschnur dienen dabei die Ansätze von Lexers "Mittelhochdeutschem Handwörterbuch".

²² Erfahrene Bearbeiter können die Präeditierung teilweise – insbesondere die Markierung von Getrennt- und Zusammenschreibung – bereits während der Texterfassung und Kollationierung durchführen, allerdings ohne daß dadurch die Präeditierung als eigener Arbeitsschritt überflüssig würde.

²³ Ein eventuelles Interpunktionszeichen muß dem Trennstrich | vorangehen.

²⁴ In älteren Indexfassungen steht *f am Anfang der Belegform.

5.1.1 Trennung von in der Handschrift zusammengeschriebenen Wörtern

Wörter, die in der Handschrift nach moderner Auffassung bzw. entsprechend den Projektkonventionen "fehlerhaft" zusammengeschrieben sind, werden durch den senkrechten Strich | (Tastatur: [AltGr+<]) voneinander getrennt (z. B. *in|daz*, *en|wolde*, *lebet|er*, *maht|u*).²⁵ Sofern zwischen die beiden durch den senkrechten Strich | getrennten Wörter zusätzlich Interpunktionszeichen eingefügt werden, so müssen diese vor dem Trennstrich gesetzt werden (etwa *so,,|daz*).

5.1.2 Krasis

Sofern direkt aufeinander folgende Wörter nicht nur ohne Spatium geschrieben sind, sondern miteinander verschmolzen sind, sind besondere Trennungsregeln zu beachten:

Gehört der **Vokal im Bereich der Krasis** eindeutig (auch) zu dem zweiten Wort, so wird vor ihm getrennt; gehört er nur zu dem ersten Wort, wird nach ihm getrennt.

Beispiele:

sageter → *saget|er* (< *sagete er*)

zen → *z|en* (< *ze den*)

aber: *zun* → *zu|n* (< *zuo den*)

Gehört ein **Konsonant im Bereich der Krasis** eindeutig (auch) zu dem ersten Wort, so wird nach ihm getrennt; gehört er nur zur zweiten Form, so wird vor ihm getrennt.

Beispiel:

alsi → *als|i* (< *alse si*)

Besondere Obacht ist bei der 2. Pers. Sg. mit folgendem Personalpronomen *du* / *tu* vonnöten:

- Ein *d* wird zum Personalpronomen gezählt.
- Bei einem *t* ist zu unterscheiden:
 - Bei den Präterito-Präsentien gehört das <t> auf jeden Fall zur Flexionsendung, beim verbum substantivum *bist* und bei *wil(t)* 'willst' ziehen wir es auch nach links.
 - Sonst zählen wir das *t* zum Personalpronomen; begründend ist hier die frmhdt. und md. Bildung der 2. Pers. Sg. auf –es (im Obd. eher *est*).

Beispiele:

mahtu → *maht|u* (< *maht du*); entsprechend auch *weist|u*, *muost|u*, *solt|u*, *bist|u*, *wilt|u* etc.

aber: *wildu* → *wil|du* 'willst du' und *sagestu* → *sages|tu* (< *sages(t) du*) und *nimes|tu*, *welles|tu*.

5.1.3 Zusammenfügung von in der Handschrift (fehlerhaft) getrennten Wörtern

In der Handschrift getrennt geschriebene, aber nach modernem Verständnis und / oder Projektkonventionen zusammengehörige Wortbestandteile werden durch das "Gatter" # zusammengefügt (z. B. *ge#lebet* für *ge lebet*).

5.1.4 Komposita

Komposita werden dann angesetzt und ihre Bestandteile bei der Präeditierung durch # als zusammengehörig markiert, wenn entweder bereits bei "Lexer" ein entsprechender Ansatz vorhanden ist oder wenn eine Konstruktion vorliegt, bei der das Erstglied nicht mehr als vorangestellter Genitiv aufgefaßt werden kann und die eindeutig nicht anders als ein Kompositum aufgefaßt werden kann.

Beispiele:

hërze-küniginne: Kompositum, da bei "Lexer" angesetzt; zudem keine Genitivfügung

hërzen küniginne: kein Kompositum, da nicht bei "Lexer"; Genitivfügung

²⁵ Die in alten Dateien eventuell vereinzelt noch vorkommende Abtrennung von Klitika durch das Gleichheitszeichen = anstelle des senkrechten Strichs wird nicht fortgesetzt; das Gleichheitszeichen muß ggf. durch den Strich | ersetzt werden.

hërzen-ouge: Kompositum, da bei "Lexer" angesetzt (obwohl Genitivfügung!)

5.1.5 Worttrennung am Zeilenende

Sofern Wortbestandteile durch einen Zeilenumbruch in der Handschrift getrennt sind, aber ihre Zusammengehörigkeit bereits in der Handschrift durch ein Trennzeichen – welcher Form auch immer – markiert wird, so wird dies durch das Gleichheitszeichen = angezeigt. Ist in der Handschrift kein Trennzeichen gesetzt, erfolgt die Zusammenfügung durch das Gleichheitszeichen in runden Klammern (=).

Sofern durch ein Trennzeichen in der Handschrift über die Zeilengrenze hinweg Wörter als zusammengehörig gekennzeichnet werden, die nach den Projektkonventionen nicht als ein Wort aufgefaßt werden, so wird dies durch das Gleichheitszeichen in Verbindung mit dem senkrechten Strich =| codiert. Das Gleichheitszeichen zeigt die Existenz des Trennzeichens in der Handschrift an, während der senkrechte Strich für die nachfolgenden Programme die Wörter als getrennt markiert.

Die Kennzeichnung der Worttrennung am Zeilenende bzw. der Zusammengehörigkeit von Wortbestandteilen auch über das Zeilenende hinaus ist die entscheidende Voraussetzung dafür, daß das Programm LEM bei der Erstellung des WOK-Indexes die zusammengehörigen Teile zusammenführen kann. Dabei wird ein Wort immer der Zeile zugerechnet, in der sein erster Bestandteil steht.

5.1.6 Partikelverben (trennbare Präfixverben)

Unter Partikelverben oder trennbaren Präfixverben versteht man die Verbindung eines Verbs mit einem (Präpositional-)Adverb – etwa *abe, ane, bî, ûf* –, das dem Verb in Kontaktstellung unmittelbar vorangehen kann, aber auch in Fernstellung dem Verb nachfolgen kann. Trotz der möglichen Fernstellung besteht eine feste syntaktische und semantische Einheit von Basisverb und Partikel (nhd. *jdn. ansprechen - er spricht jdn. an*). Dabei sind als Partikeln alle Adverbien möglich, zu denen es eine gleichlautende Präposition gibt (also *ane, abe, zuo, durch* ...), außerdem *nider* und *wider*.

Um die Partikelverben / trennbaren Präfixverben von festen Präfixbildungen zu unterscheiden und um der möglichen Fernstellung von Basisverb und Partikel Rechnung zu tragen, können die Konstituenten nicht durch das Gatter # zusammengefügt werden, sondern sie müssen einerseits als selbständige Wörter behandelt, aber zugleich auch als zusammengehörig gekennzeichnet werden. Dementsprechend wird bei der Vorlemmatisierung jedem der beiden Bestandteile, der Partikel wie dem Basisverb, ein eigener Lemmaansatz zugeordnet; bei einem späteren Bearbeitungsschritt werden dann die beiden Lemmaansätze halbautomatisch als Teile eines Partikelverbs gekennzeichnet (dazu siehe unten 10.3).

Die Kennzeichnung der beiden Konstituenten des Partikelverbs erfolgt durch die Anfügung von "*1" an die Partikel und von "*2" an das Verb.

Beispiele:

*dô er den brief ane*1 sach*2*

*er sach*2 den brief ane*1*

Sofern Partikel und Basisverb in Kontaktstellung stehen und als ein Wort geschrieben sind, wie im Nhd. üblich, muß zunächst die Partikel vom Basisverb durch den Trennstrich | abgetrennt werden; anschließend werden Partikel und Basisverb mit "*1" und "*2" markiert. Dabei ist darauf zu achten, daß "*1" vor dem Trennstrich steht (z. B. *ane*1/sehen*2*, nicht *ane/*1sehen*2*).

Ferner darf kein Partikelverb angesetzt werden, wenn die Partikel Bestandteil eines Pronominaladverbs (z. B. *dâr ane, dâr nâch, dar zuo*) ist. Die Pronominaladverbien lassen sich daran erkennen, daß eine Umformung in eine kontextentsprechende Präpositionalphrase möglich ist.

Beispiele:

ein teppet ... da sie vf solde sizzen: kein Partikelverb *ûf sitzen*, sondern Pronominaladverb *dâr ûf* (umformbar zu *sie solde sitzen ûf einem teppete*)

dar vf saz d' cune degen: kein Partikelverb, sondern Pronominaladverb (Umformung: *der küene degen saz dâr ûf* [sc. *ûf dem ors*])

wilt vnde zam man da vure truch: Partikelverb *vüre tragen* (keine Umformung möglich)

Nicht immer ist eine Unterscheidung eindeutig möglich, wenn ein Adverb sowohl Teil des Pronominaladverbs als auch als trennbare Verbalpartikel aufgefaßt werden kann.

Beispiel:

da bleip der greue inne: sowohl *dâr inne* als auch *inne*1 belîben*2* möglich.

In solchen Zweifelsfällen wird **kein Partikelverb** angesetzt.

5.2 Mehrgliedrige Konjunktionen

Mehrgliedrige Konjunktionen werden von Komposita unterschieden, indem ihre Konstituenten nicht durch das Gatter #, sondern durch die Tilde ~ verbunden werden (z. B. *sît~daz*). Sind die Konstituenten in der Handschrift nicht durch ein Spatium getrennt, wird durch den senkrechten Strich die Wortgrenze markiert, aber durch die Tilde die Zusammengehörigkeit angezeigt. Dabei ist darauf zu achten, daß die Tilde ~ vor dem Trennungsstrich | geschrieben wird (z. B. *sît~/daz*). Falls die Konstituenten einer mehrgliedrigen Konstruktion durch das Zeilenende voneinander getrennt sind, muß aus programmtechnischen Gründen zusätzlich zur Tilde das geklammerte Gleichheitszeichen verwendet werden (z. B. *sît~(=)daz*).

Ob eine mehrgliedrige Konjunktion anzusetzen sei oder nicht, läßt sich in der Lemmaliste (unter alt y) prüfen. — Die Ansetzung der mehrgliedrigen Konjunktionen richtet sich nach der „Mittelhochdeutschen Grammatik“ von Hermann Paul in der Bearbeitung durch Hugo Moser und Ingeborg Schröbler; dort sind folgende mehrgliedrige Konjunktionen verzeichnet: *biz~daz*, *dâr~mite~unde*, *dâr~wider~unde*, *die~wîle~daz*, *durch~daz*, *ê~daz*, *ê~danne~daz*, *vür(e)~daz*, *innen~dës*, *noch~denne~daz*, *nû~daz*, *sît~daz*, *sît~dô*, *sô~iè*, *ûf~daz*, *umbe~daz*, *under~dës*, *unz~daz*.²⁶ Eine weitere mehrgliedrige Konjunktion, die bei Moser und Schröbler nicht genannt wird, ist *wan~daz*. Darüberhinaus dürfen weitere mehrgliedrigen Konjunktionen nur nach Absprache angesetzt werden.

5.3 Fremdsprachliches

Einzelne **fremdsprachliche Wörter und Namen** – in aller Regel lateinische oder altfranzösische, ganz selten griechische –, werden bei der Präeditierung durch vorangestelltes *f markiert. Alle durch *f markierten Wörter werden von den Programmen zur Vorlemmatisierung und Indizierung ignoriert; ihnen werden dann keine mhd. Lemmata oder grammatische Bestimmungen zugewiesen. Diese Regelung gilt jedoch nur für fremdsprachliche Wörter oder Namen, die syntaktisch in einen mhd. Kontext integriert sind. Syntaktisch eigenständige fremdsprachliche Phrasen, Sätze oder Satzfolgen werden bei der Präeditierung dagegen in L-Klammern gesetzt (Beginn: +L, Ende @L). Bei der Tokenisierung der Textdatei durch LEM (1) werden alle Wörter innerhalb von L-Klammern automatisch durch ein ° am Anfang der Wortform gekennzeichnet.²⁷

In den annotierten Indizes wird dem Lemma *a n s a t z* ein * hinzugefügt, z. B. *esse**,²⁸ wenn die Wortform mit *f oder ° beginnt. *f wird getilgt; ° bleibt als Kennzeichen der fremdsprachli-

²⁶ Vgl. Hermann Paul: *Mittelhochdeutsche Grammatik*. 21., durchgesehene Auflage. Bearb. von Hugo Moser und Ingeborg Schröbler. Tübingen 1975 (Sammlung kurzer Grammatiken germanischer Dialekte A 2), S. 494f.

²⁷ L-Klammern haben also eine doppelte Funktion: 1. Sie kennzeichnen geschlossene fremdsprachliche Textabschnitte. 2. Sie erleichtern die Präeditierung, da sonst alle Wörter solcher Abschnitte einzeln durch *f markiert werden müssten.

²⁸ Fremde Eigennamen bekommen in den jetzigen Indizes wie sonstige Fremdwörter ein * nach dem Lemma.

chen Sätze oder Absätze erhalten. Für die Annotationsarbeit bedeutet dies: Alle mit ° beginnenden Wortformen werden nicht annotiert, alle anderen – auch die, deren Lemma durch finalen * als fremdsprachlich ausgewiesen ist – werden annotiert.

Als Beispiel für die Markierung von Lateinischem durch *f bzw. durch L-Klammern diene ein Abschnitt aus Willirams Hohelied-Paraphrase:

I·O·P_Will-03r,22 {007G,03} *(U*)uír ne uuóllen nîeth uergézzan,,
 I·O·P_Will-03r,23 daz díu gnâda dîne\$ {007G,04} *feuangelii
 I·O·P_Will-03r,24 \$ûozer e\$st. dânnē díu *fau\$te(=
 I·O·P_Will-03r,25 rita\$ déro êuuo.(.) {007G,05} *fLex,, díu quît.(.)
 I·O·P_Will-03r,26 +L "Si|q(ui)\$ hoc (uel) illud fecerit.,, morte
 I·O·P_Will-03r,27 moriat(ur).(.)" @L {007G,06} Dú quíde\$ ábo in dî(=
 I·O·P_Will-03r,28 nemo *fe\,uang(eli)o.(.) +L "Non ueni {007G,07} uo(=
 I·O·P_Will-03r,29 care iu\$to\$. \$ed peccatore\$.)(.)" @L
 I·O·P_Will-03r,30 {008G,01} *{D*}íe réhton. mínnonť díh.(.)

 I·O·P_Will-03v,03 {009G,01} *{I\}*}h bín \$álo. \$ámo dîe héré(=
 I·O·P_Will-03v,04 berga cedar: unte bín ábo
 I·O·P_Will-03v,05 {009G,02} uuâthlîch. \$ámo díu gezélt
 I·O·P_Will-03v,06 *f\$alomoni\$.)(.) Cedar,, *fqui *fin(=
 I·O·P_Will-03v,07 terp(re)tatur {009G,03} *ftenebre\,, er uuá\$
 I·O·P_Will-03v,08 *fi\$maheli\$ \$ún.,, uóne demo
 I·O·P_Will-03v,09 {009G,04} *fi\$mahelite\, cúman \$ínt.,,
 I·O·P_Will-03v,10 die der hû\$er ne|hábent.
 I·O·P_Will-03v,11 \$unter {009G,05} ókkeret uílz#hû\$.
 I·O·P_Will-03v,12 unte ándera únuuatlîche
 I·O·P_Will-03v,13 hereberga.(.) {009G,06} Mít cedâr \$ínt
 I·O·P_Will-03v,14 *ffilii *ftenebrarum bezêichenet.,,
 I·O·P_Will-03v,15 uon dén {009G,07} íh mih chlágon.(.)

Als Faustregel kann dienen, dass in L-Klammern maximale Folgen von lateinischen Wörtern gesetzt werden, die an Anfang und Ende von punktwertigen Interpunktionszeichen begrenzt sind, also von (.) (: (!) (?). Daher steht der in einen mhd. Satz integrierte Relativsatz *qui interpretatur tenebre* (03v,06f.) nicht in einer L-Klammer.

5.4 Interpunktion

5.4.1 Moderne Interpunktionszeichen und Interpunktionsregeln

Für die moderne Interpunktion eines Textes können die modernen Interpunktionszeichen Punkt, Fragezeichen, Ausrufezeichen, Semikolon, Doppelpunkt, Komma und Anführungszeichen verwendet werden, außerdem die öffnende und schließende runde Klammer. Nicht möglich ist die Verwendung des Geviertstrichs ("Gedankenstrich"). Zur Unterscheidung von möglichen handschriftlichen Interpunktionszeichen werden die modernen Interpunktionszeichen bis auf das Komma und das Anführungszeichen allerdings in runde Klammern gesetzt. Folgende Interpunktionszeichen sind also möglich:

(.) (: (!) (?) (: , , ,) , (,) “

Die Interpunktion folgt im wesentlichen den gleichen Regeln, die auch in der neuhochdeutschen Schriftsprache (nach der traditionellen Rechtschreibung und Interpunktion!) gelten; eine Besonderheit ist die im folgenden erläuterte Differenzierung zwischen einfachem und doppeltem Komma. Außerdem ist darauf zu achten, daß der Punkt, das Semikolon und der Doppel-

punkt nur vollständige Sätze abschließen; innerhalb eines Satzes zur Kennzeichnung von Oberbegriff und folgender Aufzählung steht der Doppelpunkt nicht zur Verfügung.

5.4.2 Doppelkomma, einfaches Komma, kein Komma

Die Unterscheidung eines einfachen und eines doppelten Kommas ist notwendig, um die Grenzen von Elementarsätzen eindeutig markieren zu können; dies ist die Voraussetzung dafür, daß die Programme bei der grammatischen Bestimmung wenigstens ansatzweise Subjekt, zugehöriges Prädikat und Objekt(e) unterscheiden können.

Grundsätzlich gilt – wie in der heutigen Interpunktion –, dass das stärker trennende Zeichen das schwächere aufsaugt. Z. B. setzt sich an einer Stelle, an der sowohl ein schließendes Appositionskomma als auch ein Doppelkomma stünde, das Doppelkomma durch:

Beispiel:

Do un\$' herre do geborn wart von \$ce\ marien., d' ewigín megede,, Do ir\$cheín ein \$t're in o\$t' (=)ri-che.(.)

5.4.2.1 Das Doppelkomma

Sofern die Grenzen von Elementarsätzen nicht durch Punkte oder punktwertige Zeichen markiert werden, wird dort stets das doppelte – nicht das einfache! – Komma gesetzt, und zwar in Fällen von Koordination, Subordination, Parenthese, Anrede, Interjektion und Herausstellung.

Beispiele:

Koordination:

Daz da leidis ge\$chach.,, daz liden di mu\ottere(:) di trunkín daz leít.,, di hate\ den \$merzen.,, alleín \$turbín di kí\t.(.)

Subordination:

Do er da riche wart.,, da \$ant er den \$vn vber mer. zv rome.(.)

Daz v'nam d' kvni\c herodes vn\ vorte,, daz er des kvnící(=)chis v'triben wurde.,, vnd dah(=)te dar vmbe in alle wis,, wi er daz kint ír\$luge.(.)

Herausstellung:

Daz andere.,, Daz was di irwelun(=)ge nach deme hertzzen.(.)

Anrede:

Liben,, wir begen hvte den dac.,, daz di \$vgínden kint durch vn\$írs h'ren willen worden ír\$lagín. von deme kvnige herode.(.)

Außerdem steht das Doppelkomma nach *wan* 'außer' mit nachfolgendem Relativ- oder *daz*-Satz, an dessen Stelle auch eine Nominalphrase stehen könnte.

Beispiel:

Bit der warheide in mach niman bliuen wan,, de warehtich is.(.)
si jæhen nihtes wan,, alse si gheorten unde gesæhen.

5.4.2.2 Das einfache Komma

Das einfache Komma wird nur innerhalb von Elementarsätzen in Fällen von Aufzählung und Apposition verwendet sowie zwischen koordinierten Elementarsätzen, die ein Satzglied teilen.

Beispiel:

Aufzählung:

Di drí kvnínge vonden daz kint. vn\ brah(=)tin íme Golt., mir(=)ren vn\ wirauch.(.)

Koordinierte Elementarsätze mit gemeinsamem Satzglied:

Er ervur dise mere, reit hin wider sa ze stvnt vnd tet sie tristande kvnt.(.)

Vor einer Apposition wird ein einfaches Komma gesetzt, ebenso zwischen mehreren Appositionen.

Das Ende einer Apposition(sreihe) wird durch ein einfaches Komma, gefolgt von einer schließenden runden Klammer, bezeichnet, sofern kein anderes Interpunktionszeichen folgt.

Beispiel:

Di drī kvnīnge vonden daz kint. vn\ - brah(=)tin īme ir ophir., Golt., mir(=)ren vn\ - wirauch.,) vn\ - wurde\ - gemanīt īn ir trovme. von deme heiligen engeli.(.)

Sofern eine Apposition nicht in Kontaktstellung zu ihrem Bezugswort steht, sondern davon durch wenigstens ein Wort getrennt ist, wird sie nicht wie eine Apposition, sondern wie eine Herausstellung behandelt und mit doppelten Kommata abgegrenzt. Enge Appositionen (z. B. *sin neve Tantrisel*) werden überhaupt nicht durch Komma abgetrennt.

Vor *sundern*, *nur*, *wan* in der Bedeutung 'sondern' (nicht Nebensatzleitend!) wird einfaches Komma gesetzt.

5.4.2.3 Kein Komma

In Aufzählungen wird – wie im Neuhochdeutschen – kein Komma gesetzt, wenn eine koordinierende Konjunktion vorhanden ist.

Beispiel:

Di turtilduben vn\ - di andīrn duben., daz šī\ - t kv(=)šhe vogele. vn\ - reīne vogle.(.)

Koordinierte Elementarsätze, die ein gemeinsames Satzglied haben, werden nicht mit einem doppelten Komma voneinander getrennt.

Beispiel:

*Vn\ - h' štu\ - t*2 uf*1 vn\ - nam daz kint vn\ - šīne\ - muter. des nachtes. Vn\ - intweich ī\ - egy\ 'pte\ - vn\ - was da. biz zu herodes voršcheidu\ - ge., Uf~\ daz irfullit w%or(=)de., daz gešprochi\ - išt vo\ - dem h're\ - durch de\ - p\phete\ - šprechinde.(.)* "Vz egy\ 'pte\ - han ich gerufe\ - mine\ - švn.(.)"

Einfache Infinitiv- und Partizipialkonstruktionen werden wie im Neuhochdeutschen nicht mit Kommata abgetrennt.

5.4.3 Unterbrochene Elementarsätze

Wenn ein Elementarsatz durch eine Apposition, eine Parenthese, eine Interjektion oder einen eingeschobenen Haupt- oder Nebensatz unterbrochen und später fortgesetzt wird, kann zur Markierung am Beginn des Einschubs ein doppeltes Komma und eine öffnende runde Klammer und am Ende ein doppeltes Komma und eine schließende runde Klammer gesetzt werden.

Beispiele:

Unterbrechung des Elementarsatzes durch eine ferngestellte Apposition:

Y\ 'ešše abir gebar dauid, den ku\ onic.(.) David ab'.,(der ku\ onig.,) gebar šalomon uz ir., di vrie was. (.)

Unterbrechung des Elementarsatzes durch eine Interjektion:

**C*(U*)n\ - nach dem tode herodis.,(šeht.,) d' engil des h're\ - iršchein iošeph ī\ - dem šlafe ī\ - egy\ 'pte\ - šprechinde.(.)*

Unterbrechung des Elementarsatzes durch einen Satz:

*Vnde din vatir.,(d' iz šiht ī\ - d' v'borge\ - heit.,) gibet*2 dir wid'*1.(.)*

Wan vwir vatir weiz.,(wes uch not išt.,) er., wan ir en bittet.(.)

Dar\ u\ - %e., ob ir.,(šo ir boše šit.,) Ku\ onnit gute gabin gebin vwere\ - švne\ -., Wie vile m' vwir va(=)ter.,(d' ī\ - den himele\ - išt.,) gibet gut den., di en bitten.(.)

Von diesen beiden Markierungen können dann u. U. auch mehrere eingeschobene Elementarsätze umschlossen werden. Solche Einschübe in Einschüben werden nicht geklammert, sondern mit bloßen Doppelkommata abgetrennt.

5.4.4 Direkte Rede

Die Interpunktion der direkten Rede erfolgt entsprechend den Regeln der neuhochdeutschen Schriftsprache. Vor der direkten Rede steht also der Doppelpunkt, die direkte Rede wird von öffnenden und schließenden Anführungszeichen eingefasst. Wenn die direkte Rede durch eine Redeeinleitung unterbrochen wird, wird diese wie ein Einschub behandelt, d. h. durch Doppelpunkt mit öffnender und schließender Rundklammer umfaßt.

Beispiele:

Wörtliche Rede:

**(U*)n\ - nach dem tode herodis., (\$eht.,) d' engil des h're\ - ir\$chein io\$eph i\ - dem \$lafe i\ - egy\ 'pte\ - \$prechinde.(.) "Ste*2 uf*1,, vn\ - nim daz kint vn\ - \$ine muter., Vn\ - ge i\ - daz lant (Israel). (.)"*

Unterbrechung der wörtlichen Rede durch eine Redeeinleitung:

Vn\ - (Jesus) antw'te ime.(.) "Jz i\$st ge(=)\$cribe\ -", (\$p%%ach h'.,) "Wan nicht alleine i\ - dem brote le(=)bit d' me\ - \$che., \$und'n i\ - alle de\ - worte gotis.(.)"

Einschub wörtlicher Rede in erzählenden Text:

*Vn\ - \$p%%ach zu\o ir.(.) "Du \$alt nicht weine\ -", vn\ - ginc*2 zu\o*1 Vn\ - ru\orte daz \$tetelin.(.)*

6 Segmentierung von Belegen im fertigen Index

Nachträglich ist in den korrigierten Indizes eine Segmentierung mit Leerzeichen vorgenommen worden. Die Segmentierung betrifft die Wortarten pron, art, num, adj, p(art) und wurde nach folgenden Regeln vorgenommen:

pron, art, num:

1. Auch bei ganz oder teils unleserlichem Flexionsmorph wird segmentiert:

ein <e>m

un\$er [s]

ander e[<m>]

2. Ligaturzeichen und sonst nicht zum Lautkörper der Endung gehörige Zeichen werden nach links gezogen:

d_ en

and' me wie ander me

ien\ - em

3. Nicht segmentiert wurden fremdsprachliche Formen.
4. Nicht segmentiert wurden echte Pronomina.
5. Nicht segmentiert wurden endungslose Formen.
6. Nicht segmentiert wurden Formen mit unscharfer Flexionsmorphgrenze: Das heißt, es wurde nur segmentiert, wenn der Stammauslaut klar erkannt werden kann, also, wenn der stammschließende Konsonant oder Vollvokal eine eindeutige Entsprechung im Beleg hat.

Nicht segmentiert wurden daher folgende Fälle:

- Die Form ist suppletiv flektiert. Bsp. *zwêne, drî*
- (Assimilierter) Stammauslaut und (assimilierter) Endungsanlaut fallen zusammen.

Bsp. *dirre*

7. Aber: Bei doppeldeutigem e ist doch segmentiert worden; es wird der Endung zugeschlagen: *dis eme*

Weitere Beispiele für Segmentierung:

ein em

un\$er s

d '

d az

w er
w az

adj, part, pV, pG:

1.-3. wie bei pron, art, num.

4. entfällt.

5.-7. sind anders geregelt:

- Endungslose Formen haben ein schließendes Leerzeichen
- Eine Zeichenfolge, die noch zum Stamm, aber auch schon zur Endung gehören kann, ist links und rechts mit Leerzeichen abgetrennt. Das gilt auch für doppeldeutiges *e*.

Beispiele:

quot e,,

luter e,, (z. B. NSmw)

lut er e,, (z. B. GSf)

wild e ,, (z. B. NPMf; mit schließendem Leerzeichen)

krist en ,, (z. B. NPw; mit schließendem Leerzeichen)

7 Datensatzgestalt. Lemma, Kategorie, Bestimmung

Die Lemmaansätze, die Angaben zu den Kategorien und die grammatischen Bestimmungen müssen formal streng standardisiert sein, damit die diversen Programme zur Vorlemmatisierung und Indizierung korrekt arbeiten und später die Indices gezielt und treffsicher durchsucht werden können. Daher sind sämtliche bislang angesetzten Lemmata in einem Lemmaregister zusammengestellt, auf das – direkt oder indirekt – sowohl die Programme als auch die Bearbeiter beim sogenannten “Löcherstopfen” (siehe 8.4) und bei der Vorlemmatisierungskorrektur (siehe 9) zugreifen können.²⁹

7.1 Unveränderlicher und veränderlicher Ansatz. [!], [!!] und *

Jeder Datensatz hat zwei Lemmaansätze:

- einen, der immer mit dem Eintrag im Lemmaregister genau übereinstimmt und nie an die jeweilige Stelle angepasst wird,
- und einen, der unter bestimmten Umständen an die jeweilige Stelle angepasst wird.

Bsp.:

Unveränderliche Angaben	Veränderliche Angaben								
`LEMMATA @WORTART	~LEMMATA	#WORTART	^FLEXION	~NORMALFORM	\BELEG	_STELLE	°STAND	:ANM	GRAPHO
`dër @pron	~dër	#art	^NSn	~daz	\daz	_V·1·P_IrPr-32v,12{S.234}.06	°	:	...
`schar(e) @mf	~schar(e)	#n	^NS	~schar<e>	\\$char	_V·1·P_IrPr-32v,12{S.234}.07@	°	:	...
`... @...	~...	#...	^...	~...	\...	_...	°...	:...	...

Es gibt zwei Platzhaltermarken:

[!!] in Feldern, die bei der Anzeichnung noch nicht betrachtet wurden.³⁰

²⁹ Vor den weitreichenden Lemmaansatzprüfungen und Vereinheitlichungen, die das Lemmaregister (= die Lemmaliste) betroffen haben, galt: Das Lemmaregister umfaßt derzeit gut 75 000 Lemmata, und zwar – soweit nicht Projektkonventionen andere oder weniger Ansätze erfordern – sämtliche Lemmata aus Lexers “Handwörterbuch”, den Nachträgen dazu und aus dem “Findebuch” sowie alle Lemmata, die darüber hinaus in den bislang indizierten Texten angesetzt worden sind. Das Lemmaregister befindet sich im Verzeichnis C:\LEM-WB, und zwar als Gesamtdatei LRG-GES und als Folge von etwa 30 Teildateien LRG-x.

³⁰ [!!] steht nicht in den Feldern, die ungefüllt bleiben können: also nicht im Flexionsfeld, Statusfeld, Anmerkungsfeld und Graphofeld; Beleg- und Stellefeld sind auch in den [!!]-Fällen schon gefüllt.

[!] in Feldern dann, wenn man nach Betrachtung der Stelle nichts weiter ansetzen will oder kann.

Diese Platzhaltermarken füllen alleine ein Feld; einzig erlaubte Ergänzung ist der * für Fremdsprachliches.³¹

Fremdsprachliche Lemmata (auch fremde Eigennamen) sind durch einen unmittelbar angeschlossenen Stern (*) gekennt.³²

Bsp.: `!!! @!!! ~!!! #!!! ^ ~!!! \sv\ozichede
`!! @!! ~!! #!! ^ ~!! \knrslkfu
`! @! ~esse* #!
`sanctus* @adj ~sanctus* #adj

Im Feld der unveränderlichen Angabe steht [!] genau dann, wenn kein Ansatz in der Lemmaliste gefunden wird.

Bsp.: `! @! ~! #! ^ ~! \knrslkfu
`! @! ~esse* #!
`! @! ~Agathon #PN

Für das @-Feld gilt ferner: Es entfallen Angaben, die im #-Feld nach dem Leerschritt stehen:

- a 'Partizip (I oder II)'; z. B.
`loben @swv ~loben #swv pG ^3SGI 'schwaches Verb im Partizip II, 3. Sg. Geg. Ind.)
- b 'Infinitiv'; z. B.
`loben @swv ~loben #swv i ^- 'schwaches Verbum im Infinitiv'³³
- c 'Komparativ'; z. B.
`alt @adj ~alt #adj comp ^-
- d 'Superlativ'; z. B.
`alt @adj ~alt #adj sup ^-

doch 'Komparativ' und 'Superlativ' entfallen dann nicht, wenn schon der Lemmaansatz im Lemmaregister kompariert ist, z. B.

`bèzzet @adj comp ~bèzzet #adj comp ^-

7.2 Der Lemmaansatz

Die Gestaltung der Lemmaansätze und die Bestimmung der grammatischen Kategorien richtet sich meist nach dem jeweils entsprechenden Ansatz in Lexers "Handwörterbuch". Zusätzlich gelten folgende Regelungen, die die Unterschiede zwischen den Lemmaansätzen bei Lexer und im Lemmaregister erklären.

1. Nach dem Vorbild des "Lexer"-Ansatzes werden durch Segmentierungsstriche Präfixe, Suffixe und Kompositionsglieder abgetrennt. Diese morphologische Segmentierung wird konsequent auch bei vielgliedrigen Zusammensetzungen durchgeführt, während Lexer etwa bei mehrfacher Präfigierung nur das erste Präfix segmentiert. Beispielsweise wird gegen-

³¹ Aufgegeben sind alle anderen Arten von []-Klammerung in den ersten fünf Feldern. Anmerkungen (auch zu Inkongruenzen und Abweichungen vom Erwarteten) stehen immer nur im Anmerkungsfeld.

³² Aufgegeben sind die verschiedenen anderen, letztlich nicht mehr einheitlich eingesetzten Kennzeichen dafür, dass eine Form fremdsprachlich sei.

³³ Vgl. auch `loben @swv ~loben #swv i ^D 'schwaches Verbum im Infinitiv, als Nominalglied im Dativ' wie in ze sehen(ne).

über dem Ansatz *un-gesellschaft* bei "Lexer" konsequent segmentierend *un-ge-sèlle-schaft* angesetzt.

Das Programm LEM, Option 8, überprüft in einem späteren Arbeitsschritt die Abtrennung von Präfixen und Suffixen im Lemmaansatz, nicht aber die Segmentierung von Komposita. So würde etwa die Abtrennung von *un-*, *ge-* und *schaft-* bei einem Ansatz *un-gesèlleschaft* vorgeschlagen, nicht aber die Segmentierung von *diènest-mann* statt fälschlich *diènestmann*.

2. Während "Lexer" nur das alte germanische /e/ als <ë> kennzeichnet, aber nicht weiter zwischen /e/ als Umlaut von /a/ einerseits und Schwa andererseits unterscheidet, werden im Grammatikprojekt alle drei /e/ konsequent geschieden. Dabei bezeichnet <ë> wie bei Lexer das alte /e/, <è> den Primärumlaut aus /a/ und <e> das schwachtonige /ə/. Beispielsweise lauten gegenüber "Lexer" *gegen-rede*, *her-bërge* dementsprechend die Lemmaansätze *gègen-rède* und *hèr-bërge*.

Als Deckelzeichen für ein e unklarer Qualität wird É verwendet, also ein großes, akutiertes e, so dass das Zeichen sowohl im Lemma- als auch im Normalformfeld eindeutig ist. Dieses É wird vor allem für Fremdsprachliches gebraucht (s.).

3. Fremdsprachliche Wörter und Wortglieder werden vereinfacht geschrieben:

- É für alle e-Laute,
- keine Längenangaben,
- keine Segmentierstriche innerhalb solcher Wörter oder Wortglieder.

Fremdsprachlichkeit liegt mindestens dann vor, wenn die Qualität eines etwaigen e-Lautes fraglich ist.

Zu den fremdsprachlichen Wörtern zählen unter anderem viele Namen und Herkunftsbezeichnungen.

4. Zum Ansatz von Eigennamen und Herkunftsbezeichnungen:
7.4.10.

5. Varianten:

- Vielfach setzt Lexer bei einem Lemma Varianten mit und ohne auslautendes /-e/ an, so stets im Falle der mhd. Apokope nach Kurzvokal + Liquid (z. B. "tür, türe"). Besonders bei mitteldeutschen und frühen oberdeutschen Texten muß dem – hier sehr oft (noch) erhaltenen – finalen /-e/ Rechnung getragen werden; dies geschieht, indem das /-e/ im Lemmaansatz in runden Klammern angefügt wird (etwa *tür(e)*).
- Bei den "Lexer"-Ansätzen wird bei rein phonologischen Varianten der erste Ansatz in Lexers "Handwörterbuch" gewählt, z. B. wenn das "Handwörterbuch" einen Ansatz mit und einen ohne Umlaut bietet (z. B. "lügen, lügen").³⁴

6. Die in den "Lexer"-Ansätzen durchgeführte Auslautverhärtung wird rückgängig gemacht, und zwar sowohl im absoluten Auslaut als auch bei Kompositionsgliedern:

- -t > -d, z. B. *wald* statt *walt* (vgl. Gen. Sg. *waldes*); dagegen *quot* (vgl. Gen. Sg. *quotes*)
- -c > -g, z. B. *tag* statt *tac* (vgl. Gen. Sg. *tages*)³⁵
- -p > -b, z. B. *wîb* statt *wîp* (vgl. Gen. Sg. *wîbes*)³⁶
- -f > -v, z. B. *hov* statt *hof* (vgl. Gen. Sg. *hoves*); dagegen *slôf* (Gen. Sg. *slôfes*)³⁷
- Auch wird im Auslaut -h und nicht -ch angesetzt, wo in flektierten Formen im Inlaut -h- steht, z. B. *hôh* (Gen. Sg. *hôhes*).

³⁴ Früher galt mehr: "Bei den "Lexer"-Ansätzen ohne und mit durchgeführter Lenisierung von ahd. /t/ nach Nasal oder Liquid, also Ansätzen mit /t/ oder /d/ nach /l/ oder /n/ (z. B. "walten, walden"; "winter, winder") wird grundsätzlich der erste Ansatz in Lexers "Handwörterbuch" gewählt. Entsprechend wird verfahren, wenn das "Handwörterbuch" einen Ansatz mit und einen ohne Umlaut bietet (z. B. "lügen, lügen")."

³⁵ Nach Vokal ist auslautend fast immer <g> statt <c> zu setzen.

³⁶ Von vereinzelt Ausnahmen abgesehen ist auslautend <p> regelmäßig in umzuwandeln.

³⁷ Auslautend <f> ist meist beizubehalten, ausgenommen fast ausschließlich *brièv*, *bischof*, *hov*, *wolv*, *vûnv*, *èi(n)l(i)v*, *zwe(l)i(v)*.

7. Die in vielen "Lexer"-Ansätzen zu findende Sonorierung der Art *nd* < *nt* wird rückgängig gemacht.
8. Die Vereinfachung auslautender Geminaten bei Lexer wird rückgängig gemacht; zum Beispiel lauten die Lemmaansätze also *mann*, *sinn*, *voll*, *blick* gegenüber *man*, *sin*, *vol*, *blic* in Lexers "Handwörterbuch".³⁸
9. Auch der Auslaut von *wa-/wō*-Stämmen wird in der Inlautslautung angesetzt, also mit *w* (nicht *e* oder (*e*)), vgl. *varw* #adj, *sêw* #m.
10. Das Adjektivsuffix *-ig*, das bei Lexer meist als *-ec* erscheint, wird grundsätzlich – im Inlaut wie im Auslaut – als *-ig* wiedergegeben, z. B. *manig* statt *manec*, *manig-valt* statt *manec-valt*, *manig-valtig-hêt* statt *manec-valtec-heit*.
11. Bei Substantiven mit regelmäßigem Pluralumlaut wird der Genusangabe <u> hinzugefügt, bei nur fakultativem Pluralumlaut <(u)>, z. B. "gast #m<u> (Plural *gêste*); "balg #m<(u)> (Plural *bêlge*, selten *balge*).
12. Einheitlicher Ansatz von *ent*-.³⁹
13. Feminina auf *-inne/-în* werden immer auf *-inne* angesetzt; nur die Normalform wird in einem späteren Arbeitsschritt angepaßt, wenn die handschriftliche Form auf *-în* ausgeht.
14. Nomina agentis mit dem Suffix ahd. *-âri*, die Lexer teils mit *-ære*, teils mit *-er* anführt, werden einheitlich mit *-ære* angesetzt.
15. Vor den Suffixen *-inne*, *-în* (Adj.), *-ede*, *-lîn* und *-chîn* wird das Lemma grundsätzlich mit Umlaut angesetzt⁴⁰, vor *-ære*, *-lich* und *-liche* grundsätzlich ohne Umlaut. Vor den Suffixen *-ig*, *-nisse* richtet sich der Umlaut des Lemmaansatzes nach dem entsprechenden Ansatz bei Lexer.
16. Die Bestandteile von Komposita werden im Lemmaansatz möglichst so geschrieben wie die jeweiligen Simplicia; so wird z. B. *wünne-brunne* angesetzt (gegen *wunne-brunne* bei Lexer), weil das Simplex als *wünne* angesetzt ist.⁴¹
17. Komposita mit *voll-*, *volle-* oder *vollen-* als erstem Kompositionsglied werden stets mit *vol-* angesetzt; Komposita mit *all-*, *alle-*, *allen-* oder *allent-* stets mit *all-*. Später werden die Normalformen an die jeweilige handschriftliche Wortform angepaßt.
18. Feste Fügungen mit *in-* oder *en-* (z. B. *en-triuwen*, *en-gêgen*) werden als ein Lemma angesehen, wenn es in Lexers "Handwörterbuch" einen entsprechenden Ansatz oder wenigstens einen entsprechenden Verweis gibt.
19. Verbindungen aus *aller* und dem Superlativ eines Adjektivs werden als ein Wort angesetzt (z. B. *aller-schoènest*).
20. Pronominaladverbien wie *dâr zuo* werden grundsätzlich als zwei Wörter angesetzt und später durch das Programm LEM (7) als zusammengehörig markiert.
21. Adverbien mit verschiedenen erstarrten Flexionsendungen (z. B. *all-ze-hande* neben *all-ze-handes*) werden jeweils als ein eigenes Lemma angesetzt, auch wenn Lexer sie als ein Lemma anführt.

³⁸ Früher galt fernerhin: "Die Vereinfachung wird beibehalten bei auslautend /-ff/ (etwa *schiff*), ebenso im Auslaut von Indeklinabilia und innerhalb von Komposita (z. B. *man-haftig*)." Auch bei *z* scheint die Vereinfachung fast immer beibehalten worden zu sein.

³⁹ Aufgegeben wurde die alte, schon bisher nicht ausnahmslos verwirklichte Regel: "Der Ansatz der mit *ent-* präfigierten Verben folgt Lexers "Handwörterbuch" (ungeachtet gelegentlicher Inkonsistenzen): Das präfigierte Verb wird angesetzt auf *enph-* (ohne Bindestrich!), wenn das Basisverb mit *v* oder *ph* beginnt (z. B. *enphallen*, *enphâhen*, *enphlihten* statt *ent-vallen*, *ent-vâhen*, *ent-phlihten*); es wird angesetzt auf *en-* (ohne -t-, aber mit Bindestrich), wenn das Basisverb mit *b*, *p*, *g*, *k*, *d*, *t*, *r*, *z* beginnt (z. B. *en-binden* 'losbinden', *en-prîsen* 'entwerten', *en-gûrten* 'entgürten'); es wird angesetzt auf *ent-* in allen anderen Fällen."

⁴⁰ Bei *-lîn* und *-chîn* wird das Lemma nur mit Umlaut angesetzt, wenn die Basis einsilbig ist.

⁴¹ Hier ist in bereits vorhandenen Indices durchaus mit Inkonsistenzen zu rechnen, die ggf. zu beseitigen sind. Sollten Inkonsistenzen oder eindeutige Fehler festgestellt werden, sollten die betreffenden Lemmata in der "Lemmaänderungsliste" notiert werden, damit eine Korrektur in allen Indices erfolgen kann.

- 2.2. Adverbien auf *-lîche* oder *-lîchen* aber werden stets auf *-lîche* angesetzt, ungeachtet der handschriftlichen Form.
- 2.3. *()schèpfen* und *()schaffen*:

	Gegenwart	Vergangenheit
<i>()schèpfen #stv6b</i> für:	<i>schepf... / scheph... / scheff...</i>	<i>schuof(...)</i>
<i>()schèpfen #swv</i> für:		<i>schepf... / scheph... / scheff...</i> mit Dentialsuffix
<i>()schaffen #swv</i> für:	<i>schaff...</i>	<i>schaff... mit Dentialsuffix</i>

- 2.4. *(in) latîn / latîne / latînen* u. a. Sprachbezeichnungen

<i>in der latîne</i>	ːːlatîne #f ^DS
<i>in der latînen</i>	ːːlatîne #f ^DS
<i>in latîn</i>	ːːlatîn #n ^AS
<i>in latîne</i>	ːːlatîn(e) #fn ^DAS
<i>in latînen</i>	ːːlatîne #f ^DAS
<i>latîn</i>	ːːlatîne #adv
<i>latîne</i>	
<i>latînen</i>	ːːlatînen #adv
<i>in ...sch</i>	ːː...sch #adj ^OASn% (z. B. ːːdiut(i)sch)
<i>in ...sche</i>	ːː...sch #adj ^ASf%
<i>in ...schen</i>	ːː...sch #adj ^DASmfnw!%
<i>...sch</i>	ːː...sche #adv
<i>...sche</i>	
<i>...schen</i>	ːː...schen #adv

- 2.5. Für ‘oder’ wird (abweichend von der sonst ausdrucksseitig ausgerichteten Anzeichnung) auch für hsl. *ove*, *obe*, *ave(r)*, *ald(er)* u. s. w. nichts anderes als ːːoder angesetzt, wenn die Bedeutung ‘oder’ ist.
- 2.6. Unterscheidung von ːːbuoze @f und ːːbuoz @f (so trotz Lexers *m*) nach Lexer:
- *buoze* ‘Buße, Strafe, Wiedergutmachung, Heilung’
 - *buoz* nur in folgenden Wendungen: *buoz werden / machen / tuon* mit Dat. d. Pers. und Gen. d. Sache ‘jmdn. befreien von etwas’; *mir wirt buoz eines dinges* ‘ich werde von etwas befreit’.

7.3 Die Kategorien

Hinsichtlich der grammatischen Kategorien gelten folgende Abkürzungen und Regelungen (die in einzelnen Punkten von Lexer abweichen).

Wortart	Kürzel ⁴²	Erläuterungen
Substantiv ⁴³	m	starkes oder schwaches ⁴⁴ Maskulinum ohne Umlaut
	m<u>	starkes oder schwaches Maskulinum mit Umlaut
	m<(u)>	starkes oder schwaches Maskulinum mit fakultativem Umlaut
	f	starkes oder schwaches Femininum ohne Umlaut
	f<u>	starkes oder schwaches Femininum mit Umlaut
	f<(u)>	starkes oder schwaches Femininum mit fakultativem Umlaut
	n	starkes oder schwaches Neutrum ohne Umlaut
	n<(u)>	starkes oder schwaches Neutrum mit fakultativem Umlaut
	PN	Personenname
	ON	Ortsname
	EN	Eigename
Adjektive	adj	Adjektiv (im Positiv)
	adj comp	Adjektiv im Komparativ; der Lemmaansatz steht im Positiv, wenn vorhanden (z. B. <i>alt #adj</i> und <i>alt #adj comp</i>)
	adj sup	Adjektiv im Superlativ; der Lemmaansatz steht im Positiv, wenn vorhanden (z. B. <i>alt #adj sup</i>), außer bei den Superlativen mit <i>aller-</i>
	adj_part	Partizipialadjektiv ⁴⁵ oder
	... pV ⁴⁶	Partizip der Vergangenheit
	... pG	Partizip der Gegenwart
Adverb ⁴⁷	adv	Adverb (auch zu adj part angesetzt: <i>un-müèdende</i> , <i>un-ge-zogene</i>); <i>dannen</i> , <i>dâr</i> , <i>dar(e)</i> auch Nebensatzleitend ⁴⁸
		Pronominaladverb: keine eigene grammatische Kategorie; die Kennzeichnung der beiden Konstituenten erfolgt durch halbautomatisch erzeugte Lemmazusätze ⁴⁹ , z. B. <i>dâr/+umbe</i> , <i>umbe/dâr+</i> (bei Kontaktstellung); <i>dâr/.+umbe</i> ,

⁴² Ein zusammengesetztes Wortartkürzel wie "adj_part" ist mit Unterstrich (" _ ") verleimt. Dagegen steht eine belegabhängige Bestimmung der Wortart wie "comp" in "adj comp" nach Leerzeichen (" "). Die Fälle lassen sich gut unterscheiden, indem man den nicht abgekürzten Wortartnamen denkt: Ist er ein Kompositum, steht ein Unterstrich ("Partizipialadjektiv", also "adj_part"), ist er aber kein Kompositum, steht ein Leerzeichen ("Adjektiv im Komparativ", also "adj comp"; "Partizip 2 eines schwachen Verbs", also "swv pG").

⁴³ Manche Wörter sind grundsätzlich in mehreren Genera belegt. Sofern in diesen Fällen aus dem Kontext eine sichere Genuszuweisung nicht möglich ist, werden entsprechend dem Ansatz in Lexers "Handwörterbuch" alle grundsätzlich möglichen Genera angesetzt, z. B. "mf", "mn", "fn", "mfn". – Bei der Angabe des Genus hat grundsätzlich die konkrete Form Vorrang vor den Angaben in Lexers "Handwörterbuch" oder anderen Wörterbüchern.

⁴⁴ Eine Unterscheidung zwischen starken und schwachen Substantiven erfolgt nicht bereits bei der Lemmatisierung, sondern ist der späteren Auswertung vorbehalten.

⁴⁵ Partizipialadjektive sollen nur angesetzt werden, wenn eine Lemmatisierung des betreffenden Wortes als normales Partizip wegen einer semantischen Differenzierung nicht möglich ist. Ein häufiges Beispiel ist *gebor(e)n*, das als Partizip Präteritum zu *bër(e)n* angesetzt wird, wenn es konkret um das Geborenwerden geht, und als Partizipialadjektiv in Fällen wie *wol gebor(e)n*, wo der konkrete Vorgang deutlich in den Hintergrund getreten ist. Ähnlich ist *ge-tân* gelagert, das als Partizipialadjektiv im Unterschied zum normalen Partizip Präteritum die Bedeutung "beschaffen" hat (etwa *wol getân*).

⁴⁶ An der Stelle von ... steht stv1a oder stv1b oder (usw., siehe unten unter Verben).

⁴⁷ Zur Unterscheidung von Adverb, Partikel und Konjunktion vgl. auch Kap. 7.4.11.

⁴⁸ In alten Indices sind möglicherweise auch *dâr*, *dar(e)*, *dannen* in relativischem Gebrauch noch als *rel adv* kategorisiert. Diese Auskunft wurde durch ein #[rel/konj] im Anmerkungsfeld bewahrt.

⁴⁹ Dazu vgl. 10.2.

Wortart	Kürzel	Erläuterungen
		<i>umbe/dâr.+</i> (bei Fernstellung)
	adv_rel	Relativadverb: <i>swâr, swar(e), swannen</i> (nicht aber <i>swiè</i> , s. Konjunktion)
	adv_int	Interrogativadverb: <i>wiè, wanne, wannen, wâr, war(e)</i>
	adv_part	Partizipialverb (nur in einigen frühmhd. Texten angesetzt)
Konjunktion	konj	Konjunktionen, auch <i>swiè</i> , Nebensatzeinleitendes <i>dô</i> und in mancher Verwendung <i>danne</i> (s. 7.4.11)
Partikel	partikel	<i>danne, ne, sô, wan</i> (sofern nicht Konjunktion oder Adverb)
Artikel	art	bestimmter Artikel <i>dër (diu, daz)</i> , sofern nicht Relativ- oder Demonstrativpronomen
		unbestimmter Artikel <i>èin</i> , sofern nicht Numerale
Pronomen	pron	Personalpronomen: <i>ich</i> ⁵⁰ , <i>dû</i> ⁵¹ , <i>ër</i> ⁵² , <i>wir</i> ⁵³ , <i>ir</i> ⁵⁴
		Demonstrativpronomen <i>dër</i> (nur, wenn substantiviert), <i>dise, jèner, sëlþ</i>
		Reflexivpronomen <i>sich</i>
		reziprokes Pronomen <i>èin-ander</i>
		Interrogativpronomen: <i>wër, wëder, wël(i)ch</i>
		Relativpronomen: <i>dër, swër, swëder, swël(i)ch</i> ⁵⁵
		alle anderen Pronomina: <i>all</i> 'all', <i>ander</i> 'ander', <i>manig</i> 'manch, viel', <i>so-l(i)ch</i> 'solch', <i>al-so-l(i)ch</i> 'solch', <i>deh-èin</i> 'kein, irgendein', <i>kèin</i> 'kein, irgendein', <i>neh-èin</i> 'kein', <i>newëder</i> 'kein (von beiden)', <i>sum</i> 'irgendein, einige', <i>sume-lich</i> 'irgendein, einige', <i>ëte-lich</i> 'irgendein, einige', <i>ëte-wër</i> 'irgendein, einige', <i>ëtes-lich</i> 'irgendein, einige', <i>sich-èin</i> 'irgendein, einige', <i>de-wëder</i> 'irgendein (von beiden)', <i>ent-wëder</i> 'irgendein (von beiden)', <i>iè-ge-wëder</i> 'jeder (von beiden)', <i>iè-wëder</i> 'jeder (von beiden)', <i>ièt-wëder</i> 'jeder (von beiden)', <i>ge-lich</i> 'jeder' (mit vorangestelltem Substantiv im Gen. Pl.), <i>iè-ge-lich</i> 'jeder', <i>aller-manne-ge-lich</i> 'jeder', <i>iè-wël(i)ch</i> 'jeder', <i>iètes-lich</i> 'jeder'
	pron_poss	Possessivpronomen ⁵⁶ : <i>mîn, dîn, sîn, unser, iuwer, ir</i>
	pron_subst	Pronominalsubstantiv: <i>man, iè-man, niè-man</i>
Numerale	num	alle Zahlwörter: <i>èin</i> (sofern nicht unbestimmter Artikel; auch in <i>mit èin, in èin</i> etc.), <i>zwêne, drî</i> etc.
Verb	stv1a ⁵⁷	starkes Verb der Ablautreihe 1a
	stv1b	starkes Verb der Ablautreihe 1b
	stv2a	starkes Verb der Ablautreihe 2a
	stv2b	starkes Verb der Ablautreihe 2b

⁵⁰ Zum Lemma *ich* gehören auch die flektierten Formen *mîn, mir* und *mich*.

⁵¹ Zum Lemma *dû* gehören auch die flektierten Formen *dîn, dir* und *dich*.

⁵² Zum Lemma *ër* gehören alle flektierten Formen des Maskulinums (*ër, ës, sîn, ime, in, siè, sî, si, ire*), des Femininums (*siu, siè, sî, sî, ir(e), in*) und des Neutrums (*ëz, ës, ime, siu, siè, sî, si, ir(e), in*).

⁵³ Zum Lemma *wir* gehören auch die flektierten Formen *unser, uns* und *unsich*.

⁵⁴ Zum Lemma *ir* gehören auch die flektierten Formen *iuwer, iu* und *iuch*.

⁵⁵ Die in alten Indices noch vorkommende Kategorie "rel pron" für das Relativpronomen wird nicht mehr verwendet.

⁵⁶ Es ist genau darauf zu achten, das Possessivpronomen nicht mit dem Genitiv der Personalpronomina *ich, dû, ër, wir, ir* zu verwechseln.

⁵⁷ Manche Verben können sowohl stark als auch schwach flektiert werden. Wenn die konkrete Form nicht eindeutig ist, wird das starke Verb angesetzt.

Wortart	Kürzel	Erläuterungen
	stv2c	starkes Verb der Ablautreihe 2 mit Wurzelvokal /iu/ (z. B. <i>be-riuwen, biuzen</i> etc.)
	stv2d	starkes Verb der Ablautreihe 2 mit Wurzelvokal /û/ (z. B. <i>sûfen, sûgen</i> etc.)
	stv3a	starkes Verb der Ablautreihe 3a
	stv3b	starkes Verb der Ablautreihe 3b
	stv4	starkes Verb der Ablautreihe 4
	stv5	starkes Verb der Ablautreihe 5
	stv5b	starkes j-Präsens der Ablautreihe 5
	stv6	starkes Verb der Ablautreihe 6
	stv6b	starkes j-Präsens der Ablautreihe 6
	stv7	starkes Verb der Ablautreihe 7 (ehemals reduplizierend)
	swv	schwaches Verb
	swv-a	anomales schwaches Verb: <i>dunken, dênken, haben, sagen, wirken, vûrhten</i> und ihre Komposita
	anv	anomales Verb, d. h. Wurzelverben (<i>sîn, tuon, gân, stân</i>), Präteritopräsentien (<i>soln, mûgen, mûezen, wizzen, kunnen, gunnen, tugen, turren</i>), <i>wêllen</i> sowie <i>bringen</i> und <i>beginnen</i>

7.4 Grammatische Formenbestimmung

7.4.1 Abkürzungen

Bei der grammatischen asdf•er Formenbestimmung werden folgende Abkürzungen benutzt:

Kategorisierung	Kürzel	Kategorie
Nomina und Pronomina		
Kasus	N	N ominativ
	G	G enitiv
	D	D ativ
	A	A kkusativ
	I	I nstrumental ⁵⁸
Numerus	S	S ingular
	P	P lural
Genus	m	M askulinum
	f	F emininum
	n	N eutrum
Adjektivflexion	-	prädikativ
	0	endungslos
	w	schwach flektiert (engl. "weak")
		stark flektiert (ohne Kürzel)
Weiteres siehe in 7.4.5, 7.4.6		
Verben		
Person	1	1. Person
	2	2. Person
	3	3. Person
Numerus	S	S ingular
	P	P lural

⁵⁸ Der Instrumental ist nur noch in festen Wendungen wie *von diu, ze diu* gebräuchlich.

Kategorisierung	Kürzel	Kategorie
Tempus	G	Präsens (G egenwart)
	V	Präteritum (V ergangenheit)
Modus	I	Indikativ
	K	Konjunktiv
	B	Imperativ (B efehlsform)
Infinitiv	i	Infinitiv (ggf. ergänzt um die Kasusangabe <i>iN</i> , <i>iG</i> , <i>iD</i> , <i>iA</i>)

7.4.2 Die Reihenfolge innerhalb der grammatischen Bestimmung

Innerhalb der grammatischen Bestimmung folgen die Abkürzungen für die verschiedenen Kategorien in jeweils festgelegter Reihenfolge ohne Spatium aufeinander:

Substantiv: Kasus+Numerus (z. B. NP, GDS)

Pronomen: Kasus+Numerus+Genus (z. B. GSf)

Adjektiv: Kasus+Numerus+Genus+Deklination (z. B. GSfw)

finites Verb: Person+Numerus+Tempus+Modus (z. B. 3SVK)

Gerundium: Infinitiv+Kasus (z. B. iD)

Partizip, prädikativ: Partizip+Tempus (z. B. pG)

Partizip, attributiv: Partizip+Tempus+Kasus+Numerus+Genus+Deklination (z. B. pGDSn)

7.4.3 Nomina

Formal mehrdeutige Formen (z. B. die *edelen vrouwen*: Akk. Sg., Nom. Akk. Pl.) werden, soweit es möglich ist, nach dem jeweiligen Kontext bestimmt. Ist keine hinreichend sichere Entscheidung möglich, werden alle möglichen Bestimmungen angegeben (SP für "Singular oder Plural"; GD für "Genitiv oder Dativ"). Schwierigkeiten bereitet insbesondere die Numerusbestimmung bei den femininen Abstrakta (wie *êre*, *triuwe*, *mâze*, *minne*), v. a. im Genitiv und Akkusativ; hier gilt die Regel, bei oberdeutschen Texten auf jeden Fall Plural anzusetzen, während bei mitteldeutschen Texten auch Singular oder Singular und Plural angesetzt werden können.

7.4.4 Verben

Der einfache Infinitiv erhält nur die Bestimmung i; bei substantivischer Verwendung wird die Kasusangabe hinzugefügt, insbesondere beim Gerundium (iG, iD). Der Infinitiv nach *ze* wird immer als iD lemmatisiert, unabhängig davon, ob die handschriftliche Wortform noch eine Flexionsendung aufweist oder nicht.

Person, Numerus und Tempus werden bei formal mehrdeutigen Formen soweit wie möglich nach dem syntaktischen Kontext bestimmt. Da verlässliche syntaktische Regeln für den Gebrauch des Konjunktivs weithin fehlen, wird der Modus allerdings nur bei formal moduseindeutigen Formen angegeben:

1./2. Sg. Präs. der starken Verben Kl. II	Indikativ bei <i>u</i> -haltiger Graphie des Wurzelvokals (<i>u, v, iu, iv, u\o, v\o</i>)
1./2. Sg. Präs. der starken Verben Kl. IIIb, IV, V	Indikativ bei <i>i</i> -haltiger Graphie des Wurzelvokals
3. Sg. Präs.	Indikativ <i>-(e)t</i> , Konjunktiv <i>-e</i>
3. Pl. Präs.	meistens: Indikativ <i>-ent</i> , Konjunktiv <i>-en</i> ⁵⁹
3. Sg. Prät. der starken Verben	Indikativ <i>-ø</i> , Konjunktiv <i>-e</i> (und ggf. mit Umlaut)
2. Sg. Prät. der starken Verben	Indikativ <i>-e</i> , Konjunktiv <i>-es(t)</i> ⁶⁰
Pl. Prät.	Konjunktiv mit Umlaut, sofern die jeweilige Handschrift den Umlaut bezeichnet

Stets Indikativ ist ferner die westmitteldeutsche, besonders mittelfränkische 1. Sg. Präs. mit der Endung *-en* (z. B. *ich sagen*). In allen anderen Fällen erfolgt keine Modusbestimmung.

In den Fällen, in denen eine Unterscheidung zwischen Indikativ und Konjunktiv nicht formal, sondern nur syntaktisch möglich ist, kann die Modusbestimmung in runden Klammern vorgenommen werden, z. B. *sagete*: 3SV(K). Diese Bestimmung ist fakultativ; eine Modusbestimmung nach der Syntax muß nicht erfolgen. Wo klar Indikativ vorliegt, sollte auch dies nach Möglichkeit markiert werden, z. B. 3SV(I). Vor allem im Präteritum Plural der starken Verben II-VI wird durch einen Zusatz von (K) die Änderung der Normalform nötig, die dann mit Umlaut angesetzt werden soll (z. T. auch Präterito-Päsentien, Prät. Sg. und Pl.). Änderung der Normalformen im Zusammenhang der (K)-Markierung, die den Umlaut betreffen, müssen aber nicht durch den Einzelnen erfolgen, sondern sollen mit Hilfe eines Programms systematisch geleistet werden.

7.4.5 Adjektive und Partizipien

Wortart: Handelt es sich um ein Adjektiv? Attributive und prädikative Adjektive

in alten mæren: **attributives** Adj. = Adj. steht in einer Nominalgruppe und bezieht sich auf den Kopf dieser Gruppe (hier *mæren*). (Der Kopf ist meist ein Substantiv. Er kann auch ein substantiviertes Wort sein: *ir vil mortlîchez lachen*. Sogar ein Pronomen ist möglich: *ich armer*.)

die wîsen und die tumben sunge einen tōn: **substantiviertes** Adj. = Adj. ist selbst der Kopf der Nominalgruppe. Substantivierte Adjektive werden einfach genau wie attributive lemmatisiert, hier ^{ANPW}. (Ein Ansatz wie *“wîse #m* (swm.) stattdessen sorgt leicht für Unstimmigkeiten, ausgehend von den Fragen des Genus, des Vorhandenseins eines Ansatzes überhaupt und des Wechsels zwischen starker und schwacher Flexion.)

Substantiv dagegen ist: *vil* ^(n.NS) *vröuden* ^(abhängiger GP) *was* ^(3SVI) *ze schouwen dâr*.

wâren milte: **prädikatives** Adj. = Adj. steht in einer Verbalgruppe, und zwar meist zusammen mit einem Kopulaverb (hier *wâren*), das das Adj. an sein Bezugswort bindet: *die herren wâren milte*. Die Rolle des Kopulaverbs kann allerdings außer durch *sîn*, *wêsen*, *wêrden*, *belîben* auch durch andere Verben erfüllt werden: *sî lāgen alle tōt*, *dâ stuont êr sêre wunt*, *si sach sî vil gesunt*, *êr vant in bœsê ze grunt*.

⁵⁹ Hier ist aber auf jeden Fall der Usus der jeweiligen Handschrift zu beachten, um Ausgleichsprozesse zwischen den Flexiven des Indikativs und des Konjunktivs nicht zu verdecken.

⁶⁰ Dies gilt nicht im Mittelfränkischen, da hier Ausgleich erfolgt, indem die Endung des Konjunktivs auf den Indikativ übertragen wird.

Anhaltspunkt für ein prädikatives Adjektiv ist: Entweder Kopulaverb, oder: Das Adjektiv, z. B. *tôt*, lässt sich umschreiben mit einer Wendung wie (nhd.) *tot s e i e n d* oder *a l s T o t e r / T o t e / T o t e s ...*.

Adverb dagegen ist ein Wort, das sich auf ein Verb / eine Verbalgruppe bezieht, umschreibbar ist mit *auf eine ... Weise*: *si kam sô drâte dare* (adv.) **gegen** *sî kam sô trûrec dare* (adj.). Auch *vîl ze sêre*, *ze jungest* und *von(e) êr(e)ste/en* sind alles Adverbien.

pV / pG oder adj part?

Im Zweifelsfall pV / pG. Vor allem als Teil des Verbs: *ich hân gelobet* #swv pV ^.

Bestimmung der Wortform.

- Was ist die **handschriftliche (hsl.) Endung**? Sie besteht aus allen Zeichen, die nach dem Stamm (= Lemma) kommen, wobei aber ein auslautendes *e*, (*e*) nicht zum Lemma gezählt wird: also *quot-er*, *vrî-en*, *vrî-z*, *zës-e*, *milt-ez*, *milt-e*, *quot-NICHTS*, *vrîNICHTS*, *zësNICHTS*, *miltNICHTS* von den Lemmata *quot*, *vrî*, *zës*, *milte*.
- *sancte* bekommt immer ^0.
- Was sind Kasus, Numerus, Genus des Bezugswortes? Was ist die **erwartete Endung**? Was ist die **erwartete Flexionsklasse (stark / schwach)**? Vgl. dazu folgende Tabelle:

Durch die Aufhebung von Zellgrenzen wird angezeigt, wo nach unseren Konventionen keine Angabe des Genus und keine Angabe von *w* ('schwach') steht.

- **Aber bei Endungslosigkeit (angezeichnet durch ^0...) und Formregelverstoß (^...!) wird ...**
... auch im **s c h w a c h e n** NP und AP das **G e n u s** angegeben,⁶¹
- **Aber bei den I-0- und den II-Texten wird außerdem noch ...**
... auch im **s c h w a c h e n** NP und AP das **G e n u s** angegeben,
... im **starken und schwachen GP** das **G e n u s** angegeben,
... zwischen **ASm** und **ASmw** sowie zwischen **DP** und **DPw** unterschieden.

		Starke Deklination			Schwache Deklination		
		Neutr.	Mask.	Fem.	Neutr.	Mask.	Fem.
Sg.	Nom.	(e)z	(e)r	iu ⁶²	e	e	e
	Akk.	(e)z	(e)n	e	e	⁶³	(e)n
	Gen.	(e)s	(e)s	(e)r(e)	(e)n	(e)n	(e)n
	Dat.	(e)m(e)	(e)m(e)	(e)r(e)	(e)n	(e)n	(e)n
Pl.	Nom.	iuFehler: Verweis nicht gefunden	e	e		(e)n	
	Akk.	iuFehler: Verweis nicht gefunden	e	e		(e)n	
	Gen.	(e)r(e)			(e)n		
	Dat.	(e)n					

Erwartete Flexionsklasse (stark / schwach): Formregel: Steht die Adjektivform in einem Nominalglied nach einem Determinativ mit pronominaler Endung, so wird schwache Flexion erwartet. Sonst – wenn kein Determinativ oder keins mit prono-

⁶¹ Also dort, wo *w* 'schwach' nicht aus der Form selbst, sondern aus der Stellung geschlossen ist.

⁶² Statt *iu* kann mitteldt. auch *e* stehen.

⁶³ Statt ASmw wird immer ASm angesetzt.

minaler Endung vorausgeht – wird starke Flexion erwartet. Vgl. *daz stille kint, einem stillen kinde gegen stillez kint, ein stillez kint, stillem kinde, stillem kint disem*.

Besonderheiten:

- Obwohl als #pron bezeichnet, wird **ander** für die Formregel nicht als Determinativ aufgefasst: *ein anderez stillez kint, anderiu stilliu kint*.
- Numeralia außer *ein* werden für die Formregel nicht als Determinativ aufgefasst (das gilt für Kardinalia und Ordinalia): *zwei stilliu kint, zwêne stille man, zwô stille vrouwe; twelve stilliu kint; hundert stilliu kint; ein dritter heiliger man*.
- Abgleich zwischen handschriftlicher und erwarteter Endung. Fallunterscheidung: (XXx steht für die Bestimmung, die der erwarteten Endung entspricht.)

	ohne Endung	mit Endung
attributives Adjektiv, vorangestellt	[^] 0XXx	[^] XXx
attributives Adjektiv, nachgestellt	[^] 0XXx°	[^] XXx°
prädikatives Adjektiv	[^]	[^] XXx=

Vor = ('prädikativ') können noch andere Zusatzanzeichnungen stehen; die Reihenfolge ist die auf der Tastatur von links nach rechts. Vgl. 7.4.5.1.)

Wann liegt eine Endung vor und wann nicht?

- a Es gibt eine hsl. Endung und diese ist gleich der erwarteten:
[^]XXx mit oder ohne w.
- b Es gibt eine hsl. Endung und diese ist gleich der erwarteten, wenn man statt von einer starken Form von einer schwachen Form ausgeht oder umgekehrt. Dann wird die Bestimmung gerade so mit w bzw. ohne w angesetzt, dass die Bestimmung zur hsl. Form passt. Als Zusatzanzeichnung wird ein ! angeschlossen:
[^]XXx! mit oder ohne w gemäß der hsl. Endung.
- c Es gibt eine hsl. Endung und diese ist ungleich der erwarteten:
 - Wenn die hsl. Endung nur *e* beträgt und das Lemma auf *e*, (*e*) oder *w* endet, liegt der Fall "**ohne Endung**" vor: siehe den folgenden Fall d.
 - Sonst wird die erwartete Bestimmung angesetzt und **[^][!]** im Anmerkungsfeld vermerkt: **[^]XXx [^][!]**
- d Es gibt keine hsl. Endung:
 - attributiv:
[^]0XXx mit oder ohne w gemäß der erwarteten Endung; und mit °, wenn nachgestellt.
 - prädikativ: **[^]-**
- e Wenn ein abgekürztes Adjektiv bestimmt wird, wird es wie ein endungsloses behandelt, jedoch **?** statt **0** gesetzt:
[^]?NSm

Beispiele:

dër guot recke [^]0NSmw; *ein milte recke* [^]NSmw!; *diu guot kint* [^]0NPw.

daz edeliu wîp [^]NSnw; *dër milter recke* [^]NSm!.

die hêrren wâren guot [^]-; *die hêrren wâren guote* [^]NPm=; *daz wîp was edeliu* [^]NSn=.

dër recke milt [^]0NSmw°; *dër recke milte* [^]NSmw°; *dër recke quoter* [^]NSm!°; *ein recke quote* [^]NSmw!°.

Zu ergänzen sind gegebenenfalls die im Folgenden vorgestellten Zusatzanzeichnungen.

7.4.5.1 Zusatzanzeichnungen bei Adjektiven

Zur genaueren Beschreibung der Adjektivform, insbesondere ihrer morphosyntaktischen Verhältnisse, werden zusätzliche Auszeichnungen vorgenommen. Diese Zusatzanzeichnungen werden unmittelbar an die Kasus-Numerus-Genus-Klasse-Bestimmung angeschlossen, und zwar in der Reihenfolge, in der sie auf der Tastatur liegen, von links oben nach rechts unten gesehen (z. B.: [^]NSmw!''°). Es handelt sich um folgende Zusatzanzeichnungen:

- ° Die Adjektivform ist nachgestellt.
- ! Die Flexionsklasse der Adjektivform widerspricht der Formregel (s. 7.4.5).
- '' Die Adjektivform steht in einer Anrede.
- % Die Adjektivform wird substantiviert, das heißt, als Kopf ihrer Nominalphrase verwendet.
- & Die Adjektivform wird mit Bezug auf ein Personal- oder Reflexiv- oder *wër-waz*-Fragepronomen im gleichen Kasus gebraucht.
- = Die Adjektivform wird prädikativ verwendet und ist zugleich flektiert.

7.4.6 Pronomina und Artikel

Pronomina und Artikel werden wie Adjektive bestimmt.

Wie *sancte* werden *èinander* und *mit èinander*, *über èinander* usw. nur mit [^]0 bestimmt; gleichbedeutendes *èin* in *mit èin*, *in èin*, *über èin*, *under èin* usw. wird entsprechend als #num [^]0% angesetzt.

Zu *all* und *sëlb* siehe 7.4.6.2.

Die Possessivpronomina können nicht prädikativ verwendet (und lemmatisiert) werden; angesetzt wird der Genitiv des jeweiligen Personalpronomens (zur Problematik vgl. PWG § 364 u. 407).⁶⁴

Folgende Pronomina zeigen gewöhnlicherweise **keine schwachen Formen**: *sum*; *ein* und *sich-èin*; *de-hèin*, *kèin*, *ne-hèin*; *wël(i)ch* mit *iè-wël(i)ch*, *swël(i)ch*, *wël(i)ch*, *sol(i)ch*; *ge-lîch* mit *allermanne-ge-lîch*, *iè-ge-lîch*, *manne-ge-lîch*; *wër* mit *iè-gewër*, *iè-swër*, *ièhtes-wër*, *swër*, *wër*, *ète(s)-wër*; *wëder* mit *de-wëder*, *ent-wëder*, *iè-ge-wëder*, *iè-wëder*, *ièt-wëder*, *wëder*; *lîch* in *manne-lîch*; *-lich* in *iètes-lich*, *sume-lich*, *ète(s)-lich*; *all* und *ëben-all*.

Zur Frage, ob **dës** ein Adverb oder Gen. Sg. neutr. des Personalpronomens ist, siehe 7.4.11.

7.4.6.1 Zusatzanzeichnungen bei Pronomina

Die Zusatzanzeichnungen bei den Pronomina sind dieselben wie die für die Adjektive, hinzu kommen:

- + Die Pronominalform ist ein *R e l a t i v p r o n o m e n*.
- ® (alt 169) Die Pronominalform ist eine Form von *ëër*, die *r e f l e x i v* verwandt wird (z. B. *ëër nam im einen stuol* 'er nahm sich einen Stuhl', *si gab ir dën tô't* 'sie gab sich den Tod').

7.4.6.2 Anzeichnung von *all* und *sëlb*

Die Anzeichnung von *all* und *sëlb* erfolgten bisher nach Sonderregeln. Diese sollen aus Gründen der Vereinheitlichung aufgehoben werden, d. h. die Zusatzanzeichnungen sollen wie bei allen anderen Pronomina nach 6.3.6 und 6.3.6.1 erfolgen, also in möglichst enger Übereinstimmung mit der Adjektivregelung.⁶⁵

⁶⁴ In Sätzen wie *ich bin dîn* faßt Behaghel *dîn* als prädikatives Possessivpronomen auf; das entspricht nicht den im Grammatikprojekt verfolgten Konventionen!

⁶⁵ Bisherige, zu verwerfende Regelung:

NP00 *alle* ist nachgestellt (*die meister alle*)

NP00% *alle* ist nachgestellt und in Fernstellung (*die sullen des alle swern*)

Die Korrektur wurde durchgeführt. Änderungen sind nun nur noch nach Rücksprache durchzuführen. Die neuen Regeln sind:

0.0 Ein vorausgehendes Relativpronomen als Bezugswort führt zu keiner anderen Bestimmung von *all* oder *sëlb* als ein Determinativ. Das heißt, es ist dann % und bei starker Flexion ! zu setzen.

1.1 Steht *all* zusammen mit einem Nomen oder Determinativ in einem Nominalglied, dann gilt es als nicht substantiviert, und zwar auch dann, wenn es in Fernstellung vor oder nach dem Bezugswort steht. Die Nachstellung wird durch ° gekennzeichnet. Anders als bei den Adjektiven gilt auch ein vorausgehendes Determinativ als Bezugswort! Bsp.:

<i>alliu kint koment</i>	all ^NPn
<i>diu kint alle koment / diu kint koment alle</i>	all ^NPn!°
<i>alliu d(is)iu koment</i>	all ^NPn
<i>d(is)iu alle koment / d(is)iu koment alle</i>	all ^NPn!°

1.2 Steht *all* dagegen alleine in einem Nominalglied oder auch neben einem echten Pronomen⁶⁶, gilt es als substantiviert und wird mit % ausgezeichnet. Bsp.:

<i>diu kint,, si alle koment / si koment alle</i>	all ^NPn%&
<i>allez ist ein niht ûf êrden</i>	all ^NSn%
<i>daz ist allez</i>	all ^NSn%

Zu unterscheiden ist also zwischen

– <i>daz ist allez.</i>	all ^NSn%
-------------------------	-----------

(mit *allez* als Prädikatsnomen in der Verbalphrase *ist allez*) einerseits und

– <i>allez daz ist guot</i>	all ^NSn
– <i>daz allez ist guot / daz ist allez guot</i>	all ^NSn!°

(mit *allez* als Teil der Subjekts-NP) andererseits.

Ähnliches gilt für *sëlb*, das aber auch hinsichtlich der Substantiviertheit ganz wie ein Adj. behandelt wird:

2.1 *sëlb* wird entsprechend der Adjektivregelung bestimmt. Vgl. Fälle wie

<i>diu sëlben kint koment</i>	sëlb ^NPw
<i>diu guoten kint koment</i>	guot ^NPw
<i>diu sëlben koment</i> (auch als Relativsatz)	sëlb ^NPw%
<i>diu guoten koment</i>	guot ^NPw%

2.2 Auch im Sinn von 'selbst' wird *sëlb* wie attributives *sëlb* gekennzeichnet (eigentlich ist nach Stellung und Inhalt prädikatives *sëlb* zu unterscheiden):

<i>got sëlbe kam</i>	sëlb ^NSmw!°
<i>got kam sëlbe</i>	sëlb ^NSmw!°
<i>sëlb dër man kam</i>	sëlb ^ONSm

AP00&	nach Personalpronomen (<i>er hat sie alle dar gezogen</i>)
NP00%&	nach Personalpronomen in Fernstellung (<i>sie trinken alle</i>)
NP00!	starke Form nach best. Art./Pron. (<i>die ander alle salmen</i>)
NS11%	vorangestellt, in Fernstellung (<i>alle gotes tougenen</i>)
NP%	<i>alle</i> ist Kopf einer Nominalphrase (wie in 6.3.6.1)
(" , !	wie in 6.3.6.1)

Probleme der bisherigen Regelung:

- Abweichungen gegenüber anderen Konventionen bei Adjektiven und Pronomina.
- Keine deutliche Differenzierung von % als Zeichen der Fernstellung (nur bei *alle*) und % als Kennzeichnung für den Kopf einer Nominalphrase (sowohl bei *alle* als auch in 6.3.6.1).
- Keine Kennzeichnung von Endungslosigkeit.

⁶⁶ Personalpronomen, Reflexiv- oder Rezipropronomen sowie die Interrogativpronomen *wër* und *waz*.

<i>ër sëlbe kam</i>	sëlb ^NSmw!%&
<i>ër kam sëlbe</i>	sëlb ^NSmw!%&

2.3 Entsprechend bei '... selbst als zweiter / dritter usw.' mit prädikativer Ordinalzahl:

<i>got kam sëlbe dritte</i>	sëlb ^NSmw!°, dritte ^NSmw!=
<i>ër kam sëlbe dritte</i>	sëlb ^NSmw!%&, dritte ^NSmw!=

2.4 Eine bemerkenswerte Besonderheit sind Possessivkonstruktionen wie

<i>mîn sëlber hërze</i> (bezogen auf eine Frau)	<i>ich</i> ^GS, <i>sëlb</i> ^GSf%&, <i>hërze</i> ^NS
<i>sîn sëlbes hërze</i> (bezogen auf einen Mann)	<i>ër</i> ^GSm, <i>sëlb</i> ^GSm%&, <i>hërze</i> ^NS

Das vorausgehende Pronomen, der Numerus und das Genus werden auch – mit der Zeit zunehmend – abweichend verwendet: Vgl. *sînes sëlbes hërze* oder *sîn sëlbes hërze* bezogen auf eine Frau. Hierfür gilt:

- möglichst *#pron* statt *#pron poss* ansetzen, *-es*, *-er* wird als Genitivverstärkung am Personalpronomen gesehen, im Anmerkungsfeld */#[!]* vermerkt. *#pron poss* ist nur anzusetzen in Fällen wie *von sînem sëlbes guote* (IV-3-U-Freib-09,10), wenn also offenbare Kongruenz mit dem nichtgenitiven Bezugswort vorliegt.
- *sëlb* wird also auch hier mit & angezeichnet; eine starke Form bleibt ohne !.
- Genus und Numerus aber werden nach der Form bestimmt, eine Abweichung vom Sinn wird im Anmerkungsfeld festgehalten: z. B. */^[statt P]* oder */^[statt f]*.

7.4.7 Zahlwörter

Die Zahlwörter werden wie Adjektive lemmatisiert, das heißt, sie erhalten ggf. eine Doppellemmatisierung mit Zusatzanzeichnungen.

hundert und *tûsent* werden, soweit morphologisch und syntaktisch möglich, als num angesetzt:

<i>hundert/tûsent wildiu ros</i>	<i>hundert/tûsent</i> #num ^ONPn
<i>mit hundert/tûsent knëhten</i>	<i>hundert/tûsent</i> #num ^ODP, <i>knëhten</i> ^DP
<i>tûsente wîghafter manne</i>	<i>tûsente</i> #num ^NPM%

Auch wenn *hundert/tûsent* endungslos ist und der gezählte Gegenstand als partitiver Genitiv vorliegt, wird neuerdings (anstelle des bisher angenommenen Suffixes #n) num angesetzt, das als ^0..n% bestimmt wird. Das Genus ist stets n, unabhängig von dem des abhängigen Genitivs.

<i>èin tûsent/hundert wîghafter manne</i> Gen.attr.	<i>èin</i> ^ONSn, <i>tûsent/hundert</i> #num ^ONSn% +
<i>zwèi tûsent/hundert wîghafter manne</i> Gen.attr.	<i>zwèi</i> ^NPN, <i>tûsent/hundert</i> #num ^ONPN% +
<i>zwèi tûsent/hundert manne</i> Gen.attr.	<i>zwèi</i> ^NPN, <i>tûsent/hundert</i> #num ^ONPN% +

(In diesem Beispiel ist nicht *manne* ausschlaggebend, sondern die neutrale Form *zwèi*.)
tûsent/hundert manne wart erslagen *tûsent/hundert* #num ^ONPN% + Gen.attr.

(In diesem Beispiel ist nicht *manne* ausschlaggebend, sondern die Singularform *wart*.)

Im Unterschied dazu richtet sich das Genus bei allen anderen substantivierten Kardinalzahlen nach dem des partitiven Genitivs:

<i>vièr guoter dëgene</i>	<i>vièr</i> ^ONPM%
---------------------------	--------------------

#n wird nur dann angesetzt, wenn die Endung dazu zwingt, wie etwa in:

<i>sô gibet man iè von dëm hunderte</i>	<i>hunderte</i> #n ^DS
---	------------------------

Ein besonderer Fall liegt vor, wenn der gezählte Gegenstand auf *èin hundert/èin tûsent* folgt und nicht als Genitiv angesetzt werden kann. *èin* wird dann immer als $\wedge 0...Sn$ angesetzt, *hundert/tûsent* dagegen als mit dem gezählten Gegenstand kongruierendes num im Plural:

<i>èin tûsent vröude</i>	<i>èin $\wedge 0NSn$, tûsent $\wedge 0NPf$</i>
<i>èin hundert wort</i>	<i>èin $\wedge 0NSn$, hundert $\wedge 0NPn$</i>

Steht *èin* zusammen mit mehrzahlbezeichnenden Zahlwörtern, wird wie folgt verfahren:

- In *èin und ...* wird Numeruskongruenz erzwungen, indem auch *èin* mit P gekennt wird; zugleich wird die Folge als Phrase mit nur einem Kopf angesetzt:

<i>èin und zwèinzig schillinge</i>	<i>èin $\wedge 0NPm$, zwèinzig $\wedge 0NPm$, schillinge $\wedge NP$</i>
------------------------------------	---

- Flektiertes *èin* wird mit S gekennt; und ineins mit der Numeruskongruenz werden zwei Köpfe angesetzt: ein singularischer und ein pluralischer. Beispiel:

<i>èinez unde niunzig jâr</i>	<i>èinez $\wedge NSn\%$, zwèinzig $\wedge 0NPn$, jâr $\wedge NP$</i>
-------------------------------	---

7.4.8 Übungsbeispiele zur Anzeichnung und Zusatzanzeichnung von Adj, Pron, Num

"einiz ge\$prochin unde zwei uirnommin.", *èin* #num \wedge _____

zwêne #num \wedge _____

"ich vngetrivwe,, waz to\vn ich%.(?) un-ge-triuwe #adj \wedge _____

"Ir holden mine\$ trehtíne\$,, hold #adj \wedge _____

"ir,, chvních riche,, ... ríche #adj \wedge _____

"Nvtta,, rekchen hohgemv\ot(!)" hôh-ge-muot #adj \wedge _____

\$i liezen in unge\$unden ligen un-ge-sund #adj \wedge _____

\$inget,, ir liebesteden,, liëb #adj sup \wedge _____

(D)at corn en/ |\$ede ni(=)man in dich,, reine maget inde \$uuerlich,,

rèin(e) #adj \wedge _____

sûber-lich #adj \wedge _____

<E>in iegilicher,, der [[ze]] [mir] ... iè-ge-lich #pron \wedge _____

an/ |de\ - an(=)deren od' an/ |de\ - trítten mano\ede.

ander #pron \wedge _____

dritt #num \wedge _____

bedenthalp víelen*2 die degen Tote nid'*1 ovf daz lant(.)

tôt #adj \wedge _____

daz er im Demetriv\ - gevan(=)gen vnd lebndigen bræhte.(.)

vâhen #stv7 ^ _____

lëbendig #adj ^ _____

daz in allen landen. niht \$chon's mohte \$ín. schoène #adj comp ^ _____

daz/ |er vns an di(=)\$en u\oche geleret hat. dise #pron ^ _____

lêren #swv ^ _____

de\$ \$ult ir \$in gwi\$\$e(.) ge-wiss #adj ^ _____

di is geliche einem frei\$lichem tiere(.)) ge-lîch #adj ^ _____

vrëis-lich #adj ^ _____

die chvnen helde gv\ote Ovfeín#ander \$lv\ogen(.) küène #adj ^ _____

guot #adj ^ _____

Diz i\$st ze tu\one dise #pron ^ _____

du reine., \$cho(=)ne. ^ _____

er wa\$ rechte kuninc#\$lahte(.) küni(n)g-slaht #adj ^ _____

gîb im daz/ |ze trînchen <a>I\$o lawez.(.) lâw #adj ^ _____

ich trivwelo\$er tri\$tran.(!) triuwe-lôs #adj ^ _____

ich weiz doch warz al\$ den tot.(.) wâr #adj ^ _____

Ir blinden., wederz/ |i\$st mere., ... blind #adj ^ _____

Mîne k(arissi)mi., alle., dêe da gôt meínent., mîn #pron poss ^ _____

vn\ - \$imîliche wurden tîvel(=)\$uhtic.(.) sume-lich #pron ^ _____

tiuvel-sühtig #adj ^ _____

vn\ - leg/ |e\$ al\$to tovmî(=)ge vb' di wnden. toumig #adj ^ _____

vn\ - nim beien⁶⁷., (di man tote vinde in/ |dem honege.,) tôt #adj ^ _____

wolfhart, der mære(.) mære #adj ^ _____

7.4.9 Indeclinabilia

Unflektierbare Wortarten (Konjunktionen, Adverbien, Partikel, Präpositionen) können naturgemäß keine grammatische Bestimmung erhalten; das Feld zur grammatischen Bestimmung bleibt daher leer. Die Feldeingangsmarkierung ^ muß auf jeden Fall erhalten bleiben.

⁶⁷ 'Bienen'.

7.4.10 Eigennamen

Regelungen, die bei der letzten Bearbeitung des Namenbestandes als Richtschnur dienten und auch künftig befolgt werden sollten:⁶⁸

1. Für Personengruppennamen (Kollektivnamen) gilt folgende Regelung: Sie werden im Lemma mit der Endung *-er(e)* (< ahd. *-ari*), in der Normalform dagegen mit *-ære* (< ahd. *-āri*) angesetzt, da sie sich spätestens in frühmhd. Zeit weitgehend den Nomina agentis auf *-ære* < *-āri* angeschlossen zu haben scheinen.
2. Namen mit kontrahiertem *Rèin-* < *Regin-* < *Ragin-* sind mit *Rèin-* und nicht mit *Règen-* angesetzt; daher z. B. *Rèinolt* neben *Règenolt*.
3. Namen auf *-olt* < *-walt* und *-olf* < *-wolf* werden nicht segmentiert; desgleichen werden Namen mit regelmäßiger stärkerer Fugenassimilation (z.B. *Lambrèht* < *Lant-berht*) in der Regel mit Assimilation und unsegmentiert angesetzt, also z.B. *Lambrèht*. Sind dagegen nur der letzte Konsonant des Erstglieds und der erste Konsonant des Zweitglieds von der Assimilation betroffen, wird wie im Beispiel *Râ<t>-bald* für hs.lich *Rapolt* verfahren.
4. Mhd. bereits regelmäßige Syn- und Apokopen sind nicht berücksichtigt, so z.B. bei den Namen mit *Wèrn-* < *Warin-* und *-hèr* < *-hari* < **-harja-* statt *Wèr(e)n-* bzw. *-hèr(e)*.
5. Namen auf *-bote* < ahd. *-boto* und *-bôt* < germ. **-baud* sind wegen der weitgehenden Vermischung beider Zweitglieder ggf. als *-bo^t(e)* zusammengefasst, z.B. *Sige-bo^t(e)* für *Sige-bote/Sige-bôt*.
6. Namen mit *-kofen* < *-ing-hofen* sind belegnah mit *-koven* und nicht mit *-ing-hoven* angesetzt.
7. Bei fremden Namen wird neben dem fremdsprachlichen Lemma dort ein (partiell) mhd. Lemma angesetzt, wo (partielle oder komplette) Integration vorliegt, und sei es auch nur in Form der Tilgung von lat. *-us*, z.B. in *Urbân* neben *Urbanus*. Eine Entscheidung ist hier oft zwangsläufig unsicher und willkürlich, z.B. bei *rome* Genitiv/Dativ →

⁶⁸ Veraltetete o d e r ergänzende Hinweise:

- Allgemein: Appellativ oder Proprium?
Die Unterscheidung zwischen Eigennamen und Appellativen ist in der Sprachwirklichkeit nicht trennscharf, das heißt, es gibt Fälle des Zweifels, ob ein Eigenname oder ein Appellativum anzusetzen sei.
- Vorzuziehen ist der Ansatz als Appellativum. Ein Appellativ wird angesetzt, wann immer sich ein passendes in der Lemmaliste findet. Bei deutlich aufstoßender Verwendung eines solchen Appellativs als Eigenname(nzusatz) ist */#[BN]* für Personen bzw. */#[FIN]* für Örtlichkeiten im Kommentarfeld zu vermerken und – falls sich diese Frage stellt – eher kein Appositionskomma zu setzen.

Bsp.: *Hinz, d̥er bëcke,); Hinz, bëcke,); der bëcke Hinz; bëcke Hinz*. – Ob Hinz wirklich Bäcker war, ist für den Ansatz belanglos und lässt sich in den meisten uns vorkommenden Fällen auch gar nicht entscheiden.

- Namen, auch Einwohnernamen, werden als Appellative mit Großbuchstaben angesetzt.
- Wenn sich kein Appellativ ansetzen lässt:
 - PN/BN oder ON/FIN? Oder EN?
Entscheidet sich danach, ob von einer Person oder einer Örtlichkeit die Rede ist. Eigennamen ohne weitere Unterscheidung (*#EN*) werden nur im Ausnahmefall angesetzt: dann, wenn nicht zwischen Person und Ort entschieden werden kann.
Eine Besonderheit sind Ortsangaben durch den DP von Gruppennamen: *swâben* '(bei den) Schwaben'. Wenn eine solche Ortsangabe vorliegt, wird z. B. angesetzt:
a) *`Vranke @m `Vranke #m ^DP ... \vraken ... :/'[als ON]*
b) *`[!] @[!] `Gabeling #m ^DP ... \gabelingen ... :/'[als ON]*
und zwar a) dann, wenn es einen Eintrag im LRG gibt, b) sonst.

- BN oder PN? FIN oder ON?
Auf diese immer schwanke Unterscheidung wird kein Berichtigungsaufwand mehr gewendet.

Roma oder *Rôme* (wo nicht der deutsche bzw. lat. Kontext zugunsten von *Rôme* bzw. *Roma* (*Romae*) spricht).

8. Fremde Namen, die sowohl lateinisch als auch mhd. flektiert werden, erhalten nur einen gemeinsamen Lemmaansatz, der im Gravis-Lemma je nach Integrationsgrad mit oder ohne * erfolgen kann; für diese Entscheidung lassen sich leider kaum alle Fälle erfassende strenge Richtlinien angeben,⁶⁹ doch ist dies wegen der folgenden Regelung auch nicht von allzu großer Bedeutung: Die Flexionsformen mit lateinischer Endung erhalten in jedem Falle im Trema-Lemma einen Stern, während die mhd. flektierten in jedem Fall ein sternloses Trema-Lemma erhalten, z. B.
``Alexander @PN `Alexander* #PN ^DS ~Alexandrô \alexandro`
``Alexander @PN `Alexander #PN ^DS ~Alexand<e>ren \alexandren`
9. Auf Angabe der Vokallänge und Präzisierung der kurzen *e*-Laute wird bei fremden Namen verzichtet, soweit nicht der Reimgebrauch eine Vereindeutigung erlaubt. Gelegentlich ist hier schon – und ähnlich auch bei unsicherer nativen Formen – das „Deckel-e“ *É* verwendet worden. Dies sollte künftig systematisch geschehen.
10. Nachgesetztes *^* wird bei nativen Namen gelegentlich dort gebraucht, wo die Vokalquantität nicht feststeht, z.B. *a^* = *a* oder *â*.

Verbleibende Inkonsistenzen und Fehler:

1. Fremde Namen (und zwar nicht nur die partiell integrierten, s.o. 1.7) sind in früheren Projektphasen – teils unter dem Einfluss der Namensschreibungen in den Standardausgaben – partiell vermittelhochdeutsch worden, insbesondere was die Verwendung von Langvokalen und von *ë* bzw. *è* anbelangt. Bei der Überarbeitung der Namenformen ist darauf weitestgehend verzichtet worden; Ausnahmen sind meist nur dort gemacht worden, wo der Reimgebrauch die Vokalquantität oder -qualität vereindeutigt, z. B. *Artûs*, *Lûnête*. In vielen Fällen sind außerdem die alten Projekt-Namenformen unverändert übernommen worden, z. B. *Jêsus*.
2. BN sollten einer neueren Absprache zufolge als Appellativa angesetzt werden, wo dies möglich ist; die BN-Verwendung soll dann im Kommentarfeld angezeigt werden. Noch günstiger wäre es wohl, sie im Gravis-Lemma als Appellativum, im Trema-Lemma dagegen also BN anzusetzen. Dies setzt allerdings voraus, dass es auch bei EN in aller Regel ein ausgefülltes Gravis-Feld gibt.
3. Eine konsequente Durchführung der Unterscheidung zwischen primärer Wortart und belegspezifischer BN-Verwendung würde außerdem eine sehr arbeitsaufwändige Prüfung aller BN- und PN-Belege erfordern.
4. Dasselbe Problem stellt sich für Ortsnamen (im weiteren Sinne, also inklusive Flurnamen und Wohnstättenbezeichnungen, die ihrerseits wieder z.B. Flurnamen oder Häusernamen sein können), wo sie in Beinamen nach der Herkunft oder der Wohnstätte erscheinen; auch sie sollen konsequent als ON klassifiziert, aber hinsichtlich der belegspezifischen Verwendung als BN markiert werden.
5. Probleme bereitet ferner – wie üblich – die Abgrenzung zwischen EN und appellativischen Monosemantika und Ähnlichem (z.B. Bezeichnungen von Gestirnen wie *Mars*, *Saturn(us)*, *Venus*); jetzt ist z.B. sowohl *Mars* PN (der Gott) und *Mars* ON (der Planet) angesetzt worden. Wenn Objektnamen wie Buchtitel (z.B. *Leviticus*, *Eneide*) als Ergonyme gewertet werden, sollte für sie das Kürzel EN verwendet werden, desgleichen für alle anderen Namen, sich nicht PN, ON oder BN zuordnen lassen. Im Zweifelsfall sollte jedoch – wenn immer dies möglich ist – ein Appellativum angesetzt werden.
6. Uneinheitlich sind bislang die lat. Flexionsendung gehalten, weil teils die lat. Quantitäten, teils auch Reimbelange (z.B. DS *o*-Dekl. *-ô* : *-ô*) berücksichtigt worden sind, teils aber auch nicht, daher: *-î* ~ *-i*, *-ô* ~ *-o*, GDS *â*-Dekl. *-ê* ~ *-ae*, *-e*. Vielleicht sollte hier zu *-î*,

⁶⁹ Schon im Gravisfeld erhalten aber alle Namen einen Stern, bei denen schon die Form des Nom. Sg. eine eindeutig fremde Endung hat, so z. B. die lat. Namen auf *-us*, *-a*, *-o* usw.

-ô, -ê hin vereinheitlicht werden.

7.4.11 “Kleine Wörter” und einzelne Wendungen

Besondere Schwierigkeiten bereitet vielfach die Lemmatisierung der “kleinen Wörter”, und zwar v. a. die Unterscheidung von Adverbien, Konjunktionen und Partikeln – schon im “Renner” heißt es: *von kleinen worten hube sich die not* (11508). Folgende Erläuterungen mögen dabei hilfreich sein, bedürfen aber sicher in noch vielen Fällen der Präzisierung. [[Doppelt eckig geklammert sind etwas abweichende, ungültige Vorschläge.]]

aber

adv: “aber, dagegen; abermals, wieder”

konj: “aber, dagegen; abermals, wieder”, einen Nebensatz einleitend (selten)

al

adv: “ganz und gar”

konj: “wie sehr”, “obgleich”; auch zur Einleitung von Konzessivsätzen, wiewohl das finite Verb direkt folgt

al-same

adv: “ebenso”, “ebenso wie”

konj: “wie”, “wie wenn”, “als ob” (mit Konj.)

al-sô

adv: “so” (im Vergleich), z. B. *alsô grôz alsô ein boum*; vgl. auch PWG §§ 464f.: “wie”, “wie wenn”, “als ob” (vergleichend); “so” – *si striten alsô sêre daz al diu burc erscal*

konj: “wie” (im Vergleich); *alse ein burc*; *alse hute*; *alse der mâne vil verre über lant liuhtet des nahtes*; vgl. auch PWG §§ 459-61: “sowie”, “als” (temporal-konditional) – *alse es danne zit si, so bin ich unde Isot da bi*; *als er die stat ersach, ... er dô sprach*; “wenn” (konditional) (selten) – *also morne ein gros hochgezit solte sin*; “wie” (konzessiv) (“wie ... auch immer”) – *also lip er deme uater was, got ... hîz ime oppheren daz kint*

al-ène

adv: “allein”

konj: “obgleich”, “obschon” als Einleitung von Konzessivsätzen

âne

adv: “ledig”, “allein” bei werden, bleiben, sein (*âne wësen* etc.), u. U. mit abhängigem Genitiv

konj: “außer” (in der Regel nur nach neg. (fragenden) und vor verkürzten Sätzen)

bèide

konj: “sowohl ... als auch”

num: “beide”

biz

adv: vor Präpositionen, z. B. *biz an*, *biz ze*, *biz ûf*

konj: “bis” Nebensatzeinleitend

präp: “bis” NP-regierend (*bis nächsten Donnerstag*)

danne

adv: “dann”, “danach”, “darauf”, “sodann”, “daraufhin”

konj: Nebensatzeinleitend: “als”, “wenn” – *diu krône ist elter*, *danne der künec Philippes si*

partikel: PWG § 396: zur Einführung des verglichenen Gegenstandes nach Komparativ – *wîzer danne snê*; PWG § 447: als formales Charakteristikum exzipierender Bedingungssätze: “es sei denn, daß” – *Ich lasse Dich nicht, Du segnest mich denn*. (Gen 32,26)

dannen

adv: “von da”, “daher”, “deshalb”, “davon”, “woher” (Herkunft)

[[konj: Nebensatzeinleitend]]

dâr

adv: “da”, “dort”, “wo” (lokal). *dâr* und *dar(e)* werden in den Handschriften auf unterschiedliche Weise unterschieden; eine Differenzierung muß nach dem Sinn und unter Beachtung des jeweiligen Usus der Handschrift erfolgen.

Bei Doppeladverbien allerdings ist grundsätzlich *dâr* als Ansatz zu wählen, nicht *dar(e)*.⁷⁰

In obd. Texten dürfen keine Formen von *dâr/+in* angesetzt werden, sondern nur *dâr/+inne* (lokal) oder *dâr/+în* (Richtung). In md. Texten darf hingegen auch *dâr/+in* vorkommen.

[[konj: Nebensatzeinleitend]]

dar(e)

adv: “dahin”, “wohin” (Richtung). Vgl. auch *dâr*.

[[konj: Nebensatzeinleitend]]

dës

adv: “daher”, “deshalb” (kausal). *dës* soll als Adverb nur angesetzt werden, wenn eine Lemmatisierung als Gen. Sg. neutr. des Pronomens *dër* nicht mehr möglich ist; im Zweifelsfall ist die Lemmatisierung als Gen. Sg. neutr. vorzuziehen (vgl. PWG § 218, Anm. 7; Beispiel: *er badet' sich in dem bluote: sîn hût wart hûrnîn. des snîdet in kein wâfen*. [NL, HS B,100,3-4]).

dô

adv: “dann”, “da”, “damals”, oft nach vorausgehendem, mit *dô* (konj.) eingeleitetem Temporal-satz

konj: Nebensatzeinleitend⁷¹

doch

adv: PWG 445: “doch”, “dennoch”, “trotzdem” (konzessiv) (in einem übergeordneten Satz) – *wirt mir dîn meister nimmer holt, dîns amts du doch geniezen solt*

konj: PWG 461: “obgleich”, “wiewohl”, “obwohl” (konzessiv) (ursprgl. mit Konj., später auch Ind.)

êr

adv: “früher”, “vormals”, “eher”, “lieber”

konj: “bevor”, “eher als”, “ehe”

präp: “vor” (temporal, mit Gen. oder Dat.)

êrest

Die Wendung *von êr(e)st(e)* wird als zwei Adverbien aufgefaßt und lemmatisiert.

halb

Zahlangaben wie *dritte-halb*, *ahte-halb* werden (derzeit) als Komposita aufgefaßt und ggf. als solche angesetzt. Ein Problem besteht allerdings darin, daß potentiell sehr viele dieser Zahlwörter angesetzt werden müßten; andererseits sind sie nicht sonderlich frequent.

hin(e)

hin(e) nider, *hin(e) ab* *a u s e i n a n d e r*.

iè-doch

adv: “doch”, “dennoch”

konj: siehe unter *doch*

in allen gâhen

Obgleich nach Lexer die Wendung auf einen ahd. Akk. Sg. zurückgehen soll (*in ala gâhûn*), ist die Beleglage im Ahd. alles andere als eindeutig (auch Dativ, adverbial). Entsprechend der für die mittelhochdeutsche Zeit anzunehmenden synchronen Auffassung soll die Wendung als Präposition mit nachfolgendem Dativ Pl. analysiert werden.

în

adv: “hinein”

inne

adv: “innen”

joch

⁷⁰ Die einzige Ausnahme ist der “Wiener Notker”, weil dort die handschriftlichen Formen entsprechend abweichende Ansätze fordern.

⁷¹ Alt: “PWG 459: “als”, auch “indem” oder “nachdem” (temporal) (nur mit dem Präteritum!)”.

adv: "auch", "sogar" (einen Konzessivsatz oder Hauptsatz ("doch") oder eine Frage einleitend
konj: "und", "und auch", "nebst", "sowie"; "noch", "weder ... noch" – *daz si endrunken joch enâzin*

lützel

adv: wenn kein Genitiv von *lützel* abhängt und es kein Adjektiv ist
n: wenn ein Genitiv von *lützel* abhängt oder wenn *lützel* Subjekt oder Objekt ist

mære (mfrk.) = mhd. *wan*

adv: "aber"

konj: "außer"

mêr(e)

adv: "mehr"

n: wenn ein Genitiv von *mêr* abhängt oder wenn *mêr* Subjekt oder Objekt ist, und zwar auch in *mêr(e) danne*.

niht

adv: "nicht"

n: wenn ein Genitiv von *niht* abhängt oder wenn *niht* Subjekt oder Objekt ist

niuwan

adv: "nur", "nichts als"

konj: "außer", "ausgenommen", z. T. ellipt. konstruiert

noch

adv: "noch" (die Fortdauer von einem Zeitpunkte an kennzeichnend; "noch mehr")

konj: "noch" als Negation ("weder ... noch")

nû

adv: "nun", "jetzt", "eben jetzt"

konj: "da", "als nun", "während"

obe

adv: "oben", "oberhalb", "über"

konj: "wenn", "wenn auch", "falls", "ob"

a b e r n i c h t f ü r ' o d e r '

präp: "über", "oberhalb", "auf" (mit Dat.)

oder

konj: 'oder', auch wenn Ausdrucksseite abweicht (*ove, obe, ave(r), ald(er)* u. a.)

ouch

adv: "auch", "überdies"; PWG 479: zur Kennzeichnung konzessiver Nebensätze – *sî ouh ih ie-mer ze allen êren verkorn, nû was ih doh ze Rôme ain rîche man*

konj: "auch", "überdies", "ferner", "noch mehr" zur Einleitung eines neuen Satzes – *ouch wart sîn herze jâmers vol*

sam(e)

adv: "so", "ebenso"

konj: "wie", "wie wenn", "als ob"

samen(e)

adj: "gesamt"

adv: "zusammen"

same(n)t

adv: "beisammen, zusammen, zugleich"

präp: "zusammen mit"

sâ(r)

ze stunt, sâ(r) ze stunt hinfort a u s e i n a n d e r .

sint (Nebenform zu sît)

adv: "seitdem", "darauf", "darnach", "späterhin"

konj: "seit", "seitdem", "nachdem", auch kausal "da", "weil"

präp: "seit" (mit Gen. oder Dat.) – *sint dem mâle*

sît

adv: "seitdem", "darauf", "darnach", "späterhin"

konj: "seit", "seitdem", "nachdem", auch kausal "da", "weil"

präp: "seit" (mit Gen. oder Dat.) – *sît dem mâle*

sô

oft Kurzform für *al-sô*, vgl. oben

adv: "so" (im Vergleich) PWG 462: im Nachsatz adverbiales Bezugswort zu einem mit *wande* eingeleiteten Vordersatz, in dieser Stellung "so gut wie ohne semantischen Wert" – *wandich den trost von dir han, sô wil ich di rede understan*; PWG 464: "so" in einem übergeordneten Satz als Bezugswort für einen mit *daß* eingeleiteten abhängigen Satz ("so ..., daß ...")

konj: "wie" (im Vergleich); PWG 459: "als", "sowie", "[dann] wenn" (temporal; auch temporal-konditional); PWG 465: "wie", "dementsprechend wie", "so wie" (Einleitung modaler Sätze); selten "in der Weise, daß"

partikel: relativische Partikel, vor allem in der Stellung nach verallgemeinerndem *swer*, *swie* etc. (entstanden aus Wendungen wie *sô swer sô*) (vgl. PWG § 451).

stunt

In *drî stunt* etc. wird das Zahlwort als endungslos (#num ^0), *stunt* als Adverb lemmatisiert.

ze stunt, *sâ(r) ze stunt* hinfort *a u s e i n a n d e r*, aber ohne nachträgliche Vereinheitlichung.

sunder

adj: "abgesondert", "allein", "eigen", "besonders"

adv: "abgesondert", "allein", "einsam", "ausschließlich", "eigen", "auf besondere Weise"

konj: "ausgenommen", "außer", "gleichwohl", "vielmehr", "indessen", "aber", "sondern"

präp: "außer", "ohne" (mit Akk.)

swanne

konj: "wann (immer)", "wenn"

Jüngeres *wanne*, *wër* etc. in der Verwendung wie älteres *swanne*, *swër* etc. (d.h. nicht als Fragewort!) werden als *swanne*, *swër* etc. lemmatisiert. Echtes *wanne*, *wër* etc. kommen mhd. nur in direkten oder indirekten Fragesätzen vor; wenn sie sich durch verdeutlichendes "wann **auch immer**, wer **auch immer**" paraphrasieren lassen, sind *swanne*, *swër* anzusetzen.

swannen

rel adv [[adv und konj]]: "woher auch immer" (vgl. PWG § 450)

swâr

rel adv [[adv⁷² und konj]]: "wo auch immer" (vgl. PWG § 450)

swar(e)

rel adv [[adv und konj]]: "wohin auch immer" (vgl. PWG § 450)

swiè

konj: ist dort anzusetzen, wo die Bedeutung 'wie auch immer' oder 'wenn auch' ist; dann ist nicht *wiè* anzusetzen.

unde

adv: "unten", "unter", "hinunter"

konj: "und"; auch als relative Partikel (vgl. PWG § 451) und als modalvergleichende Konjunktion (vgl. PWG § 465,3). Die relative Partikel sollte durch den Zusatz "[rel.]" markiert werden.

unz(e)

adv: "solange"; vor Präposition (z. B. *unz an* siehe oben unter *biz*)

konj: "bis", "solange als"

vil(e)

n: im Zweifelsfall, gewiss aber, wenn ein Genitiv von *vil(e)* abhängig ist oder wenn *vil(e)* Subjekt oder Objekt ist

adj: wenn adjektivisch flektiert

⁷² In Pronominaladverbien.

adv: wenn keine Flexion vorliegt und kein Genitiv von *vil(e)* abhängig ist

wan

konj: "sondern", "aber"

partikel: nach *niht* in Verbindung mit einem Komparativ oder mit *ander* (PWG § 396); einleitender Ausruf in einem Wunschsatz im Konj. Präs. (PWG § 322; "o, daß doch") – *wolde got wan wære ich tôt*

adv: "bis", vor Präpositionen (siehe oben unter *biz*)

wande

konj: "denn", "weil"

wanne

int adv: "wann?" – *wenne sul wir varn?*

konj: [nebensatzeinleitend] "wann irgend", "wann auch", "sobald" (auch *swanne*, *swenne*)

wannen

int adv: "woher?", "von wo?" – *wannan kument dirre sûzen rôsen smac?*

[[konj: nebensatzeinleitend]]

wâr

int adv: "wo?" – *wâ bistu Adam?* Auch in Verbindung *wâr umbe*.

[[konj: nebensatzeinleitend]]

war(e)

int adv: "wohin?" –4 *war ist nû chomen diu manicvaltîu wîsheit?*

[[konj: nebensatzeinleitend]]

wiè

s. *swiè*.

wol(e)

adv: "gut", "wohl", "sehr"

ze

adv: in Verbindungen wie *ze vil*, *ze jungest(e)* (vgl. PWG § 211, Anm. 2)

präp: "das als präp. gebrauchte adv. *zuo*" (Lexer); Präp. mit dem Dat., selten mit dem Akk., sowohl räumlich als auch zeitlich gebraucht; außerdem stets in Verbindung mit dem Infinitiv (der nach *ze* stets als iD lemmatisiert wird, unabhängig davon, ob die handschriftliche Wortform noch eine Flexionsendung aufweist oder nicht)

ze stunt, *sâ(r) ze stunt* hinfort immer *a u s e i n a n d e r*, aber ohne nachträgliche Vereinheitlichung.

zuo

adv: u. a. als Bestandteil von Partikelverben

Als mehrgliedrige **Konjunktionen** gelten – in Anlehnung an die "Mittelhochdeutschen Grammatik" von Hermann Paul in der Bearbeitung durch Hugo Moser und Ingeborg Schröbler – *biz~daz*, *dâr~mite~unde*, *dâr~wider~unde*, *die~wîle~daz*, *durch~daz*, *ê~daz*, *ê~danne~daz*, *vûr(e)~daz*, *innen~dës*, *noch~denne~daz*, *nû~daz*, *sît~daz*, *sît~dô*, *sô~iè*, *ûf~daz*, *umbe~daz*, *under~dës*, *unz~daz*, außerdem *wan~daz*.

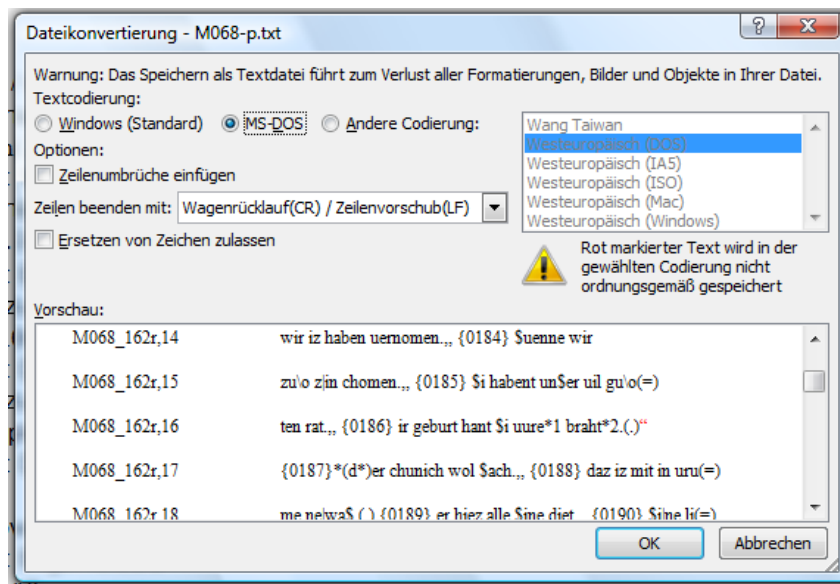
8 Die automatische Vorlemmatisierung

Nach Abschluß der Präeditierung beginnt die automatische Vorlemmatisierung. Zunächst muß jede Wortform des zu lemmatisierenden Textes zeilenweise isoliert und mit einem eigenen Ordnungsteil versehen werden; in mehreren Schritten werden dann einer jeden Wortform mögliche Lemmata und grammatische Bestimmungen zugeordnet. Das Ergebnis dieser automatischen Vorlemmatisierung muß dann manuell überarbeitet werden, damit schließlich jeder einzelnen handschriftlichen Wortform genau ein Lemma und genau eine grammatische Bestimmung zugeordnet ist.

8.1 Dateikonvertierung von Word- in ASCII-Text

Hat die Textdatei ein Word-Format (*.doc, *.docx, *.rtf), so muss sie zunächst in eine reine Textdatei umgewandelt werden.

Dazu die Option "Speichern unter" benutzen und als Dateityp "Nur Text" wählen. Nach Klick auf "Speichern" erscheint der Bildschirm "Dateikonvertierung". Statt "Windows (Standard)" muss hier "MS-DOS" angeklickt werden:



Falls die Word-Datei nicht ASCII-kompatible Elemente enthält – insbesondere Sonderzeichen –, so erscheint neben einem gelben Warndreieck die Meldung:

Rot markierter Text wird in der
gewählten Codierung nicht
ordnungsgemäß gespeichert

und in der "Vorschau" wird das erste dieser Elemente rot hervorgehoben. Meist handelt es sich um die graphischen Anführungszeichen „ “ ‘ ’ ”.⁷³

Dann genügt es in der Regel das Kästchen „Ersetzen von Zeichen zulassen“ anzuklicken. Verschwinden dann Warndeck und –meldung, so werden „ “ ” generell durch " ersetzt, was akzeptabel ist, da in unseren ASCII-Dateien nicht zwischen führenden und schließenden Anführungszeichen unterschieden werden.

Andere Nicht-ASCII-Elemente müssen vor der Konvertierung durch reine ASCII-Kodierungen ersetzt werden, z. B. ũ durch u\o.

8.2 (Zur Dateikonvertierung zwischen Excel- und ASCII-Text)

Im gegebenen Zusammenhang wichtig sind Unterschiede zwischen der Verschlüsselung in "Word-Perfect-Office-Editor"-Dateien und der Verschlüsselung in "Microsoft-Office"-Dateien (z. B. xls-Dateien und doc-Dateien). Für die ersten 128 ascii-Zeichen gibt es keinen Unterschied, wohl aber für die folgenden. In den Projektdateien kommen zwar im Wesentlichen nur Zeichen aus dem Bereich bis zum 128. Zeichen vor, allerdings auch Zeichen jenseits davon (z. B. ", ä, ö, ü). In der "Word-Perfect-Office-Editor"-Verschlüsselung (MS-DOS (PC-8)) werden diese Zeichen nicht richtig dargestellt, wenn die Datei in "Excel" oder "Word" geöffnet wird – und die "Microsoft-Office"-Verschlüsselung wird hinwiederum vom "Word-Perfect-Office-Editor" nicht richtig aufgelöst.

⁷³ Will man dieses Problem vermeiden, so kann man vor der Erstellung der Textdatei mit Word unter den *AutoKorrektur-Optionen* die Registerkarte *AutoFormat während der Eingabe* anklicken und unter *während der Eingabe ersetzen* das Häkchen bei "Gerade" Anführungszeichen durch "typographische" entfernen.

Allerdings können "Excel" und "Word" beim Öffnen und Speichern einen Verschlüsselungswechsel vornehmen, so dass Text von "Word-Perfect-Office-Editor"-Dateien auch in "Excel" oder "Word" richtig dargestellt und umgekehrt auch ein in "Excel" oder "Word" richtig dargestellter Text wieder so abgespeichert werden kann, dass er dann im "Word-Perfect-Office-Editor" richtig dargestellt wird.

Doch "Excel" führt bei Ein- und Ausfuhr oft auch einige unerwünschte Änderungen durch. Die unerwünschten Ausfuhränderungen lassen sich bei einer Ausfuhr von Hand nicht vermeiden und nur mühsam rückgängig machen. Und auch die Einfuhr von Hand ist nicht ohne überraschende Schwierigkeit. Manchmal wird Zellinhalt von Excel nämlich nur im Überblicksfenster oben richtig dargestellt, in der Zelle aber nur eine lange Folge von Gattern (####...) – und es sind diese Gatter, die bei einer Rückwandlung in eine reine Textdatei gespeichert werden! Die Fehldarstellung lässt sich nicht durch Zuweisung des Zellformates "Text" verhindern, sondern unter anderem und am besten durch die eines benutzerdefinierten Zahlenformates wie "#.#####" (andere voreingestellte Formate führen nämlich wieder zu anderen unerwünschten Änderungen).

Es ist also nur ratsam, Ein- und Ausfuhr mit den _müle-Makros "müle019a1_EinfuhrNachExcel" und "müle019b1_AusfuhrAusExcel" durchzuführen. Sie erledigen den Verschlüsselungswechsel und etwaige nachträgliche Bereinigungen, ohne dass der Nutzer eingreifen müsste. Der Nutzer muss nur gegebenenfalls am Ende den Dateinamen ändern.

8.3 Die Erstellung des WOK-Indexes

Zunächst ist aus dem präeditierten Text der WOK-Index (= **W**ortform **o**hne **K**ontext) zu erstellen. Mit dem Programm LEM, Option 1, wird der fortlaufende Text in einen Vertikaltext umgewandelt, indem die einzelnen Wortformen zeilenweise isoliert werden.

<u>Programm:</u>	LEM, Option 1
<u>Eingabe-Datei:</u>	X.TXT
<u>Ausgabe-Datei:</u>	Y-WOK ⁷⁴

Der WOK-Index beginnt mit einem Header für Indexdateien, in den bereits wesentliche Informationen aus dem Header der Eingabedatei automatisch eingefügt sind; dieser Index-Header wird von den folgenden Programmen übernommen, er kann also sukzessive mit Informationen gefüllt werden.

Im WOK-Index ist jeder Wortform der Ordnungsteil der jeweiligen Textzeile zugeordnet, erweitert durch eine Zählung der Wörter pro Zeile bzw. Vers. Diese Wortzählung wird, getrennt durch einen Punkt, unmittelbar an die Vers- oder Zeilenzahl angeschlossen; das jeweils letzte Wort einer Zeile wird durch den "Klammeraffen" @ markiert. Aus Gründen der Übersichtlichkeit können die Korpussigle und das Textkürzel im WOK-Index und den folgenden Arbeitsdateien fortgelassen werden.

Jede Zeile des WOK-Indexes beginnt mit dem Backslash \ als Feldeingangsmarkierung und der Wortform; getrennt durch einen [TAB] und nach dem Unterstrich _ als Feldeingangsmarkierung folgt der Ordnungsteil.

<u>Beispiel:</u>	
\{*S*}wer	_0001.01
\<an>	_0001.02
\rechte	_0001.03
\gv\ete.	_0001.04@
\wendet	_0002.01
\\$in	_0002.02
\gemv\ete.,,	_0002.03@

⁷⁴ Im Folgenden steht **X** für die oben in erläuterte Sigle als Dateinamen der Textdateien, **Y** für die zugehörige, oben in 3.2.3 erläuterte Sigle für Indexdatei-Namen, also z. B. beim Lucidarius **X** für **III-3-P2** und **Y** für **LUCI**.

\dem	_0003.01
\volget	_0003.02
\\$ælde	_0003.03
\vn\-	_0003.04
\êre.(.)	_0003.05@

8.4 Die automatische Vorlemmatisierung

Auf den WOK-Index werden die Programme LEM-2, Option 2, und LEM-3 angewendet, um jeder handschriftlichen Wortform mögliche Lemmatisierungen zuzuordnen.

Zunächst werden mit dem Programm LEM-2, Option 2, allen handschriftlichen Wortformen in gewissem Rahmen "normalisierte" Formen zugeordnet. Es handelt sich dabei nicht um normalisierte Formen im Sinne des Normalmittelhochdeutschen, sondern nur um gewisse Vereinfachungen und Standardisierungen, die die Vielfalt möglicher handschriftlichen Schreibungen reduzieren; so werden etwa alle Abkürzungen und Superskripte aufgelöst, Majuskeln durch Minuskeln und Schaft-s durch Rund-s ersetzt.

<u>Programm:</u>	LEM-2, Option 2
<u>Eingabe-Datei:</u>	Y-WOK
<u>Ausgabe-Datei:</u>	Y-VL.1
<u>Option 1:</u>	Tilgung weiterer Zeichen in der Normalform (i. d. R. nein)
<u>Option 2a:</u>	Text ist bereits normalisiert (i. d. R. "nein")
<u>Option 2b:</u>	keine Veränderung von â, ê, î, ô, û etc. (i. d. R. "nein")
<u>Option 3:</u>	zusätzliche Endsilbennormalisierung (nur bei frmhd. Texten "ja")
<u>Optionen 4-8:</u>	Berücksichtigung schreibsprachlicher Besonderheiten

Die "normalisierten" Formen stehen in der Ausgabedatei jeweils direkt am Zeilenanfang; die handschriftliche Wortform und der Ordnungsteil folgen jeweils durch einen [TAB] getrennt.

Beispiel:

```

swer \{*S*}wer _0001.01
an \<an> _0001.02
rehte \rehte _0001.03
gvete \gv\ete. _0001.04@
wendet \wendet _0002.01
sin \$in _0002.02
ge-mvete \gemv\ete.,, _0002.03@
dem \dem _0003.01
volget \volget _0003.02
selde \$ælde _0003.03
unn \vn\ _0003.04
ere \êre.(.) _0003.05@

```

Anschließend muß die Ausgabedatei Y-VL.1 alphabetisch sortiert werden.

<u>Programm:</u>	SORTA
<u>Eingabe-Datei:</u>	Y-VL.1
<u>Ausgabe-Datei:</u>	Y-VL.1S
<u>Feldbegrenzer:</u>	[TAB] (= [ENTER])
<u>Sortierreihenfolge:</u>	1 (= [ENTER])

Die alphabetische Sortierung der "normalisierten" Formen ist zwingende Voraussetzung für die automatische Lemmazuweisung.

Beispiel:

```

an \<an> _0001.02
dem \dem _0003.01

```

```

ere \êre.(.) _0003.05@
ge-mvete \gemv\ete., _0002.03@
gvete \gv\ete. _0001.04@
rehte \rehte _0001.03
selde \$æelde _0003.03
sin \$in _0002.02
swer \{*S*}wer _0001.01
unn \vn\ _0003.04
volget \volget _0003.02
wendet \wendet _0002.01

```

Auf die alphabetisch sortierte Datei Y-VL.1S wird das Programm LEM-3 angewendet. Es vergleicht die "normalisierten" Wortformen der Datei Y-VL.1S mit den Einträgen im jeweils gewählten Lemmatisierungswörterbuch und ersetzt ggf. die "normalisierte" Wortform durch die entsprechend zugeordneten Lemmatisierungsangebote aus dem Lemmatisierungswörterbuch. Wenn eine "normalisierte" Wortform im Lemmatisierungswörterbuch nicht gefunden wird, greift das Programm auf das jeweils zugehörige Ergänzungswörterbuch zu. Wenn dort auch kein Lemmatisierungsangebot zu finden ist, wird die "normalisierte" Form als provisorisches Lemma angesetzt und muß manuell weiterbearbeitet werden (sog. "Löcherstopfen", siehe 8.6).

Derzeit kann je nach zu lemmatisierendem Text zwischen einem (normal-) mittelhochdeutschen, einem alemannischen, einem bairischen, einem mittelfränkischen und einem rheinfränkischen Lemmatisierungswörterbuch gewählt werden. U. U. ist die Anwendung von LEM-3 mit verschiedenen Lemmatisierungswörterbüchern zu wiederholen, bis das bestmögliche Lemmatisierungsergebnis erzielt ist.

<u>Programm:</u>	LEM-3, Optionen 1 bis 5 (je nach Schreibdialekt)
<u>Eingabe-Datei:</u>	Y-VL.1S
<u>Ausgabe-Datei:</u>	Y-VL.2

Die Datei Y-VL.2 besteht wieder aus zeilenförmigen Datensätzen, die mit dem Backslash \ und der Wortform beginnen; durch einen [TAB] getrennt folgt nach dem Unterstrich _ der Ordnungsteil, und nach einem weiteren [TAB] schließen sich die Lemmatisierungsangebote an. Die Lemmatisierungsangebote bestehen ihrerseits aus drei Feldern, die durch eine jeweils unterschiedliche Feldeingangsmarkierung eingeleitet werden; nach dem Trema " [ALT+249] folgt das Lemma, nach dem Gatter # die Kategorie, nach dem Zirkumflex ^ die grammatische(n) Bestimmung(en). Sofern einer Wortform mehrere Lemmata zugewiesen werden, werden diese in geschweiften Klammern {} durchgezählt.

Beispiel:

```

\<an> _0001.02 {1}"ane #präp ^ {2}"ane #adv ^
\dem _0003.01 {1}"dër #art ^DSm/DSn {2}"dër #pron ^DSm/DSn
\êre.(.) _0003.05@ {1}"êre #f ^NS/AS/GDS/NAP {2}"êren #swv ^1SG/3SGK/1PG/2SGB/i
\gemv\ete., _0002.03@ "ge-müète #n ^NS/AS/DS/NP/GP
\gv\ete. _0001.04@ ?"güèten #swv ^1SG/3SGK/1PG/
\rehte _0001.03 {1}"rëht(e) #adv ^ {2}"rëht #n ^DS {3}"rëht #adj ^NSmfnw
\$æelde _0003.03 {1}"sæl(e)de #f ^NS/AS/GDS/NAP {2}"sèl(e)de #f ^NS/AS/G
\$in _0002.02 {1}"sîn #pron poss ^0/- {2}"sîn #anv ^i/3PG/1PG
\{*S*}wer _0001.01 {1}"swër #pron ^NSm/GP/GSf/DSf {2}"swære #adj ^
\vn\ _0003.04 {1}"unde #konj ^ {2}"unden(e) #adv ^ {3}"von #präp ^
\volget _0003.02 "volgen #swv ^3SGI/2PG/2PGB/3SV/1SV
\wendet _0002.01 "wènden #swv ^3SGI/2PG/2PGB

```

Lemmatisierungsangebote, die aus dem Ergänzungswörterbuch stammen, werden durch ein der Lemmafeldmarkierung vorangestelltes Fragezeichen ? gekennzeichnet, weil sie mitunter

fehlerhaft, selten auch geradezu abstrus sind. Dieses Fragezeichen weist also darauf hin, die entsprechenden Lemmatisierungsangebote bei der Vorlemmatisierungskorrektur besonders gründlich zu prüfen; es veranlaßt außerdem die Überprüfung der entsprechenden Lemmaan-sätze durch das Prüf- und Korrekturprogramm LEM, Option 8.

Am Ende der Datei Y-VL-2 finden sich in absoluten Zahlen und in Prozentsätzen Angaben darüber, wie erfolgreich die automatische Lemmazuordnung war.

Beispiel:

* TREFFERZAHLEN

*

* Gesamtzahl:	12219
* Treffer im Lemmatisierungswörterbuch:	11862 (= 97.15%)
* Treffer im Ergänzungswörterbuch:	80 (= 0.66%)
* Treffer-Prozentsatz insgesamt:	97.81

Angestrebt werden sollte eine Trefferquote von wenigstens 95 %; liegt sie deutlich darunter, sollte zunächst die automatische Vorlemmatisierung mit anderen Einstellungen im Programm LEM-2 (Berücksichtigung anderer schreibsprachlicher Besonderheiten) und anderen Lemmatisierungswörterbüchern wiederholt werden. Wenn auf diese Weise ein hinreichend zufriedenstellendes Ergebnis erzielt wird, schließt sich das "Löcherstopfen" an (siehe 8.6), andernfalls sollte zunächst die Vorlemmatisierung durch eine Ergänzung des Lemmatisierungswörterbuchs verbessert werden (siehe 8.5).

8.5 Verbesserung der automatischen Vorlemmatisierung

Stellt man höherfrequente "Löcher" in einer Datei Y-VL.2 fest, so können die entsprechenden Einträge in das benutzte Lemmatisierungswörterbuch eingefügt werden. Damit werden sowohl bei einer anschließenden Neulemmatisierung des aktuell in Arbeit befindlichen Textes als auch bei weiteren Anwendungen desselben Lemmatisierungswörterbuchs verbesserte Resultat erzielt. Die Verbesserung des Lemmatisierungswörterbuchs soll aber nur von Mitarbeitern mit ausreichender Erfahrung und ggf. zusammen mit diesen vorgenommen werden.

Das Ergänzungsverfahren besteht aus folgenden Schritten:

1. Zunächst ist das Tastaturlayout 2 ("Ergänzung des Lemmatisierungswörterbuchs" zu aktivieren; Makro TASTATUR). Dann wird die Datei Y-VL.2 im Editor in Datei 2 geladen; Datei 1 muß leer sein (andernfalls Fehlermeldung des Makros LOCH-KOP). Anschließend werden mit dem Makro LOCH-KOP (aufzurufen durch [ALT+C]) alle "Löcher"-Zeilen nach Datei 1 kopiert. Die so erstellte Datei wird als Y-VL.2f gespeichert.
2. Für die folgenden Schritte muß sich die Datei Y-VL.2f in Datei 1 befinden (andernfalls Fehlermeldung des Makros LEM-ERG); Datei 2 muß frei sein für das Lemmaregister, Datei 3 für die Textdatei. Die Datei Y-VL.2f wird mit dem Makro LOECHER (aufzurufen durch [ALT+M]) nach Blöcken von wenigstens $n = 5$ identischen Lemmatisierungslöchern durchsucht. Die Zahl n kann mit dem Makro N-AENDER (über [ALT+G]) verändert werden.
3. Soll dem Lemmatisierungswörterbuch ein entsprechender Eintrag hinzugefügt werden, so startet man anschließend das Makro LEM-ERG (über [ALT+L]). Es erzeugt in einer der Vorlemmatisierungskorrektur ähnlichen Weise in Datei 4 den neuen Eintrag, dessen Format der Vorstufe des Lemmatisierungswörterbuchs LW-Z.Kn entspricht. Die so in Datei 4 entstehende / entstandene Ergänzungsdatei wird unter dem Namen Y-LW.ERG abgespeichert.⁷⁵
4. Die fertige Datei Y-LW.ERG wird anschließend mit SORTA alphabetisch sortiert und als Y-LW.K gespeichert.

⁷⁵ Muß die Erstellung von Y-LW.ERG unterbrochen werden, so sollte die aktuelle Stelle mit +++ markiert werden. Bei Wiederaufnahme der Arbeit muß Y-VL.2f wieder in Datei 1 und Y-LW.ERG in Datei 3 geladen werden.

5. Dann wird Y-LW.K mit dem Programm LW-ERG, Option 3, in LW-Z.Kn, die Vorstufe des entsprechenden Lemmatisierungswörterbuchs, inseriert.

<u>Programm:</u>	LW-ERG, Option 3
1. Eingabedatei:	LW-Z.Kn
2. Eingabedatei:	Y-LW.K
1. Ausgabedatei:	LW-Z.Kne
2. Ausgabedatei:	Y-LW.ERW

Das Programm LW-ERG, Option 3, überprüft während des Programmlaufs auch, ob Y-LW.K und LW-Z.Kn korrekt alphabetisch sortiert sind. Wenn nicht, bricht das Programm mit einer Fehlermeldung ab. Dies dient dazu, die streng alphabetische Sortierung von LW-Z in jedem Falle zu sichern, da sonst fatale Folgen bei der Vorlemmatisierung auftreten können.

6. Die Datei LW-Z.Kne wird in LW-Z.Kn umbenannt. Danach wird mit dem Programm LW-ERG, Option 4, aus LW-Z.Kn das neue Lemmatisierungswörterbuch LW-Z erzeugt.

7. Zuletzt muß die Datei Y-VL.1S mit dem nun ergänzten Lemmatisierungswörterbuch erneut mit dem Programm LEM-3 vorlemmatisiert werden (siehe oben unter 8.4).

8.6 Das “Löcherstopfen”

Alle Wortformen, deren automatische Lemmatisierung gescheitert ist, weil kein entsprechender Eintrag im Lemmatisierungswörterbuch vorhanden war, sind durch das Dollarzeichen \$ vor der Lemmafeldmarkierung ` gekennzeichnet; anstelle eines richtigen Lemmaansatzes findet sich die “normalisierte” Wortform, die vom Programm LEM-2, Option 2, erzeugt worden ist. In allen diesen Fällen müssen manuell ein Lemma und mögliche grammatische Bestimmungen zugewiesen werden (das “Löcherstopfen”).

Zunächst muß im WP-Editor das Tastaturlayout 1 (“Vorlemmatisierungskorrektur”) aktiviert werden. Die Datei Y-VL.2 wird in Datei 1 geladen; Datei 2 muß frei bleiben für das Lemmaregister, Datei 3 für die Textdatei.

Nun werden der Reihe nach alle “Lemmatisierungslöcher” aufgesucht, indem man mit [F2] entweder nach der Kombination \$` oder nach @[BLANK]^ sucht. Mit dem Makro LEM-ERG (über [ALT+Y]) werden dann die korrekten Lemmaansätze und ggf. mögliche grammatische Bestimmungen eingefügt. Ausdrücklich wird auf die Regelungen zur formal korrekten Gestaltung der Lemmaansätze und der Angaben zu Kategorien und grammatischen Bestimmungen (unter 6) hingewiesen. Neben dem Lemmaregister, das über [ALT+Y] aufgerufen wird, steht auch der “Lexer” über [ALT+W] zur Verfügung. Über [ALT+T] kann die zugehörige Textdatei aufgerufen werden, da die Lemmazuweisung oft nur aus dem jeweiligen Kontext heraus möglich ist. Für alle anderen Makros im Tastaturlayout 1 wird auf die Übersicht unter 19 hingewiesen.

Es ist darauf zu achten, daß das Dollarzeichen \$ vor dem eingefügten Lemma stehenbleibt oder ggf. manuell eingefügt wird, wenn ein automatisch erzeugter Vorlemmatisierungsansatz manuell verändert wird. Alle durch das Dollarzeichen \$ markierten Lemmaansätze werden später durch das Programm LEM, Option 8, auf ihre formale Korrektheit überprüft.

8.7 Die Erstellung des VK-Indexes

Nach Abschluß des “Löcherstopfens” wird Y-V2 mit dem Programm SORTA nach dem Ordungsteil sortiert.

<u>Programm:</u>	SORTA
<u>Eingabe-Datei:</u>	Y-V2.txt
<u>Ausgabe-Datei:</u>	Y-VZ.txt
<u>Feldbegrenzer:</u>	[TAB] (= [ENTER])
<u>Sortierreihenfolge:</u>	2

Damit ist wieder die ursprüngliche Textreihenfolge hergestellt.

Beispiel (im WP-Editor allerdings ohne Zeilenumbruch innerhalb eines Datensatzes):

```
\{*S*}wer _0001.01 {1}`swër @pron ^NSm/GP/GSf/DSf {2}`swære @adj ^_/0/NSmfnw/NASf/
ASnw/NAP {3}`swèr @m ^NS/AS/DS/NP/AP/GP {4}`swër @n ^NS/AS/DS/NP/AP/GP
{5}`swër(e) @m ^NS {6}`swër(e)n @stv4 ^3SGK/1SGK/3SVK/1SVK/2SVI/1PVK/i
{7}`swër(e)n @swv ^1SG/3SGK/1PG/2SGB/i {8}`swért @n ^NS/AS/NP/AP/DS {9}`swèr(e)n
@stv6b ^1SG/2SGB/3SGK {10}`swære @m ^NS/AS/DS/NP/AP/GP
\<an> _0001.02 {1}`ane @präp ^ {2}`ane @adv ^ {3}`âne @präp ^
\rehte _0001.03 {1}`rëht(e) @adv ^ {2}`rëht @n ^DS/GP {3}`rëht @adj ^NSmfnw/NASf/
ASnw/NAP {4}`rëhte @f ^NS/AS/GDS/NAP {5}`rëhte @m ^NS {6}`rëhte @n ^NS/AS/DS/
NP/AP/GP {7}`rëhten @swv ^1SG/3SGK/1PG/2SGB/i
\gv\ete. _0001.04@ {1}`güèten @swv ^1SG/3SGK/1PG/2SGB/i/3SV/1SV {2}`güète @f
^NS/AS/GS/DS/NP/AP {3}`wëten @stv5 ^3SGK/1SGK/3SVK/1SVK/2SVI/1PVK/i
\wendet _0002.01 `wènden @swv ^3SGI/2PG/2PGB
\sin _0002.02 {1}`sîn @pron poss ^0/_ {2}`sîn @anv ^i/3PG/1PG {3}`ër @pron ^GSm/GSn
{4}`sinn @m ^NS/AS/DS {5}`sint @konj ^ {6}`sîn @n ^NS/AS/NP/AP/DS
\gemv\ete.,, _0002.03@ {1}`ge_müèt @adj part ^NSmfnw/NASf/ASnw/NAP {2}`ge_müète @n
^NS/AS/DS/NP/AP/GP
\dem _0003.01 {1}`dër @art ^DSm/DSn {2}`dër @pron ^DSm/DSn
\volget _0003.02 `volgen @swv ^3SGI/2PG/2PGB/3SV/1SV
\$/ælde _0003.03 {1}`sæl(e)de @f ^NS/AS/GDS/NAP {2}`sèl(e)de @f ^NS/AS/GDS/NAP
\vn\ _0003.04 {1}`unde @konj ^ {2}`unden(e) @adv ^ {3}`von @präp ^
\ère(.) _0003.05@ {1}`ère @f ^NS/AS/GDS/NAP {2}`ëren @swv ^1SG/3SGK/1PG/2SGB/i {3}`ër
@m ^DS/NP/AP/GP {4}`ër(e)n @stv7 ^1SG/3SGK/1PG/i/2SGB
```

Zur Verbesserung von Y-VZ dient das Programm LEM-4, das die Reihenfolge der Lemmatisierungsmöglichkeiten in bestimmten Fällen gemäß der syntaktischen Umgebung ändert. Die Reichweite dieses Programms ist begrenzt, und es arbeitet auch nicht immer fehlerfrei. Die von LEM-4 veränderten Lemmatisierungsangebote sind daher durch ein Ausrufezeichen ! vor der Lemmafeldmarkierung `` und ` gekennzeichnet.

<u>Programm:</u>	LEM-4
<u>Eingabe-Datei:</u>	Y-VZ.txt
<u>Ausgabe-Datei:</u>	Y-V.txt

In der Ausgabedatei doppelt LEM-4 das Angebot: voran steht nun ein „Trema“-Bereich, in dem die Lemmata mit einem Trema `` (teils auch als • dargestellt) beginnen; durch einen Tab und % abgetrennt folgt der „Gravis“-Bereich, in dem die Lemmata mit einem Gravis ` beginnen. Der Gravis-Bereich ist zunächst ein genaues Spiegelbild des Trema-Bereichs, nur dass ihm die Flexionsangebote fehlen, z. B.

```
{1}`ge-mach #mn ^NS/AS/NP/AP/DS {2}`ge-mach #adj ^-/0 % {1}`ge-mach @mn ^ {2}`ge-
mach @adj ^
```

Während das Trema-Lemma und seine Wortart belegspezifisch angepasst werden können, dürfen das Gravis-Lemma und seine Wortart grundsätzlich nicht verändert werden! Das Gravis-Lemma kann lediglich zusammen mit dem Trema-Lemma durch ein anderes Lemma des Lemmaregisters ersetzt werden.

Beispiel einer belegspezifischen Anpassung des #-Genus' (Kontext: *daz gemacht*):

```
{1}`ge-mach #n ^NS/AS/NP/AP/DS {2}`ge-mach #adj ^-/0 % {1}`ge-mach @mn ^ {2}`ge-
mach @adj ^
```

Für die weitere, nun manuell-interaktive Bearbeitung der Vorlemmatisierung müssen die in aller Regel überlangen Zeilen der Datei Y-VL an die Breite des Bildschirms angepaßt werden.

Programm: LEM, Option 5
Eingabe-Datei: Y-VL
Ausgabe-Datei: Y-VK
Maximale Zeilenlänge: 75

Am linken Rand wird zudem ein Freiraum eingefügt, in den bei der Vorlemmatisierungskorrektur die manuellen Korrekturen geschrieben werden.

Beispiel:

```

\{*}*wer _0001.01 {1}“swër #pron ^NSm/GP/GSf/DSf +
{2}“swære #adj ^-/0/NSmfñw/NASf/ASñw/NAP +
{3}“swèr #m ^NS/AS/DS/NP/AP/GP +
{4}“swër #n ^NS/AS/DS/NP/AP/GP +
\<an> _0001.02 {1}“ane #präp ^ {2}“ane #adv ^ +
{3}“âne #präp ^
\rehte _0001.03 {1}“rëht(e) #adv ^ +
{2}“rëht #n ^DS/GP +
{3}“rëht #adj ^NSmfñw/NASf/ASñw/NAP +
{4}“rëhte #f ^NS/AS/GDS/NAP {5}“rëhte #m ^NS +
{6}“rëhte #n ^NS/AS/DS/NP/AP/GP +
{7}“rëhten #swv ^1SG/3SGK/1PG/2SGB/i
\gv\ete. _0001.04@ {1}“güèten #swv ^1SG/3SGK/1PG/ +
{2}“güète #f ^NS/AS/GS/DS/NP/AP
\wendet _0002.01 “wènden #swv ^3SGI/2PG/2PGB
\$in _0002.02 {1}“sîn #pron poss ^0/- +
{2}“sîn #anv ^i/3PG/1PG +
{3}“ër #pron ^GSm/GSn {4}“sinn #m ^NS/AS/DS
\gemv\ete., _0002.03@ {1}“ge-müèt #adj part ^NSmfñw/ +
{2}“ge-müète #n ^NS/AS/DS/NP/AP/GP
\dem _0003.01 {1}“dër #art ^DSm/DSn +
{2}“dër #pron ^DSm/DSn
\volget _0003.02 “volgen #swv ^3SGI/2PG/2PGB/3SV/1SV
\$ælde _0003.03 {1}“sæl(e)de #f ^NS/AS/GDS/NAP +
{2}“sèl(e)de #f ^NS/AS/GDS/NAP
\vn\ _0003.04 {1}“unde #konj ^ {2}“unden(e) #adv ^ +
{3}“von #präp ^
\êre.(.) _0003.05@ {1}“ère #f ^NS/AS/GDS/NAP +
{2}“èren #swv ^1SG/3SGK/1PG/2SGB/i
  
```

9 Die Vorlemmatisierungskorrektur

Bei der Vorlemmatisierungskorrektur geht es darum, aus den Lemmaansätzen und den möglichen grammatischen Bestimmungen, die mit Hilfe der Lemmatisierungswörterbücher den einzelnen handschriftlichen Wortformen zugeordnet worden sind (vgl. oben 8), das jeweils richtige Lemma und die richtige grammatische Bestimmung auszuwählen und auf diese Weise jede Wortform genau grammatisch zu bestimmen. Unvollständige und falsche Lemmaangebote sind manuell zu korrigieren.

Die Vorlemmatisierungskorrektur der Datei Y-VK erfolgt direkt am Bildschirm im WordPerfect-Editor. Wie das obige Beispiel (siehe) verdeutlicht, besteht die Datei Y-VK aus Datensätzen, die die handschriftliche Wortform, den Ordnungsteil und eine oder mehrere Lemmaangebote beinhalten. Jeder Datensatz wird durch einen [TAB] eröffnet; die handschriftliche Wortform ist durch den Backslash \ als Feldeingangsmarkierung gekennzeichnet, der Ordnungsteil durch den Unterstrich _, der Lemmaansatz durch das Trema “, die Kategorie durch das Gatter und die

grammatische(n) Bestimmung(en) durch den Zirkumflex. Vor dem Trema `` als Feldeingangs- markierung für das Lemma können das Dollarzeichen \$, das Fragezeichen ? oder bestimmte an- dere Zeichen stehen, die Aufschluß geben über das Eingreifen bestimmter Programme oder über manuelle Veränderungen; diese Zeichen sind für die Vorlemmatisierungskorrektur ohne Belang und brauchen nicht weiter beachtet zu werden; sie dürfen aber auch nicht entfernt wer- den.

Sofern einer Wortform mehrere Lemmata zugeordnet sind, werden diese durchgezählt; ist ei- ner handschriftlichen Wortform nur ein Lemma zugeordnet, entfällt die Numerierung. Der Zei- lenumbruch innerhalb eines Datensatzes wird durch das Pluszeichen + markiert; auch dieses Pluszeichen darf unter keinen Umständen entfernt werden.

Schematischer Aufbau eines Datensatzes der Datei Y-VK:

```
[TAB]\Wortform _Ordnungsteil {1} ``Lemma1 #Wortart ^Bestimmung1/2/3/4 + {2} ``Lemma2 #Wortart  
^Bestimmung1/2/3/4
```

Zur Vorlemmatisierungskorrektur muß zunächst im WP-Editor das Tastaturlayout 1 (``Vorlem- matisierungskorrektur``) aktiviert werden; dieses Tastaturlayout stellt eine Reihe von Makros zur Verfügung, die die Vorlemmatisierungskorrektur am Bildschirm erleichtern (siehe 19.2.1). Die Datei Y-VK wird in Datei 1 geladen; Datei 2 muß frei bleiben für das Lemmaregister, Datei 3 für die Textdatei. Wie beim ``Löcherstopfen`` kann auch bei der Vorlemmatisierungskorrektur das Lemmaregister über [ALT+Y] aufgerufen werden, der ``Lexer`` über [ALT+W] und die Textda- tei über [ALT+T].

Nun wird die Datei Y-VK von Anfang an in der Textreihenfolge durchgearbeitet. Dabei wird das Ergebnis der automatischen Vorlemmatisierung zu jeder einzelnen Wortform mit Sinn und Ver- stand überprüft; bei mehreren Lemmatisierungsangeboten wird die richtige Lemmatisierung angewählt, ist kein Lemmatisierungsangebot korrekt, erfolgt die Korrektur manuell.

Sofern nun das erste Lemmaangebot und gegebenenfalls die erste dazu angebotene grammati- sche Bestimmung zutrifft, ist keinerlei Korrektur notwendig. In allen anderen Fällen ist wie folgt vorzugehen:

1. Wenn ein anderes als das erste Lemmaangebot und / oder eine andere als die erste angebo- tene grammatische Bestimmung zutrifft, wird am linken Zeilenrand vor dem Tabulatorschritt die Nummer des zutreffenden Lemmaangebotes und danach, durch Komma, aber ohne Blank getrennt, die Nummer der korrekten grammatischen Bestimmung eingetragen.

Beispiel:

```
1,4 \si _005.01 {1} ``ër #pron ^NSf/ASf/NP/AP +  
{2} ``sîn #anv ^3SGK/1SGK/i
```

Durch 1,4 ist hier zu der Wortform si das erste Lemmaangebot ``ër #pron und dazu die vierte gram- matische Bestimmung AP ausgewählt worden.

Auch wenn zu einem Lemma die erste angebotene grammatische Bestimmung zutreffend ist, ist diese grammatische Bestimmung im Korrektureintrag eindeutig mit 1 zu bezeichnen, um Verwechslungen und damit mögliche Fehler zu vermeiden. Die Angabe der Nummer der ausge- wählten grammatischen Bestimmung entfällt nur in den Fällen, wenn es zu einem Lemmaange- bot nur eine oder gar keine grammatische Bestimmung gibt.

2. Wird zu einer Wortform nur ein Lemma mit mehreren möglichen grammatischen Bestim- mungen angeboten, so entfällt natürlich die Angabe der Nummer des Lemmas; es wird am lin- ken Rand nur die Nummer der zutreffenden grammatischen Bestimmung eingetragen.

Beispiel:

```
3 \tuginde _0016.02 ``tugend #f ^NS/AS/GS/DS
```

Ausgewählt ist die Bestimmung GS zum einzigen Lemmaangebot ``tugend #f.

Enthält die angebotene grammatische Bestimmung mehrere Kasus-, Numerus- oder Genusan- gaben, so ist der Nummer der grammatischen Bestimmung das zutreffende Kasus-, Numerus- oder Genuskürzel hinzuzufügen.

Beispiele:

2,1f \rechte _107.04@ {1} ``r ht #n ^NS/AS/NP/AP/DS +
{2} ``r ht #adj ^NSmfnw/NASf/ASnw/NAP

2,4N \rechte _107.04@ {1} ``r ht #n ^NS/AS/NP/AP/DS +
{2} ``r ht #adj ^NSmfnw/NASf/ASnw/NAP

Durch 2,1f bzw. 2,4N ist zum zweiten Lemma ``r ht #adj die erste grammatische Bestimmung NSfw bzw. die vierte Bestimmung NP ausgew hlt worden.

3. Wird zu einer Wortform zwar das richtige Lemma, aber nicht auch die richtige grammatische Bestimmung angeboten, so wird nach der Nummer des Lemmas und davon durch ein Komma getrennt die gesamte richtige Bestimmung eingetragen. Gibt es nur ein Lemmaangebot, wird direkt die richtige grammatische Bestimmung angegeben.

Beispiele:

2,AP \rechte _107.04@ {1} ``r ht #n ^NS/AS/NP/AP/DS +
{2} ``r ht #adj ^NSmfnw/ASnw

Anstelle der angebotenen Bestimmungen wird der Wortform *rechte* die Bestimmung AP zugeordnet
NP \tuginde_0016.02 ``tugend #f ^NS/AS/GS/DS

Anstelle der angebotenen Bestimmungen wird der Wortform *tuginde* die Bestimmung NP zugeordnet

Ebenso kann man verfahren, um aus einer Sammelangabe wie NAP eine genaue Bestimmung auszuw hlen.

Beispiel:

2,NP \rechte _107.04@ {1} ``r ht #n ^NS/AS/NP/AP/DS +
{2} ``r ht #adj ^NSmfnw/NASf/ASnw/NAP

4. Wenn kein Lemmaangebot zutrifft, mu  der gesamte Eintrag neu erstellt werden. Dazu wird das zutreffende Lemma unter Anwendung der entsprechenden Makros im Lemmaregister gesucht und von dort in die Vorlemmatisierungsdatei eingef gt oder – wenn das Lemma noch nicht im Lemmaregister enthalten ist – entsprechend den oben beschriebenen Konventionen neu angesetzt. Erg nzt wird der Lemmaansatz um die grammatische Bestimmung. Es ist darauf zu achten, da  der Vorlemmatisierungseintrag den oben beschriebenen Konventionen genau entspricht (vgl. oben 6). Alle neu geschriebenen oder manuell ver nderten Eintr ge m ssen durch das Dollarzeichen \$ vor der Lemmafeldmarkierung `` gekennzeichnet sein, um die  berpr fung durch das Korrekturprogramm LEM, Option 8, zu veranlassen. Sofern das Lemma aus dem Lemmaregister  bernommen wird, wird das Dollarzeichen automatisch eingef gt.

5. Wird bei der Vorlemmatisierungskorrektur ein Fehler in der Interpunktion festgestellt, so mu  die betreffende Stelle in der Textdatei und in der Datei Y-VK korrigiert werden. Die Korrektur allein in der Datei Y-VK ist nicht ausreichend, weil andernfalls diese  nderung bei einem sp teren Abgleich der Textdatei mit dem Index verloren gehen w rde.

6. Wenn bei der Vorlemmatisierung ein Pr editierungsfehler festgestellt wird, der die Wortheiten betrifft, m ssen ebenfalls sowohl die Textdatei als auch die Datei Y-VK korrigiert werden.
(1) Wenn zwei oder mehr W rter zu einem Wort zusammengezogen werden, muss zun chst die handschriftliche Wortform einschlie lich der Pr editierungszeichen manuell korrigiert werden; dann werden die nun  berfl ssigen Zeilen des Index gel scht, und zuletzt ist die Konsistenz der Zeilen- und Wortz hlung zu  berpr fen. In aller Regel f llt nur eine Korrektur der Wortz hlung an; sie sollte mit dem Makro Alt+Z (=WZAHN-KO.EDM) vorgenommen werden, das die Wortformen zwischen der vorangehenden und der folgenden Zeilenende-Markierung @ neu durchz hlt.

Beispiel:

\er _27,27.01 {1} `` r #pron ^NSm/GP/GSf/DSf {2} ``d r #art ^NSm/GP/GSf/DSf ...
\unde _27,27.02 {1} ``unte #konj ^ {2} ``unte #adv ^ {3} `` nde #f ^NS/AS/GDS/NAP ...
\e\$au. _27,27.03 ``Esau #PN ^NS/DS/AS %`Esau @PN ^NS/DS/AS
\ne _27,27.04 ``ne #partikel ^ %`ne @partikel ^

```

\waren _27,27.05 {1}`wësen #stv5 ^3PVI/1PVI/3PVK/1PVK {2}`wâr #adj ^DP/ASm/
ASfw/...
\ein _27,27.06 {1}`ëin @pron ^0 {2}`ëin @art ^0 {3}`ëin @num ^0/- {4}`ëine @adv
^
\ander _27,27.07 {1}`ander @pron ^-/0/NSmfnw/NASf/ASnw/NAP {2}`ander(e)n
@swv ...
\niht _27,27.08@ {1}`niht #adv ^ {2}`niht #n ^NS/AS/DS % {1}`niht @adv ^ ...
\getr%oiv.(.) _27,28.01 ``ge-triuwe #adj ^-/0/NSmfnw/NASf/ASnw/NAP %`ge-triuwe
@adj ^- ...

```

korrigiert zu:

```

\er _27,27.01 {1}`ër #pron ^NSm/GP/GSf/DSf {2}`dër #art ^NSm/GP/GSf/DSf ...
\unde _27,27.02 {1}`unte #konj ^ {2}`unte #adv ^ {3}`ünde #f ^NS/AS/GDS/NAP ...
\esau. _27,27.03 ``Esau #PN ^NS/DS/AS %`Esau @PN ^NS/DS/AS
\ne _27,27.04 ``ne #partikel ^ %`ne @partikel ^
\waren _27,27.05 {1}`wësen #stv5 ^3PVI/1PVI/3PVK/1PVK {2}`wâr #adj ^DP/ASm/
ASfw/...
\ein#ander _27,27.06 ``ëin-ander #pron ^0/- %`ëin-ander @pron ^0/-
\niht _27,27.07@ {1}`niht #adv ^ {2}`niht #n ^NS/AS/DS % {1}`niht @adv ^ ...
\getr%oiv.(.) _27,28.01 ``ge-triuwe #adj ^-/0/NSmfnw/NASf/ASnw/NAP %`ge-triuwe
@adj ^- ...

```

(2) Wenn umgekehrt ein Wort in zwei oder mehrere Wörter getrennt wird, müssen entsprechend die notwendigen Zeilen in den Index eingefügt werden; dann müssen die handschriftlichen Wortformen samt Präeditierungszeichen korrigiert werden, und zuletzt ist die Zählung auf ihre Stimmigkeit zu überprüfen. Besondere Beachtung ist hier dem Klammeraffen @ als Markierung des letzten Wortes der Zeile oder des Verses zu schenken. Die Überprüfung bzw. Korrektur sollte wieder durch das Makro ALT+Z erfolgen.

Beispiel:

```

\riten _31,27.01 {1}`rîten #stv1a ^i/3PG/1PG/3PV/1PV/1SGI {2}`râten #stv7 ^3PV/
1PV ...
\unde _31,27.02 {1}`unte #konj ^ {2}`unte #adv ^ {3}`ünde #f ^NS/AS/GDS/NAP ...
\gen.,, _31,27.03 {1}`gân #anv ^i/3PG/1PG/1SGI {2}`gègen #präp ^ % {1}`gân @anv
...
\wi _31,27.04 {1}`wiè #adv_int ^ {2}`wê #interj ^ {3}`wî #interj ^ ...
\Shonez _31,27.05 ``schoène #adj ^NSn/ASn
\Stat _31,27.06 {1}`stân #anv ^3SGI/2PG/2PGB {2}`stat #f<u> ^NS/AS/GS/DS ...
\ge=malot.(.) _31,27.07@ ``mâlen #swv ^pV-/pV0 %`mâlen @swv ^pV-/pV0

```

korrigiert zu:

```

\riten _31,27.01 {1}`rîten #stv1a ^i/3PG/1PG/3PV/1PV/1SGI {2}`râten #stv7 ^3PV/
1PV ...
\unde _31,27.02 {1}`unte #konj ^ {2}`unte #adv ^ {3}`ünde #f ^NS/AS/GDS/NAP ...
\gen.,, _31,27.03 {1}`gân #anv ^i/3PG/1PG/1SGI {2}`gègen #präp ^ % {1}`gân @anv
...
\wi _31,27.04 {1}`wiè #adv_int ^ {2}`wê #interj ^ {3}`wî #interj ^ ...
\Shon| _31,27.05 ``schoène #adj ^-/0/NSmfnw/NASf/ASnw/NAP
\|ez _31,27.06 ``ër #pron ^NSn/ASn/GSm/GSn {2}`dër #art ^NSn/ASn ...
\Stat _31,27.07 {1}`stân #anv ^3SGI/2PG/2PGB {2}`stat #f<u> ^NS/AS/GS/DS ...
\ge=malot.(.) _31,27.08@ ``mâlen #swv ^pV-/pV0 %`mâlen @swv ^pV-/pV0

```

Alle wesentlichen Änderungen in der Textdatei müssen im Header im Feld "Fehler / Änderungen: bereits korrigiert / geändert:" dokumentiert werden, um auf diese Weise auch die anderen Arbeitsstellen und spätere Bearbeiter darüber genau zu informieren. Dabei sollten neben

der Stellenangabe sowohl die ursprüngliche, fehlerhafte als auch die neue, korrigierte Form vermerkt werden. Die Dokumentation der Änderungen wird vereinfacht durch das Makro ALTK, mit dem die gesamte Zeile, in der die Änderung vorgenommen wird, in den Header kopiert werden kann; aus Gründen der Übersichtlichkeit sollte die kopierte Zeile dann nach Möglichkeit manuell auf das unbedingt Notwendige reduziert werden.

10 Von der Vorlemmatisierung zum Normalformindex

Nach Abschluß der Vorlemmatisierungskorrektur werden die Korrektureinträge in der Datei Y-VK automatisch ausgewertet, so daß dann jeder Wortform nur noch genau eine grammatische Bestimmung zugeordnet ist. Nach gewissen Überprüfungen kann dann der Normalformindex Y-N0 erzeugt werden, in dem jede Wortform und grammatische Bestimmung durch eine automatisch erzeugte, weitgehend im Sinne des "Normalmittelhochdeutschen" normalisierte Form ergänzt wird.

10.1 Auswertung der Korrektureinträge

Auf die Datei Y-VK wird nach Abschluß der manuellen Vorlemmatisierungskorrektur das Programm LEM, Option 6, angewendet; es wertet die Korrektureinträge aus und übernimmt in die Ausgabedatei nur die jeweils angewählten Lemmatisierungen.

<u>Programm:</u>	LEM, Option 6
<u>Eingabe-Datei:</u>	Y-VK
<u>Ausgabe-Datei:</u>	Y-L0

Die Ausgabedatei besteht aus zeilenförmigen Datensätzen mit fünf Feldern, die jeweils durch einen [TAB] voneinander getrennt sind und mit einer Feldeingangsmarkierung beginnen. Der Datensatz beginnt mit dem Lemmaansatz, der durch das Trema `` markiert wird; danach folgen die Angaben zur Kategorie (nach dem Gatter #), die grammatische Bestimmung (nach Zirkumflex ^), die handschriftliche Wortform (nach Backslash \) und der Ordnungsteil (nach Unterstrich _).

Beispiel:

```
``swër #pron ^NSm \{*S*}wer _0001.01
``ane #präp ^\<an> _0001.02
?``rëht #adj ^ASf \rehte _0001.03
?``güète #f ^AS \gv\ete. _0001.04@
``wènden #swv ^3SGI \wendet _0002.01
``sîn #pron poss ^0 \$in _0002.02
``ge-müète #n ^NS \gemv\ete.,, _0002.03@
``dër #pron ^DSm \dem _0003.01
``volgen #swv ^3SGI \volget _0003.02
``sæl(e)de #f ^NS \$ælde _0003.03
``unde #konj ^\vn\ _0003.04
``êre #f ^NS \êre.(.) _0003.05@
```

10.2 Markierung der Pronominaladverbien

In einem interaktiven Arbeitsschritt werden nun die Pronominaladverbien markiert. Das Programm LEM, Option 7, sucht in der Datei Y-L0 nach Elementarsätzen mit wenigstens zwei Adverbien wie dâr, nâch, zuo und schlägt vor, sie als zusammengehörig (*dâr-zuo* etc.) zu markieren. Sofern der Vorschlag vom Nutzer interaktiv bestätigt wird, werden die Lemmaansätze entsprechend ergänzt.

<u>Programm:</u>	LEM, Option 7
------------------	---------------

Eingabe-Datei: Y-L0
Text-Datei: X.TXT
Ausgabe-Datei 1: Y-LD
Ausgabe-Datei 2: Y-DAR

Der jeweils ersten Konstituente wird nach der Zeichenfolge Slash, Punkt, Pluszeichen /.+ die zweite Konstituente angefügt; der jeweils zweiten Konstituente wird die erste Konstituente direkt nach dem Slash angefügt, und der Punkt und das Pluszeichen .+ folgen am Ende.

Beispiel:

"dâr/.+ane	#adv	^	\da	_0027.01
"kêren	#swv	^3SV	\chert	_027.02
"ër	#pron	^NSm	\er	_0027.03
"sîn	#pron poss	^Asm	\\$inen	_0027.04
"vlîz	#m	^AS	\<fl>iz	_0027.05
"ane/dâr.+	#adv	^	\an.(.)	_0027.06@

Sofern die beiden Konstituenten in Kontaktstellung stehen, entfällt der Punkt vor dem Pluszeichen:

Beispiel:

"dâr/+ane	#adv	^	\da	_0027.01
"ane/dâr+	#adv	^	\an.(.)	_0027.02

10.3 Markierung der Partikelverben

In ähnlicher Weise wie die Pronominaladverbien werden auch die Partikelverben markiert. Auch dieser Arbeitsschritt erfolgt interaktiv, und zwar mit dem Programm LEM, Option 4:⁷⁶

Programm: LEM, Option 4
Eingabe-Datei: Y-LD
Ausgabe-Datei 1: Y-LP
Ausgabe-Datei 2: Y-PV

Nach Markierung der Pronominaladverbien sind nahezu alle verbleibenden Präpositionaladverbien (d.h. Adverbien, mit einer Präp. homonym sind, z. B. *ane*, *umbe*) Verbpartikeln. Das Programm sucht nach ihnen (gleich ob sie durch *1 markiert sind oder nicht) und schlägt die Verbformen in ihrer Umgebung als Partikelverbkandidaten vor. Nutzerseitig muss dann das zugehörige Verb bestimmt werden.

Das Programm markierte dementsprechend die Partikelverbkonstituenten je nach Stellung der Partikel in folgender Weise:

(1) die Partikel steht unmittelbar vor dem Verb:

⁷⁶ Alternativ könnte die Markierung der Partikelverben wie bisher im WP-Editor mit Hilfe des Makros PART-V erfolgen: Das Makro sucht nach Adverbien, die durch *1 markiert sind; interaktiv muss dann der Bearbeiter entscheiden, ob das jeweilige Adverb wirklich Bestandteil eines Partikelverbs ist und ob das zugehörige, durch *2 markierte Verb in der Datei weiter oben oder weiter unten steht. Dann erzeugt das Makro automatisch die entsprechenden Lemmazusätze: Dem Lemmaansatz des Verbs wird der Slash, das Lemma der Partikel und das Pluszeichen hinzugefügt, dem Lemmaansatz der Partikel der Slash, direkt das Pluszeichen und dann das Lemma des Verbs.

Beispiel:

"dër	#pron	^NSm	~dër	\der	_0106.01
"springen/ûf+	#stv3a	^3SVI	~sprang	\\$pranch*2	_0106.02
"en-gègen #präp	^	~en-gègen	\engegen		_0106.03
"ër	#pron	^DSf	~ir<e>	\ir	_0106.04
"ûf/+springen	#adv	^	~ûf	\vf*1	_0106.05
"ze-hant	#adv	^	~ze-hant	\zehant.,	_0106.06@

Nach Erzeugung der Lemmazusätze startet das Makro erneut durch [ENTER].

```
`ane @adv      ``ane/>+bëten #adv ^      \ane|*1
`bëten @swv    ``bëten/ane>+ #swv i ^-    \|beten*2
```

(2) die Partikel steht unmittelbar hinter dem Verb:

```
`rëcken @swv   ``rëcken/ûf<+ #swv ^3PVI \racten*2
`ûf @adv      ``ûf/<+rëcken #adv ^      \ûf*1
```

(3) die Partikel steht in Fernstellung hinter dem Verb:

```
`sprëchen @stv4 ``sprëchen/zuo<.+ #stv4 ^3SVI \sprach*2
```

....

```
`zuo @adv      ``zuo/<.+sprëchen #adv ^      \zuo*1(:)
```

10.4 Formale Überprüfung

Alle Lemmaansätze, die durch das Dollarzeichen oder das Fragezeichen vor der Lemmafeldmarkierung `` gekennzeichnet sind, werden vom Programm LEM, Option 8, auf formale Korrektheit überprüft. Das Programm schlägt ggf. Korrekturen vor, die angenommen, ausgewählt oder abgelehnt werden können. Zweifelsfälle können für eine Überprüfung im Kontext markiert und im Anschluß an die interaktive Bearbeitung manuell korrigiert werden.

<u>Programm:</u>	LEM, Option 8
<u>Eingabe-Datei:</u>	Y-LP
<u>Ausgabe-Datei:</u>	Y-L

10.5 Überprüfung der Zählung

Spätestens an dieser Stelle unmittelbar vor der Erzeugung des Normalformindex muß die Datei Y-LEM auf Vollständigkeit bzw. auf mögliche Störungen und Fehler in der Zeilen- bzw. Vers- und Wortzählung überprüft werden. Diese Überprüfung erfolgt mit dem Programm ZAHL, Option 5 (‘‘Ordnungsteil eines vertikal sortierten Index’’).

<u>Programm:</u>	ZAHL, Option 5
<u>Eingabe-Datei:</u>	Y-L
<u>Ausgabe-Datei:</u>	Y-LF

Störungen in der Zählung werden in der Ausgabedatei Y-LF dokumentiert; alle dort angegebenen Stellen müssen in der Datei Y-L aufgesucht und ggf. korrigiert werden.

10.6 Erzeugung des Normalformindexes

Das Programm NFORM erzeugt aus den Lemmaansätzen und den grammatischen Bestimmungen in der Datei Y-L zu jeder handschriftlichen Wortform eine in Anlehnung an das ‘‘Normalmittelhochdeutsche’’ normalisierte Wortform. Diese normalisierten Formen sind die Grundlage für die spätere graphonematische Analyse. Bei der Erzeugung der normalisierten Formen muß der Bearbeiter bei Fällen von Kontraktion, Synkope und Apokope teilweise interaktiv über deren Gestaltung entscheiden.⁷⁷

⁷⁷ Das Modul zur Berücksichtigung von Synkope und Apokope in den normalisierten Formen wurde dem Programm NFORM erst Anfang 2002 inseriert. Ältere Normalformindices müssen daher mit dem Programm NF-SYNK bearbeitet werden. Dieses Programm sucht nach Fällen, in denen entweder Schwa synkopiert ist, diese Synkope aber in der Normalform noch nicht berücksichtigt ist, oder aber eine Normalform mit Schwa in Spitzwinkelklammern fehlerhaft ist. In diesen Fällen schlägt NF-SYNK eine korrigierte Ersatzform vor, die ggf. manuell weiter korrigiert werden kann.

<u>Programm:</u>	NF-SYNK
<u>Eingabe-Datei:</u>	Y-NF.0
<u>Ausgabe-Datei:</u>	Y-NF.On

Die Ausgabedatei Y-NF.On ist anschließend wieder in Y-NF.0 umzubenennen. – Mittlerweile sind fast alle älteren Normalformindices mit dem Programm NF-SYNK bearbeitet.

Programm: NFORM

Eingabe-Datei: Y-L

Ausgabe-Datei: Y-NO

Gegenüber der Eingabedatei Y-L enthält der Normalformindex drei zusätzliche Felder: Zwischen der grammatischen Bestimmung und der handschriftlichen Wortform ist die normalisierte Form eingefügt (mit der Tilde ~ als Feldeingangsmarkierung); nach dem Ordnungsteil folgen – jeweils durch einen [TAB] getrennt – das Statusfeld und das Kommentarfeld.

Zur **Gestaltung der normalisierten Formen** im einzelnen siehe 11.

Das **Statusfeld** wird eingeleitet durch den Gradkringel °. Der ursprüngliche Zweck des Statusfeldes⁷⁸ hat sich erübrigt, weil 1. die abschließende NF-Index-Korrektur inzwischen bei jedem Text erfolgt ist, 2. die Maßnahme der Schlussüberprüfung schon vor längerer Zeit aufgegeben wurde, und 3. die Auszeichnung mit * nicht nur der alten Regel gemäß verwendet wurde und bei der Auswertung auch nicht berücksichtigt wird.

Das **Kommentarfeld** wird eingeleitet durch einen Doppelpunkt; in dieses Kommentarfeld werden mit dem Programm IND-KLAM alle Kommentare aus den anderen Feldern des Datensatzes verschoben.

Programm: IND-KLAM

Eingabe-Datei: IND-LIST.X

Das Programm IND-KLAM kann mehrere Indices auf einmal bearbeiten; daher kann nicht direkt der Pfad und Dateiname des zu bearbeitenden Index' angegeben werden, sondern nur eine Datei mit einer Liste von Indices. Diese Index-Liste IND-LIST muß im ASCII-Format gespeichert sein; sie muß die Eingabedatei(en) mit genauem Pfad und nach einem [TAB] die Ausgabedatei(en) mit genauem Pfad nennen. – Nach Anwendung des Programms IND-KLAM muß die Ausgabedatei Y-NF.0n unbedingt wieder in Y-NF.0 umbenannt werden!

Die Kommentare im Kommentarfeld werden jeweils eingeleitet durch einen Slash sowie die Feldeingangsmarkierung des Feldes, auf das sich der Kommentar bezieht (z. B. /~ zur Markierung eines Kommentars zum Lemmaansatz).

10.7 Überprüfung des Normalformindexes

Nach der Erstellung des Normalformindex sollte darauf das Programm PRUEF-NF angewandt werden, das in gewissem Rahmen in der Lage ist, Fehler in den grammatischen Bestimmungen zu ermitteln, v. a. Inkongruenzen verschiedener Art. Aufgrund der Komplexität von Sprache wird das Programm immer kompliziertere syntaktische Strukturen als vermeintlich inkongruent monieren. Daher kann das Programm Fehler nicht automatisch korrigieren, sondern der Benutzer muß interaktiv entscheiden, ob überhaupt ein Fehler vorliegt, und hat dann Gelegenheit, die Lemmatisierung manuell zu korrigieren.

Programm: PRUEF-NF

Eingabe-Datei: Y-NO

Ausgabe-Datei: Y-NOn

Sofern eine Korrektur ad hoc nicht möglich ist, etwa weil dazu der größere Kontext eingesehen werden muß, kann die fragliche Stelle markiert werden. Automatisch markiert werden zudem auch sämtliche Stellen, an denen eine Korrektur erfolgt ist.

Nach Abschluß der Überprüfung des Normalformindex mit dem Programm PRUEF-NF muß daher die Ausgabedatei Y-NOn im Editor geladen und mit dem Makro PRUEF-NF weiterbearbeitet

⁷⁸ Früher galt: Das Statusfeld besteht aus zwei Ziffern und ggf. dem Asterisk *. Die beiden Ziffern geben Auskunft über den Bearbeitungsstand des Datensatzes: Die erste Ziffer bezieht sich auf die handschriftliche Wortform und gibt an, ob die Textdatei bereits abschließend überprüft ist (1) oder noch nicht (0); die zweite Ziffer gibt an, ob der NF-Index bereits abschließend durchkorrigiert ist (1) oder noch nicht (0). Der Asterisk kennzeichnet alle Wortformen, die die Gesamtanzahl von 12.000 bis 12.500 Wortformen pro Korpustext überschreiben (vgl. 2).

werden: Dieses Makro sucht sämtliche Markierungen auf und ermöglicht dann durch eine Makropause dem Bearbeiter, die automatisch vorgenommenen oder noch ausstehenden Korrekturen abschließend zu bearbeiten. Insbesondere müssen in diesem Arbeitsschritt die Normalformen an die ggf. veränderten grammatischen Bestimmungen angeglichen werden. Nach Beendigung auch dieses Arbeitsschrittes muß die Datei Y-N0n wieder in Y-N0 umbenannt werden.

10.8 Abgleich des Normalformindexes mit der Textdatei

Um sicherzustellen, daß die maßgebliche Textdatei und der Normalformindex hinsichtlich der handschriftlichen Wortformen und des Ordnungsteil exakt übereinstimmen, ist nach Abschluß der Indizierung ein Abgleich dieser beiden Dateien notwendig. Zunächst muß aus der Textdatei ein neuer WOK-Index erzeugt werden (siehe 8.1). Standardmäßig wird bei der Erzeugung des WOK-Index' der Ordnungsteil auf die Zählung im engeren Sinne reduziert; für den folgenden Abgleich aber ist der **vollständige** Ordnungsteil notwendig, der daher ggf. wiederhergestellt werden muß.

Der eigentliche Abgleich erfolgt dann durch das Programm NFO-WOK, das unter Orientierung am Ordnungsteil und der Wortzählung Wortform für Wortform vergleicht. Wenn eine Differenz registriert wird, schreibt das Programm die betreffende Wortform aus der Textdatei bzw. dem WOK-Index neben die entsprechende Wortform in den Normalformindex und markiert die betreffende Stelle durch ALT+178. Ist eine Einfügung wegen Differenzen im Ordnungsteil nicht möglich, wird nur die betreffende Zeile durch den Hinweis "nicht ersetzt" markiert.

<u>Programm:</u>	NFO-WOK
<u>Eingabe-Datei 1:</u>	Y-N0
<u>Eingabe-Datei 2:</u>	Y-W
<u>Ausgabe-Datei:</u>	Y-N0p

Nach der Anwendung von NFO-WOK müssen in der Ausgabedatei alle Stellen überprüft werden, an denen das Programm eine Diskrepanz zwischen Textdatei und Normalformindex festgestellt hat oder an denen eine Ersetzung gar nicht möglich war. Dazu wird die Ausgabedatei Y-NF.Op im WordPerfect-Editor geladen, und dann werden alle Stellen, die mit ALT+178 oder dem Hinweis "nicht ersetzt" markiert sind, aufgesucht und manuell bearbeitet. Zunächst muß festgestellt werden, welche der beiden Wortformen, die der Textdatei oder die des Normalformindex' korrekt ist; die jeweils andere ist dann entsprechend zu ändern. Sofern die Wortform im Normalformindex geändert wird, kann dies Änderungen des Lemmaansatzes, der grammatischen Bestimmung oder der Normalform nach sich ziehen. Abschließend ist die Ausgabedatei Y-NF.Op wieder in Y-NF.0 umzubenennen.

Achtung: Wenn das Programm Differenzen nur in der Interpunktion registriert, wird ungefragt die Interpunktion des WOK-Index' übernommen und die Interpunktion des Normalformindex gelöscht bzw. überschrieben. Vor Anwendung des Programms ist also sicherzustellen, daß die Textdatei hinsichtlich der Interpunktion auf dem aktuellen Stand ist.

11 Die Normalform

Das Programm NFORM erzeugt anhand des Lemmaansatzes und der grammatischen Bestimmung zu jeder Wortform eines Indexes eine in Anlehnung an das "Normalmittelhochdeutsche" normalisierte Wortform als Voraussetzung für die weiteren Auswertungsschritte. Bei der Gestaltung dieser Normalformen sind gegenüber dem "Normalmittelhochdeutschen" in Einzelfällen Abweichungen zu beachten.

Bei der Gestaltung der Normalform bestehen oft zwei grundsätzlich verschiedene Möglichkeiten; dazu sei folgendes Beispiel gegeben:

a ~kir<e>chén \chilchún /l=r/ sagt etwas über die Lautung.

Zuordnung Graph = Referenzallophon.

tb ~kil<e>chén \chilchún /l=l/ sagt etwas über die Verschriftlichung.

Zuordnung Graph = Phon.

Handelt es sich um einzelne Laute wie im Beispiel, ist Lösung *a* (ohne Angleichung) vorzuziehen. Handelt es sich um ganze Formen, die variieren – zum Beispiel *oder* – *ald(e)* u. a. – muss eine Einzelfallentscheidung getroffen werden. Zu beachten sind insbesondere mfrk. Fälle, in denen eine Angleichung an mfrk. Sonderformen gewollt ist. Diese Angleichung ist in den Normalformfeldern bisher noch nicht vollständig durchgeführt (etwa *van* gg. *von*).

11.1 Synkope, Apokope, Elision und Kontraktion

In der Normalform werden drei verschiedene Typen von Klammern benutzt, um damit bestimmte sprachliche Erscheinungen oder Differenzen zwischen der Normalform und der handschriftlichen Wortform zu kennzeichnen.

Mit **runden Klammern** werden im Lemmaansatz und in der Normalform alle Fälle von möglicher, aber in der handschriftlichen Wortform nicht realisierter Apokope oder Synkope markiert.⁷⁹ Es ist darauf zu achten, daß nicht sämtliche Rundklammerungen aus dem Lemmaansatz übernommen werden, sondern nur dann, wenn auch in der konkret vorliegenden grammatischen Form die Möglichkeit zu Synkope oder Apokope besteht.

Apokope oder Synkope gelten als “möglich” nach Kurzvokal + Liquid, danach also werden *e* und *é* rund geklammert, wenn sie nicht spitz geklammert werden.

Durch **Spitzwinkelklammern** werden in der Normalform alle jene Grapheme eingefasst, die keine Entsprechung in der handschriftlichen Wortform haben; dabei kann es sich um Fälle von Synkope und Apokope, um graphische Besonderheiten jedweder Art und auch um fehlerhafte Schreibungen handeln.

Durch **geschweifte Klammern** werden Kontraktionen gekennzeichnet, und zwar in der Weise, daß das Kontraktionsergebnisse außerhalb der Klammern, die kontrahierten Vokale und Konsonanten in den Klammern stehen, z. B. ge-sèi{age}t, ge-lèi{ge}t, lì{ige}t. Die Schweifklammerung wird bei Formen von “lâzen aber weggelassen.

Wird von zwei im Lemmaansatz aufeinanderfolgenden Lauten nur einer in der Handschrift geschrieben, so wird in der Normalform im Zweifelsfall der erste geklammert. Der Zweifelsfall liegt vor, wenn es sich um zwei gleiche oder durch Assimilation angegliche Laute handelt, z. B. vrèvel-lich ~vrève<l>-lich, maged-tuom ~mage<d>-tuom.

Wenn in der hsl. Wortform ausl. Schwa steht, in der Nform aber nicht, wird nach folgendem Kriterium über die Ausnahme in die Nform entschieden:

- Lässt sich finales Schwa (auch) als lautgesetzlich entstandenes erklären, dann wird Schwa in die Nform aufgenommen; z. B. *e* bei prädikativen Adjektiven, die **ja*-stämmig gebildet sein können: ~schoèn<e>.
- Lässt sich das finale Schwa nicht als lautgesetzlich entstandenes erklären, sondern beruht wahrscheinlich auf Analogie, dann wird es nicht in die Nform aufgenommen. Dies ist z. B. der Fall, wenn Substantive der fem. *i/u*Dekl. im N/AS oder starke Verben in 3SVI ein *e* aufweisen: ^NS ~zît \zíte; ^3SVI ~kom \kome.

11.2 Unorganische Vokale und Konsonanten

Früher galt:

⁷⁹ Zu runden Klammern im Lemmaansatz vgl. .

Unorganische Vokale (anaptyktische) und Konsonanten (epenthetische oder hiattilgende), die nicht im Lemmaansatz, aber in der hsl. Form stehen, werden in der Normalform grundsätzlich nicht eingefügt. Das betrifft:

- *g* als Hiattilger (z. B. in obliquen Kasus von *zwî* oder *drî*).
- *j* als Hiattilger
- *w* als Hiattilger (z. B. in Formen von *kuo* f. ‘Kuh’ oder in Verba pura wie *sæjen* / *sæwen*)
- epenthetisches *d* wie in *\inder* ‘ièner
- epenthetisches *t* wie in *\ellent* ‘èllen
- anaptyktisches *e*

Inzwischen gelten subtilere Verhältnisse, die hier nicht zu beschreiben sind.

11.3 Auslautverhärtung

Die Auslautverhärtung /d/ > /t/ wird in den Normalformen grundsätzlich durchgeführt, die Auslautverhärtungen /b/ > /p/ und /g/ > /c/ dagegen nicht. Diese Vorgehensweise ist darin begründet, daß die Auslautverhärtung /d/ > /t/ in allen Dialekten in aller Regel durchgeführt ist, während die Reflexe von auslautend /b/ und /g/ insbesondere im Mitteldeutschen sehr unterschiedlich sein können. Damit über die Normalform problemlos auf alle einschlägigen Belege zugegriffen werden kann, müssen daher auslautende /b/ und /g/ unverändert bleiben.

Anmerkung: Die Markierung der Auslautverhärtung /d/ > /t/ erscheint in Hinblick auf die Auswertung ebenfalls problematisch und sollte überdacht und ggf. geändert werden!

11.4 Volle Nebensilbenvokale im Beleg

a, o, u in der Nebensilbe eines Beleges werden in der Normalform mit akutiertem *é* wiedergegeben, z. B. *~salé* *šalo*.

11.5 Vereinfachung von Geminaten im Auslaut und vor Konsonant

Vereinheitlichungsbedürftig.

11.6 Besonderheiten bei den schwachen Präterita

Wo der Mittelvokal schon vorahd. synkopiert worden ist – wie besonders bei allen Rückumlautfällen und den anomalen schwachen Verben –, wird die Synkope nicht durch Spitzwinkelklammern <*e*> markiert, sonst schon. Die Grenze ist vom Mhd. nicht immer genau zu ziehen, so daß NFORM hier zwangsläufig immer wieder Formen konstruieren wird, die man eigentlich anders ansetzen könnte oder müßte. Bei der KTX-Korrektur ist besonderes Augenmerk darauf zu achten, daß im Falle des sogenannten Rückumlauts das möglicherweise von NFORM produzierte spitzgeklammerte <*e*> in den Formen des Präteritums getilgt wird – der Rückumlaut ist ja gerade dadurch begründet, daß der Themavokal bereits sehr früh ausgefallen ist. Dies gilt auch bei den Verben zu beachten, die Präteritalformen sowohl mit als auch ohne Rückumlaut bilden, z. B. *zèl(e)n*.

Steht hsl. *t* gegen *zt* (wie in *sazte* ‘setzte’), wird weder das *z* noch das *t* spitzgeklammert, so daß *zt* gegen *t* stehen bleibt.

Besonders problematisch sind in diesem Zusammenhang die langwurzeli gen schwachen Verben. Ausgehend von der Silbenstruktur konstruiert NFORM derzeit deren Präteritum grundsätzlich ohne Bindevokal (z. B. *tèilte*, *wèinte*). Dies ist aber nur bei den **ja*-Verben (Klasse 1) korrekt, während bei den *ên*- und *ôn*-Verben Fälle von Synkope durch Spitzklammerung markiert werden müssen (korrekt wäre z. B. *wèin<e>te*). Bis eine automatisierte Lösung für diese

Problemfälle entwickelt ist, sind die betroffenen Formen bei der KTX-Korrektur anzupassen. Zur Unterscheidung der *ên*- und *ôn*-Verben siehe ggf.⁸⁰:

Frithjof Raven: Die schwachen Verben des Althochdeutschen. Bd. 1: Lang-, mehr- und kurzsilbige jan-Verba. Giessen 1963; Bd. 2: Die *ôn*- und *ên*-Verben, einschließlich eines ausführlichen lateinisch-alt-hochdeutschen Glossars zu beiden Bänden. Giessen 1967.

Bei langwurzigen schwachen Verben, die auf *t* oder *d* enden, folgt NFORM folgenden Regeln:

1. Wenn in der Handschrift Doppeldental steht, wird auch die Normalform mit Doppeldental geschrieben (z. B. *~huotte* zu handschriftlichem *\hutte*; *~bèitte* zu handschriftlichem *\beitte*).
2. Wenn in der Handschrift dagegen kein Doppeldental steht, wird die Normalform mit zwei Dentalen, aber Spitzklammerung des wurzelauslautenden Dentals geschrieben (z. B. *~huo<t>te* zu handschriftlichem *\hute* oder *~san<d>te* zu handschriftlichem *\sante*). Entsprechend werden von NFORM auch Formen wie *~rat<t>te* zu handschriftlichem *\ratte* als Präteritum von *rètten* erzeugt, obwohl Schreibungen wie *ratte* höchstens als Schreibfehler begehren dürften.

11.7 Auflösung des Nasalstriches

Wenn der finale Nasal im Dat. Sg. mask. und neutr. nur durch den Nasalstrich vertreten ist, wird in der Normalform *-m* angesetzt (nicht *-n*, wie von NFORM vorgegeben).

11.8 Einzelfälle

11.8.1 *-ære, -er* (Nomina agentis, *rîtare*)

Nomina agentis auf *ære, er* bekommen einen Lemmaansatz auf *`...ære*, die Normalform wird, wenn das schließende *e* fehlt, zu *\...ær<e>* angepasst. Dagegen werden folgende Fälle auf *`...er* angesetzt:

`...er < lat. *-er*, vgl. *`mèister* < lat. *magister*

`...er < lat. *-or*, vgl. *`phister* < lat. *pistor*

`...er < ig. *-êr*, vgl. *`muoter* < ig. **mâtēr* und andere Verwandtschaftsnamen.

Auch *\riter, \ritter* u. ä. werden als *`rîtare* lemmatisiert. Wenn aber eine Schreibung mit *\...tt...* vorliegt, dann wird die Normalform stärker, nämlich zu *~ritter* angepasst. Die Sonderbehandlung rechtfertigt sich durch die zu nhd. *Reiter* und *Ritter* führende Spaltung des Paradigmas.

11.8.2 *bi-gihte, bîhte*

Zum Lemma *bi-gihte* und handschriftlichem *bihte* lautet die Normalform *bî{igi}hte*.

11.8.3 *ëht, oht, ëckert, ockert*

Die Normalform wird großzügig an die handschriftliche Form angepasst, da kein allgemeiner Lautwandel, sondern lemmaspezifische Varianz vorliegt:

<i>\eht</i>	<i>~ëht</i>	<i>\ot</i>	<i>~ot</i>
<i>\et</i>	<i>~ët</i>	<i>\eckert</i>	<i>~ëckert (u. a.)</i>
<i>\oht</i>	<i>~oht</i>	<i>\ockert</i>	<i>~ockert (u. a.)</i>

11.8.4 *ent-* (Verbalpräfix)

Bei Verben, die mit *ent-* präfigiert sind, aber im “Lexen” und damit auch nach Projektkonventionen nur mit *en-* angesetzt sind, wird die Normalform in jedem Fall mit *ent-* angesetzt. Dabei muß das *t* ggf. geklammert werden:

⁸⁰ Möglicherweise wird auch speziell für die Bedürfnisse des Projektes eine Liste der betreffenden Verben erstellt.

Lemma	Handschrift	Normalform
<i>enphâhen</i>	<i>envie</i>	<i>en<t>-viè<ng></i>

Weiterhin ist darauf zu achten, daß in der Normalform *ph-* dann als *v-* angesetzt werden muß, wenn das entsprechende Simplex im “Lexer” mit *v-* angesetzt ist.

11.8.5 *heiden, kristen* und ähnliche Leute

Der Plural von Wörtern wie *hèiden*, *kristen* u. ä. wird häufig ohne erkennbare Flexionsendung gebildet (*die heiden* statt *die heidene*; *den heiden* statt *den heidenen*). In solchen Fällen werden in den Normalformen das finale *-e* bzw. *-en* geklammert: *~hèiden<e>*, *~hèiden<en>* (nicht *~hèid<en>en* oder *~hèide<ne>n*).

11.8.6 *-hèit* (Substantivsuffix)

- nach folgenden Regeln:

Hs. Wortform	Normalform
<i>snellekeit</i>	<i>snëllig-<h>èit</i>
<i>snellkeit</i>	<i>snëll-kèit</i>
<i>snellichkeit</i>	<i>snëllig-hèit</i>
	m f r k . a b e r : <i>snëllig-<h>èit</i>
<i>snellickeit</i>	<i>snëllig-<h>èit</i>
- Vereinheitlichen der Fälle von *...h-hèit* nach folgenden Regeln:

Hs. Wortform	Normalform
<i>richheit</i>	<i>rîch-hèit</i>
<i>richeit</i>	<i>rîch-<h>èit</i>
<i>mensheit</i>	<i>mèn<ni>sch-hèit</i>
<i>mensgheit</i>	<i>mèn<ni>sch-hèit</i>
<i>menscheit</i>	<i>mèn<ni>sch-hèit</i>
<i>mensgeit</i>	<i>mèn<ni>sch-<h>èit</i>

11.8.7 *-inne* bzw. *-în* (movierte Feminina)

Die Bildungen auf *-în* sind meistens oberdeutsch (bairisch), die auf *-inne* oder kurz *-in* eher mitteldeutsch. Das Lemma wird stets auf *-inne* angesetzt; wenn die Endung der handschriftlichen Wortform *-in* lautet, soll die Normalform je nach Gegebenheiten des Textes (vgl. evtl. Reimbindung) möglichst eine Anpassung erfolgen zu *-in<ne>* oder *-în*.

11.8.8 *iuwer*

Für die Normalformen des Possessivpronomens *iuwer* gelten folgende Regelungen:

Lemma	Handschrift	Normalform
<i>iuwer</i>	<i>ivr</i>	<i>iu<we>r</i>
<i>iuwer</i>	<i>iwr</i>	<i>iuw<e>r</i>
<i>iuwer</i>	<i>iuwer</i>	<i>iuwer</i>

Folgt dem *w* ein Vokal, wird die volle Form angesetzt, folgt kein Vokal, wird *w* geklammert.

11.8.9 *komen*

Beim Verb *komen* erfolgt keine Differenzierung des Lemmaansatzes nach *komen* und *quëmen*; angesetzt wird immer *komen*. Die Normalformen werden hingegen hinsichtlich Konsonantismus (Ankaut *k-* oder *qu-*) und Vokalismus (*komen* – *quëmen*; *kam*, *kâmen* – *kom*, *kômen*) an die handschriftlichen Formen angeglichen.

11.8.10 *lich* und *lîch-*

lich, wenn *i* in geschlossener, *lîch*, wenn *i* in offener Silbe steht.

11.8.11 *lâzen, liè* (nicht *liè<z>*) bei *\lie*

11.8.12 *oder* oder *eder*

Bei handschriftlichem *eder* 'oder' wird als Lemma *oder* angesetzt, die Normalform aber zu *èder* angepaßt.

11.8.13 *sîneme*

Bei der handschriftlichen Form *simme* zum Possessivpronomen *sîn* lautet die Normalform *sîn<e>me*.

11.8.14 *sol(e)n*

Die Normalformen zum Verb *sul(e)n* werden hinsichtlich Vokalismus und anlautendem Konsonanten an die handschriftlichen Formen angepaßt. Wenn die handschriftliche Wortform mit anlautend *sch-*, *sh-* oder *sc-* gebildet ist, wird die Normalform mit *sch-* gebildet. Der Vokalismus der Normalform wird ebenfalls an die handschriftliche Form angeglichen (z. B. 3SGI *sal* – *sol* – *schal* – *schol*). Die Graphie *v\o*, *u\o* für den Wurzelsilbenvokal wird in der Normalform einheitlich mit *u* wiedergegeben (z. B. Handschrift *fcu\olt*, Normalform *schul<e>t*).

11.8.15 *unde* (nicht *un<de>*) bei *un*-

11.8.16 *vrouwe* (Nebenformen)

Bei der seltenen Form *fer* für *vrouwe* wird in der Normalform *ver* angesetzt.

11.8.17 *vröuwen*

Vorschlag: Präteritum von *vröuwen* in der Normalform einheitlich *~vröuwete*, ggf. mit Spitzklammerung. Aber nicht *~vröu<w>t...* wie bislang gelegentlich.

11.8.18 *w* für *wu*

Bei der Schreibung *<w>* für normalmittelhochdeutsch *<wu>*, *<wuo>* (z. B. *wnder* = *wunder*; *whs* = *wuohs*) soll in der Normalform *wu* bzw. *wuo* ohne Klammerung stehen (ohne Spitzwinkelklammer, wie auch von Nform erzeugt).

Die graphonematische Segmentierung erfolgt aber *|w=w|=u|*; in Gra-par soll an entsprechender Stelle die Umgebungsbedingung ‚nach *w'*‘ angegeben werden.

11.8.19 *wa-* und *wô-*Stämme

Für die Normalformen der alten *wa-*Stämme (Lemmaansatz endet auf *-w*) gelten folgende Konventionen:

Endungslose Formen Nom. Sg. und Akk. Sg.:

Lemma	Handschrift	Normalform	
<i>sêw</i>	<i>sê</i>	<i>sê</i>	(ohne <i>-w</i>)
<i>sêw</i>	<i>seo</i>	<i>sêé</i>	(alter, voller Endsilbenvokal)

Prädikative Adjektive:

Lemma	Handschrift	Normalform	
<i>salw</i>	<i>salo</i>	<i>sal(é)</i>	(alter, voller Endsilbenvokal)
<i>varw</i>	<i>var</i>	<i>var<e></i>	

Endungslose Adjektive:

Lemma	Handschrift	Normalform
<i>varw</i>	<i>var</i>	<i>var<e></i>

An allen anderen Fällen muß die Zugehörigkeit des betreffenden Wortes zu den *wa-* / *wo-*Stämmen durch eine entsprechende Klammerung angezeigt werden, z. B. im Dat. Sg.:

Lemma	Handschrift	Normalform
<i>kniew</i>	<i>knie</i>	<i>knie<we></i>

11.8.20 *zer-* (Verbalpräfix)

Die Lemmata werden immer mit *zer-* angesetzt, die Normalformen aber ggf. durch Klammerung *<r>* an *r-*lose handschriftliche Formen angeglichen.

12 Anmerkungsfeld

Anmerkungen (auch zu Inkongruenzen und Abweichungen vom Erwarteten) stehen immer nur im Anmerkungsfeld.

12.1 Anzeichnung von Dittographien

/\([doppelt] im Anmerkungsfeld; Lemma, Wortart- und Normalformfeld werden mit [!] nach der Eingangsmarke gefüllt; das Formbestimmungsfeld bleibt leer (wie z. B. bei Adverbien).

13 Der Kontrolltext und seine Korrektur

Die KTX-Korrektur ist der letzte Arbeitsschritt, bei dem sämtliche Fehler der Kollationierung und Lemmatisierung auffallen und korrigiert werden sollen. Aus diesem Grund erfährt jeder Index grundsätzlich zwei aufeinander aufbauende KTX-Korrekturen, und zwar durch zwei besonders erfahrene Bearbeiter, von denen keiner schon für frühere Arbeitsschritte zuständig gewesen ist (wenn nicht mehrere Jahre dazwischen liegen). Eine etwas andere Möglichkeit besteht darin, dass die zwei KTX-Korrekturen unabhängig voneinander zur selben Zeit stattfinden und ihre Ergebnisse dann miteinander und mit der Ausgangsfassung verglichen werden (mit einem "Excel"-Vergleichsmakro). Von den beiden KTX-Bearbeitern oder einem dritten Bearbeiter wird dann für jede Abweichung entschieden, welche Variante die richtigere ist.

Zu überprüfen und ggf. zu korrigieren sind zu jeder einzelnen Wortform der Lemmaansatz, die grammatische Bestimmung und die Normalform, in Zweifelsfällen außerdem die Kollationierung. Alle Änderungen im Feld "handschriftliche Wortform" müssen unbedingt auch in die Textdatei übertragen werden!

13.1 KTX-Korrekturausdruck

Zur abschließenden Korrektur aller bisherigen Arbeitsgänge dient der Kontrolltext Y-KTX. Er bietet zeilenweise fortlaufend die Transkription, in der Zeile darunter die vom Programm NFORM erzeugten normalisierten Formen, darunter die grammatischen Bestimmungen und zuunterst die Lemmaansätze.

Der Kontrolltext Y-KTX erlaubt es also, die gesamte Lemmatisierung fortlaufend zu lesen und zu überprüfen. Zunächst muss sich der Bearbeiter für den betreffenden Index auf einer Liste eintragen, um Bearbeitungsüberschneidungen zu verhindern. Dann ist je eine Kopie der Normalformdatei Y-NF.0 und der Textdatei X.TXT, wie sie im aktuellen P-Verzeichnis zu finden sind, in ein eigenes Verzeichnis zu ziehen. Hier ist nun die Korrekturausdruckdatei Y-KTX aus dem Normalformindex Y-NF.0 zu erstellen:

<u>Programm:</u>	LEM, Option 9
<u>Eingabe-Datei:</u>	Y-NF.0
<u>Ausgabe-Datei:</u>	Y-KTX
<u>Option:</u>	Für Bildschirmbearbeitung (b, Enter)

Die Datei Y-KTX kann in Word geöffnet werden: Word öffnen, unter "Extras" "Optionen" "Allgemein" "Konvertierung beim Öffnen bestätigen" aktivieren, Datei öffnen: Y-KTX, "Codierter Text", dann "MS-DOS" oder "Westeuropäisch (DOS)" wählen. Die Word-Datei kann ausgedruckt werden.

(Andere Möglichkeit: Y-KTX wird mit WordPerfect (nicht mit dem Editor!) geöffnet; die Druckformatierung erfolgt dann durch das Makro KTX-D. Anschließend kann der Korrekturtext ausgedruckt werden. Es ist nicht unbedingt notwendig, die druckformatierte Datei zu speichern; wenn sie gespeichert werden soll, etwa weil zunächst nur ein Teil ausgedruckt worden ist, muß sie im WordPerfect-Format und unter dem Namen Y-KTX.D gespeichert werden.)

(Die Korrektur kann nun auf dem Ausdruck erfolgen. Doch müssen dann noch durch weitere Bearbeiter sämtliche Korrekturen in den Normalformindex Y-NF.0 eingearbeitet werden, was zeitaufwändig und fehlerträchtig ist. Mit dem Programm NF-KTX kann der Normalformindex umsortiert werden, um das Eintragen der Korrekturen zu erleichtern; nach Abschluß der Korrekturteintragungen muss auf jeden Fall die ursprüngliche Sortierung wiederhergestellt werden (ebenfalls mit dem Programm NF-KTX). Der Normalformindex Y-NF.0 wird nach Abschluß dieser Arbeit in Y-NF.K umbenannt.)

13.2 KTX-Korrektur am Bildschirm

Um Fehler besser erkennen zu können, ist es empfehlenswert, sich den gemäß 13.1 erzeugten Korrekturausdruck von Y-KTX vorzunehmen und auf Fehler durchzulesen, die anfallenden Berichtigungen aber gleich in eine Kopie der Normalformdatei Y-NF.0 und in eine Kopie der Textdatei X.TXT einzutragen. (Im Editor können die Dateien am einfachsten so verteilt werden: Y-NF.0 auf Datei 1, X.TXT auf Datei 3. In Datei 2 kann dann die Lemmaliste mit alt y geladen werden, ohne dass sie den Index oder die Textdatei überschreibt.) Der Normalformindex Y-NF.0 wird nach Abschluß dieser Arbeit in Y-NF.K umbenannt und kann in das Verzeichnis P-Ablage gestellt werden.

(Ferner gibt es die Möglichkeit, auch Y-KTX gleich auf dem Bildschirm im WordPerfect-Editor statt auf Papier zu lesen (nicht empfohlen): Zunächst ist das entsprechende Tastaturlayout zu aktivieren (Makro TASTATUR, Option 3). Die Datei Y-KTX ist in Datei 1 zu laden, der zugehörige Normalformindex Y-NF.0 in Datei 2, die Textdatei X.TXT in Datei 3. Diese Zuordnung ist genau einzuhalten. Mit dem Makro KTX-NEU kann nun das Makro KTX erzeugt werden, mit dem zu Beginn aller folgenden Arbeitssitzungen die drei Dateien automatisch in die richtigen Fenster geladen werden. Wenn sich mehrere Bearbeiter einen Computerarbeitsplatz teilen, sollte das Makro KTX umbenannt werden, damit jeder auf sein eigenes personalisiertes Makro zugreifen kann. — Bei der KTX-Korrektur am Bildschirm müssen die Lemmatisierungen und die Normalformen in der Datei Y-KTX gelesen werden; die Korrekturen erfolgen direkt in der Normalformdatei Y-NF.0. Der Vorteil der KTX-Korrektur am Bildschirm besteht also darin, daß gegenüber der KTX-Korrektur im Ausdruck das zeitraubende nachträgliche Eintragen der Korrekturen entfällt; nach Abschluß auch nur eines Korrekturdurchgangs steht direkt ein verbesserter Normalformenindex zur Verfügung. Die KTX-Korrektur am Bildschirm wird durch mehrere Makros unterstützt (siehe 19.2.3.)

14 Die Lemmaänderung

14.1 Änderungen in der Lemmaliste und Einarbeitung in die Indizes

Aus der laufenden Korpuspflege- und -erweiterung ergeben sich Änderungsvorschläge für die im Projekt geltenden Lemmmaansätze, wie sie in der Lemmaliste erfasst sind. Anmerkungen, Einwände und Änderungsvorschläge werden zunächst gesammelt, und zwar in der xls-Datei Ar-tus\\C:\PP\P_Lemmatisierung\LRG_Fragliches.xls.

Nötige Änderungen werden gelegentlich eingearbeitet:

- Das alte Lemma in C:\lem-wb\LRG wird geändert oder getilgt, wie sachlich angezeigt ist.
- Die neue LRG wird auf alle Rechner (unter C:\lem-wb) verteilt. (In C:\lem-wb befinden sich auf den meisten Rechnern nur LRG und LRG-best, auf dem Zentralrechner aber auch die Lemmatisierungswörterbücher.)
- Das alte Lemma wird im Korpus gesucht und ersetzt – unter Berücksichtigung etwa nötiger Änderungen im Wortart-, Bestimmungs-, Normalform- und (soweit vorhanden) Grapho-Feld.
- (– Anpassung des betroffenen Lemmatisierungswörterbuches.)

14.2 Große Änderung

14.2.1 Einarbeitung der angesammelten Änderungsvorschläge in die Lemmaliste

- Derzeitigen Ordner \lem-wb auf eigenen Rechner oder eigenes Verzeichnis kopieren.
- Die Datei \lem-wb\LRG in \lem-wb\LRG_neu umbenennen. Diese Datei enthält alle Lemmaeinträge. Die Dateien \lem-wb\LRG-... enthalten noch einmal alle Einträge, aufgespalten auf verschiedene Dateien, um einen einfachen Zugriff vom WP-Editor aus zu ermöglichen. Doch zwei Dateien in der LRG-Reihe enthalten etwas anderes: LRG-bst enthält Bestimmungsvorschläge, LRG-EN enthält Bestimmungsvorschläge und die Eigennamen-Einträge. (Die Eigennamen sind nicht in LRG enthalten.) — Die Dateien \lem-wb\LW-... enthalten Lemmatisierungswörterbücher.
- Sammelblätter mit Änderungsvorschlägen vornehmen.
- Änderungsvorschläge prüfen und gegebenenfalls in \lem-wb\LRG_neu einarbeiten, und zwar so:
 - Einen vorhandenen Ansatz korrigieren (Trennung zwischen alt und neu: Tabulator):⁸¹
`alt-lëmma #n niuwe-lëmma #n
 - Einen neuen Ansatz hinzufügen (Trennung zwischen alt und neu: Tabulator):
` niuwe-lëmma #n
 - Einen vorhandenen Ansatz tilgen, wenn möglich mit Angabe des stattdessen einzusetzenden Ansatzes – der auch ein sonst noch zu buchendes *niuwe-lëmma* sein kann (Trennung zwischen alt und neu und ander: Tabulator):
`alt-lëmma #n - ander-lëmma #n
- \lem-wb\LRG_neu wird für die spätere selbsttätige Änderung der Indizes im P-Verzeichnis und für die spätere Angleichung der Lemmatisierungswörterbücher aufgehoben.
- Aus \lem-wb\LRG_neu eine umgearbeitete Liste \lem-wb\LRG_neuneu erzeugen: Ein Programm setzt die Zeilen von \lem-wb\LRG_neu wie folgt in Zeilen einer Datei \lem-wb\LRG_neuneu um:

`alt-lëmma #n	niuwe-lëmma #n	(\lem-wb\LRG_neu)
`niuwe-lëmma #n		(\lem-wb\LRG_neuneu)
`alt-lëmma #n	-	(\lem-wb\LRG_neu)
Zeile wird getilgt		(\lem-wb\LRG_neuneu)
`	niuwe-lëmma #n	(\lem-wb\LRG_neu)
`niuwe-lëmma #n		(\lem-wb\LRG_neuneu)
- Neuen Ordner mit LRG_neuneu, LRG-bst und den lw-...-Dateien anlegen.
- LRG_neuneu in LRG umbenennen.
- LRG-EN an LRG anhängen, die anderen LRG-... nach lem-wb-alt (nur auf dem Zentralrechner) verschieben, alty.edm ändern, so dass alty.edm auf LRG zugreift, und nur LRG, LRG-bst und die lw-... (im neuen Ordner lem-wb) sowie alty.edm und die frühere Fassung altyalt.edm (im

⁸¹ ` ist der Gravis; man drücke Umschalt ´ (Akut) und dann Leerzeichen.

Ordner Editor) auf alle Rechner verteilen (die alten Ordner lem-wb auf den Rechnern sind vorher zu löschen!).

- Die möglichen selbsttätigen Änderungen in den Hauptkorpus-Indizes durchführen.

14.2.2 Anpassung der Ansätze des Korpus an die neue Lemmaliste

- Umwandlung der (adj) BA.
- (automatisch:) Aus den Hauptkorpus-Indizes werden alle Lemmaansätze (mit Wortart) gesammelt, die nicht in der derzeitigen Lemmaliste gegeben sind. Ein abweichender Lemmaansatz wird nur einmal aufgenommen, nicht jedes seiner Vorkommen; die Anzahl der Vorkommen wird aber mitaufgenommen. — Der Abgleich berücksichtigt nicht: Fremdsprachliches und Varianzen, die auf einer kontextbedingten Einengung der Genusangabe beruhen. Bezugsformen hinter “/” wie in “sehen/ane+ werden so berücksichtigt, als wären sie selbst eigene Stichwörter (allerdings bleibt hier die Wortart unberücksichtigt). — Zu den abweichenden Lemmata gehören von vornherein alle Namen (#...N); als automatischer Korrekturansatz für sie wird eine Form angeboten, in der alle e als e erscheinen, alle ^ und alle - getilgt sind. — Zu den anderen abweichenden Lemmata werden mit dem Levenstein-Algorithmus ermittelte ähnliche Ansätze aus der derzeitigen Lemmaliste vorgeschlagen. — Die Liste ist nach der Lautgestalt des betreffenden Simplex’ oder Hintergliedes sortiert.
Die Zeilen der Liste setzen sich wie folgt zusammen:

Zahl ⁸²	abweichender-ansatz	wortart	\$"	vorschlag #wortart
--------------------	---------------------	---------	-----	--------------------

oder, wenn eine Bezugsform hinter “/” wie in “sehen/ane+ eingesammelt wurde:

Zahl ⁸³	abweichender-ansatz	/-Zusatz!	\$"	vorschlag #wortart
--------------------	---------------------	-----------	-----	--------------------

oder, wenn ein Name eingesammelt wurde, mit automatischem Korrekturvorschlag:

Zahl ⁸⁴	abweichender-ansatz	PN ⁸⁵	\$"Neuernamensansatz	PN ⁸⁶
--------------------	---------------------	------------------	----------------------	------------------

(a_p_a_MhdAbgleichWortbestand.py mit LRG auf P anwenden.)
- Die Sammeldatei kann im Editor geöffnet und bearbeitet sowie für die Bearbeitung vorher auf mehrere Bearbeiter aufgeteilt werden. Nach dem Öffnen ist das Makro “tabu.edm” (liegt in Artus\\C:\PP\Editor-Makros neu, ist auf den gerade benutzten Rechner in C:\Editor zu kopieren) anzuwenden: **alt-F10, tabu, Enter**.
- Bearbeitung der Liste (**vorher neueste LRG von Artus\\C:\PP\P_Lemmatisierung holen!**):
 - In jeder Zeile darf nur rechts vom "\$" etwas verändert werden, nie links davon.
 - Wenn sich in der Lemmaliste ein mit Sicherheit richtiger Ersetzungsansatz ausmachen lässt, wird dieser hinter dem alten Lemma (nach dem "\$") eingefügt: **alt-y, Lemma eingeben, F7, F7 (und beim ersten Mal nach dem Öffnen des Editors nun noch: e)**; der so eingefügte neue Ansatz soll nicht mehr verändert werden:

Zahl ⁸⁷	abweichender-ansatz	wortart	\$"neuer-ansatz #wortart ^
--------------------	---------------------	---------	----------------------------

oder, wenn es sich um eine Bezugsform hinter “/” handelt:

Zahl	abweichender-ansatz	/-Zusatz!	\$"++neuer-ansatz #xyz ^ ⁸⁸
------	---------------------	-----------	--
 - Wenn man
 - den Kontext bräuchte, um den Lemmaansatz zu entscheiden, oder
 - den Kontext bräuchte, um den Wortartansatz zu entscheiden (nämlich zur Vereindeutigung des Genus), oder
 - die Wortart sich geändert hat, so dass das Bestimmungsfeld vielleicht geändert werden müsste (vor allem bei Änderung hin zu #adj oder #pron)**so wird ein ++ hinter das "\$" geschrieben und dahinter ein vermutlich oder möglicherweise richtiger Ansatz** (der dann im Nachbearbeitungsschritt f richtig gestellt werden wird):

Zahl	abweichender-ansatz	wortart	\$"++neuer-ansatz #wortart ^
------	---------------------	---------	------------------------------

oder, wenn es sich um eine Bezugsform hinter “/” handelt:

Zahl	abweichender-ansatz	/-Zusatz!	\$"++neuer-ansatz #xyz ^ ⁸⁹
------	---------------------	-----------	--
 - Wenn man einen Ansatz spalten will, schreibt man die beiden neuen Ansätze rechts und links von | und verfährt ebenso mit der Wortart:

Zahl	abweichender-ansatz	wortart	\$"neu-eins neu-zwei #wortart1 #wortart2 ^
------	---------------------	---------	--
 - Namen: BN / FIN werden, wenn möglich, Appellativ (und damit auch in die Lautauswertung einbezogen). Ansonsten bleiben sie BN / FIN (und werden der Lautauswertung entzogen). Auf die immer schwanke Unterscheidung BN gg. PN und FIN gg. ON wird kein Berichtigungsaufwand mehr gewendet.
Groß-/Kleinschreibung: Alle Namen (#...N) und die Einwohnerbezeichnungen auf -er(e) werden großgeschrieben, alles andere klein.
- Automatische Anwendung der Korrekturliste auf das Korpus: Ersatz durch Ersetzungsansatz. Marken (~ nach TAB~) für Nachbearbeitung (g) hinterlassen (wenn nicht schon TAB~~ vorliegt). Etwaiges ++ ins Statusfeld umsetzen; bei

⁸² Anzahl, wie oft der abweichende Ansatz im Korpus vorkommt.

⁸³ Anzahl, wie oft der abweichende Ansatz im Korpus vorkommt.

⁸⁴ Anzahl, wie oft der abweichende Ansatz im Korpus vorkommt.

⁸⁵ Oder ON, BN, FIN.

⁸⁶ Oder ON, BN, FIN je nach dem, was vorgefunden wurde.

⁸⁷ Anzahl, wie oft der abweichende Ansatz im Korpus vorkommt.

⁸⁸ Was bei “/”-Bezugsformen im Feld Wortart steht, ist beliebig; nur muss das Feld vorhanden sein und das “#” darf nicht fehlen. Möglich ist also z. B. auch \$"++neuer-ansatz # ^ ohne Inhalt.

⁸⁹ Was bei “/”-Bezugsformen im Feld Wortart steht, ist beliebig; nur muss das Feld vorhanden sein und das “#” darf nicht fehlen. Möglich ist also z. B. auch \$"++neuer-ansatz # ^ ohne Inhalt.

- Wortartänderungen ein ++ ins Statusfeld schreiben (wenn nicht schon gesetzt). In Fällen von BN / FIN > Appellativ ein `/#[BN]` oder `/#[FIN]` ins Kommentarfeld setzen. (**a_p_a_MhdAbgleichWortbestandErsetzenMitListe.py**)
- f Nachbearbeitung (halbautom. mit Makro):
1. Bei Marke | im Lemmafeld oder Wortartfeld (vgl. d.3) wird die Trennung des betroffenen Wortes eingeleitet. Von Hand ist im Belegfeld ein | an die Stelle der Trennung zu setzen (und sei es vor ein (=) oder vor ein #), ebenso im Normalformfeld und in einem etwaigen Graphofeld (vor oder nach dem Grapho-|). Wird kein | gesetzt, fragt das Makro nach, ob tatsächlich keine Trennung durchgeführt werden sollte; wenn nicht, ist dann nur noch wie unter f.2 zu verfahren. Ansonsten wird die Trennung soweit möglich automatisch durchgeführt, dann ist von Hand nachzubessern und wie unter f.2 zu verfahren. (**müle 042a, Eingabekürzel 967a**)
 2. Zu prüfen und gegebenenfalls zu berichtigen sind anhand des Kontextes und der Auskünfte in (...):
 - der Wortansatz (anhand der Lemmaliste),
 - die Wortart (vor allem zur Genusvereindeutigung) und
 - die Bestimmung (samt Zusatzanzeichnungen!).

Entweder:

Berichtigung bei Lauf über das P-Verzeichnis: Bei Marke ++ im Statusfeld (vgl. d.2) prüfen und berichtigen. Wenn dann das Makro weiterlaufen gelassen wird, wird das ++ von alleine gelöscht. (**müle 042a, Eingabekürzel 967b**)

Oder:

Berichtigung in einer xls-Datei, in die die ++-Zeilen zusammen mit Nachbarzeilen geschrieben sind (**a_p_a_Mhd_Auszug.py**):

 - Nur die mit >>> in der ersten Spalte ausgezeichneten Zeilen sind zu ändern.
 - Der Ordnungsteil darf nicht geändert werden.
 - Wenn mehr Kontext nötig ist als gegeben, markiere man das Feld mit dem Ordnungsteil und starte über einen Knopf in der Leiste das Makro **müle029c_StelleInPAufsuchen**.
 - Wenn fraglich ist, was man eintragen soll, oder wenn sonst eine zweite Bearbeitung der Stelle nötig erscheint, trägt man in die erste Spalte nach dem >>> seine Initialen ein (z. B. *m/* = Matthias Lexer).

Diese Bearbeitung wird zweimal unabhängig voneinander durchgeführt. Dann werden (in beiden Bearbeitungsergebnissen) die Initialenmarken und etwaige dahinterstehende Anmerkungen hinter das ++ ins Statusfeld geschrieben, dann die erste Spalte gelöscht. Die beiden Ergebnisdateien werden verglichen (und zwar nur die Stellen mit ++ im Standfeld), abweichende Stellen werden entschieden. (**ad-hoc-Python, müle008aa_VergleichVonNfOderGrIndizes**)

Im Vergleichsergebnis werden etwaige Anmerkungen in den ++-Zeilen und die nicht-++-Zeilen gelöscht. Die +-Zeilen werden wieder in das Korpus eingeflochten; das ++ wird gelöscht, wenn die abwesend gewesenen Zeilen nicht mehr unterschieden werden müssen. (**ad-hoc-Python, a_p_a_MhdZeilenEinflechten.py**)
- g Überspring- und aufschiebbar: Fortsetzung der Nachbearbeitung:
3. (autom.) Marke ^{TAB} durch "# ersetzen für Nform-Programm. (**ad-hoc-Python**)
 4. (autom.) Mit Nform-Programm bei Marke " neue Nform erzeugen; wenn mit alter gleich, dann Marke tilgen, sonst neben alte schreiben. (**NFORM.SPX**)
 5. (halbautom. mit Makro) Bei Leerzeichen im Normalformfeld Ansätze prüfen. (**müle 042a, Eingabekürzel 966**) (Oder: Zeilen mit Marke sammeln und Nformfelder im Block ändern.) Richtiges Angebot nehmen, falsches löschen.
- h Wiederholung des Abgleiches. Ausgegebene Ansätze durchsehen und die richtigen der Lemmaliste hinzufügen.

14.3 Änderungen mit lem-aend

Es ist unvermeidlich, daß sich einzelne Lemmaansätze auch nach Abschluß der Vorlemmatisierung und Indizierung als formal fehlerhaft erweisen. Solche fehlerhaften Lemmaansätze werden von Zeit zu Zeit möglichst in allen Indices auf einmal und auf die gleiche Weise korrigiert. Die Korrektur, die nur von dem damit beauftragten Projektmitarbeiter durchgeführt werden soll, erfolgt so weit wie möglich mit Hilfe des Programms LEM-AEND.

<u>Programm:</u>	LEM-AEND, Option 3 oder 4
<u>Eingabe-Datei:</u>	Y-NF.0 / Y-NF.K oder IND-LIST
<u>Ausgabe-Datei:</u>	Y-NF.LAE
<u>Änderungsliste:</u>	L-AEND.xxx im Verzeichnis C:\LEM-AEND

Die Änderungsliste besteht aus zeilenförmigen Datensätzen nach folgendem Muster:

ä" falsches Lemma #Kat. ^[TAB]x "korrigiertes Lemma #Kat. ^

Das Programm ersetzt in der Indexdatei / den Indexdateien jeweils das fehlerhafte Lemma und seine Kategorie durch das korrigierte Lemma und dessen Kategorie; die eventuell vorhandene grammatische Bestimmung wird nicht verändert. Jede geänderte Zeile wird durch ein doppeltes Fragezeichen ?? markiert; alle markierten Zeilen sind aufzusuchen und daraufhin zu überprüfen, ob die Ersetzung zurecht vorgenommen wurde und ob die Normalform korrigiert oder

angepaßt werden muß. Abschließend müssen die korrigierten Indices wieder unter ihrem ursprünglichen Namen gespeichert werden.

15 Der Grapho-Index

Das Programm GRAPHO erstellt eine graphonematische Analyse der handschriftlichen Wortformen eines lemmatisierten Index'. Dazu wird jede Wortform in die kleinstmöglichen Segmente aufgeteilt, die einem Phonem oder einer Phonemfolge der Normalform entsprechen (z. B. t=t, ch=h, v=uo, '=er). Die Folge dieser Eins-zu-Eins-Entsprechungen wird jeder Zeile des Normalformindex' als eigenes Feld angehängt.

<u>Programm:</u>	GRAPHO
<u>Eingabe-Datei:</u>	Y-NF.K
<u>Ausgabe-Datei:</u>	Y-GRA

Bei Mißlingen der automatischen Segmentierung schaltet GRAPHO auf eine manuelle Segmentierung um. Zeilen, die auch auf diese Weise nicht in Ordnung zu bringen sind, können mit einem Ausrufezeichen ! markiert werden.

Unter bestimmten Bedingungen können auch offensichtlich fehlerhafte Normalformen durch GRAPHO geändert werden; geänderte Normalformen werden von dem Programm mit einem Asterisk * und die zugehörigen Grapho-Felder durch zwei nachgesetzte Ausrufezeichen !! markiert. Diese Zeilen müssen nach Programmende makrogestützt überprüft und u.U. korrigiert werden.

Grapho ändert die Normalform in einigen Fällen selbständig, z. B. wenn eckige Klammern (vermeintlich) fehlen. Alle von Grapho veränderten und mit * gekennzeichneten Normalformen müssen im Anschluss sorgfältig überprüft werden. Wenn NF und handschriftliche Wortform stark voneinander abweichen, setzt Grapho häufig überflüssige eckige Klammern, daher sollte bei der manuellen Segmentierung darauf geachtet werden, dass die Normalform im NFO-Feld von Grapho mit der Normalform des Index' (erscheint in Grapho über dem NFO-Feld) übereinstimmt. Bei Superskripten kommt es manchmal vor, dass Grapho im WOFO-Feld zusätzliche Vokalzeichen generiert. Auch hier sollte immer ein Abgleich mit der handschriftlichen Wortform des Index' (steht ebenfalls über dem NFO-Feld) erfolgen.

Wenn das Programm problemlos durchläuft, erscheint am Schluss eine Erfolgsmeldung.⁹⁰ Sollte Grapho ohne diese Meldung abbrechen, ist der erzeugte Grapho-Index wahrscheinlich unvollständig.

Falls Grapho mit {esc} unterbrochen wird, ist es möglich, an dieser Stelle später fortzufahren, indem man in die abgebrochene Y-Gr-Datei die fehlenden Datensätze aus Y-NF.K kopiert und mit ## vor der Feldeingangsmarkierung " kennzeichnet (su: <FNZ>"; ers.: <FNZ>##"). Anschließend sollte die Datei umbenannt werden (keine Datei darf den Namen der Ausgabedatei tragen, ansonsten stoppt Grapho). Sie dient nun als Eingabedatei im Programm Grapho (Ausgabe: Y-Gr), Option "nur Datensätze, die mit ## markiert sind, bearbeiten".

15.1 Zur Gestaltung des Graphofeldes

Es ist darauf zu achten, dass Phoneme, die in der Normalform durch eine Zeichenfolge dargestellt werden, nicht getrennt werden: *sch, ch, qu, pf, tz*, alle Geminaten u. Diphthonge.

Wichtigste Richtlinie für die Gestaltung des Grapho-Feldes muss seine möglichst problemlose Auswertbarkeit, insbesondere auch mit Hilfe des Programms gra-par sein: Wenn die gra-par-Ergebnisdatei primär nach den normalmhd. Segmenten sortiert wird, so muss Zusammengehöriges auch zusammen erscheinen, z. B. erhaltenes festes, erhaltenes unfestes (rundgeklammertes) und geschwundenes (spitzgeklammertes) Schwa. Auf diesem Grundsatz basieren einige der folgenden Neuregelungen:

⁹⁰ Die Erfolgsmeldung muss vielleicht noch von Herrn Klein programmiert werden.

15.1.1 Nasalstrich über *m* und *n*

Der Nasalstrich kann – v.a. im 14. Jh. – eine Kürzung für *en* bedeuten (vgl. Römer, Geschichte der Kürzungen 45; Schneider, Got. Schriften I, 16) – bei mdt. Texten ist dies in 11l·5·V1_Aeg und IV·4b·P_PrMG nicht selten, sonst nur vereinzelt der Fall. In oberdeutschen Texten begegnet *m*-, *n*- insgesamt deutlich häufiger; hier handelt es sich wahrscheinlich meistens um Synkope von *e*, durch die der nachfolgende Konsonant silbisch wird.

In der Nform wird der Nasalstrich in beiden Fällen bislang als *en* aufgelöst.

Beispiel: "name ~namen \nam\-

Gestaltung des grapho-Feldes:

~en | \=-en |

Die Interpretation in Hinblick auf die lautlichen Verhältnisse muß noch vom gra-par-Bearbeiter geleistet werden.

15.1.2 Hsl. Schreibung <w> für normalmittelhochdeutsch <wu>, <wuo>

Bei hsl. Schreibung <w> für normalmittelhochdeutsch <wu>, <wuo> (z. B. *wnder*, *whs*) etc. der Normalform *wu* etc. stehen ohne Spitzwinkelklammer, weil *u* graphisch durch die Graphie <w> mitabgedeckt wird.

Gestaltung des grapho-Feldes:

~wu | w=w | u | Angabe der Umgebungsbedingung in "gra-par": *u* = 0 nach *w* bzw. *v*

Zuordnung bei der gra-par-Sortierung zu *w* und *u*. Es besteht keine Diskrepanz zwischen Nform und grapho-Feld.

15.1.3 Wortformen mit rundgeklammerten (*en*) und (*ni*) in der Nform:

~n(ni) | nn=nn | i=i | [i] |, da *i* in Synkopierungsposition]

15.1.4 Hsl. Wortformen \menscheit etc., Nform ~mèn(ni)sch-hèit etc.

Gestaltung des grapho-Feldes:

| sc=sch | gra-par: vor *h*

| h=h | gra-par: nach *sc* für *sch*

Begründung:

- Die Gestaltung muss einheitlich sein und kann nicht von der relativen Häufigkeit der <sc>-Schreibung für *sch* abhängig gemacht werden.
- Die alternativen Regelungen a) |sch=sch|=h| und b) |sc=sch|h=h| sind im Prinzip gleichwertig, doch spricht für b), dass es dem älteren –insbesondere ahd. – Stand entspricht, der auch nach Durchsetzung der Graphie <sch> für *sch* beibehalten wurde.

15.1.5 Klammern im grapho-Feld

- Spitzwinkelklammern

Spitzwinkelklammern sind im Grapho-Feld redundant, da die entsprechende Information auch schon durch das "Nichts" links vom Gleichheitszeitszeichen und durch 0 (Null) in der Graphie-Spalte der gra-par-Ergebnisdatei ausgedrückt wird. Außerdem führt die Spitzwinkelklammer links dazu, dass erhaltene und geschwundene normalmhd. Segmente in der gra-par-Ergebnisdatei auseinander sortiert werden.

Deswegen wird in Zukunft auf Spitzwinkelklammern im Grapho-Feld verzichtet. In gra-par werden z. B. alle ausgefallenen Schwa – letztlich ohne Informationsverlust – unter e – 0 aufgeführt (vgl. Beispieldateien BLK).

Anmerkung: Es könnte vorteilhaft sein, wenn nur die erste Spitzwinkelklammer getilgt würde. In Gra-Par gibt es die Umgebung "vor <e>", also vor geschwundenem Schwa, die unbedingt von der Umgebung vor erhaltenem Schwa unterschieden werden sollte.

- Rundklammern

Aus denselben sortiertechnischen Gründen entfällt künftig die linke Rundklammer normalmhd. Segmente im Grapho-Feld, z. B. |e=e)| statt |e=(e)|.

15.1.6 Sonderformen

Stark abweichende handschriftliche Wortformen werden in den Fällen morphologisch bedingter Variation (wo also nicht nur andere Dialektlautung vorliegt) angeglichen. Dies gilt v. a. für folgende Fälle (s. auch die Sonderfälle bei der NF-Erstellung):

- *van* statt *von*

15.1.7 Fragliches

Selten generiert Grapho zusätzliche Vokale: Rapp 17803.01 g|e| |t|o|e|er|r|ent| für WoFo \ geto\errent. oder Konsonanten: BaGB 147,06 wértlichî --> wértlichîc

bei handschriftlichem sk fügt Grapho z.T. h zw. s und k ein (Phys?)

Bei PN erzeugt Grapho einen Asterisk vor dem ersten Segmentierungsstrich |. Dieser Asterisk wird automatisch gelöscht, wenn die betreffende Form in die manuelle Segmentierung kommt.

In der Normalform von Beinamen korrigiert Grapho automatisch Großbuchstaben in Kleinbuchstaben (mit Markierung *), aber ohne Aussternung des Graphofeldes

Projektbesprechungen:

Grapho-Programm wird daraufhin überprüft, ob es Folgen in *[*]-Klammern auslässt.

Bisher geht bei vorzeitigem Abbruch des Grapho-Programmes leicht der Rest des Index unbemerkt verloren. Warnmeldung einbauen?

Erfolgsmeldung nach nicht-vorzeitigem Ende des Programmes. [schon eingebaut?]

ht: ~naht \nath: |n=n|a=a|th=ht| [so in automat. Segmentierung], nicht: |n=n|a=a|=h|th=t| [so in automatischer Segmentierung, wenn in NF h geklammert --> vorher korrigieren].

Komplett eckig Geklammertes wird automatisch ausgesternt; Problem: Teilklammerung wird nicht immer richtig umgesetzt --> eigentlich eckig geklammerte Schreibungen können so in Auswertung gelangen.

Ausgesterntes: in Grapho nicht zusätzlich gekennzeichnet (wird also in Gra-Par mitgezählt); im Fall von Meri sind 2 Seiten nicht aus der Hs., sondern aus der Ed. übernommen u. daher ausgesternt!

Gra-Par: Klammern hinter Kommata werden in Par 1 übernommen. Bsp.: wibe,,) zu wibe) in Par 1.

16 Das Reimregister

Auf der Grundlage des korrigierten Normalformindexes Y-NF.K können mit dem Programm REIM und seinen verschiedenen Unterprogrammen die Reimregister erzeugt werden. Eingriffe des Bearbeiters sind nur bei der Erstellung der Isolierung der Reimwörter in der Datei Y-REI.0 und bei der abschließenden Überprüfung der ausgedruckten Reimindices nötig. Letztere ist vor allem bei den Indices für die Reimkonsonanten und -endsilben erforderlich, da REIM noch nicht richtig funktioniert, wenn wie z. B. im Reim *gên* : *sê* nur eines der beiden Reimwörter einen Reimkonsonanten enthält. Volle Nebensilbenvokale im Reim sollten zur adäquaten Analyse in den Normalformen mit einem Akzent versehen sein (z. B. *todés* : *is*, *wandelét* : *got*); diese Markierung der vollen Nebensilbenvokale sollte bereits durch das Programm NFORM oder spätestens bei der KTX-Korrektur erfolgt sein.

16.1 Die Isolation der Reimwörter

In einem ersten Schritt müssen die Reimwörter isoliert werden.

<u>Programm:</u>	REIM, Option 1
<u>Eingabe-Datei:</u>	Y-NF.K
<u>Ausgabe-Datei:</u>	Y-REI.0

Die Ausgabedatei beinhaltet zeilenförmige Datensätze, die aus der Normalform, einer Zeilen- oder Verszählung, teilweise der handschriften Wortform und der grammatischen Bestimmung besteht.

Beispiel:
ge-nësen|_0001 (stv5 i)
wësen|_0002 (stv5 i)
minnen|_0003 (mjnnen, swv i)
wërlt-dingen|_0004 (werlt#dingen, n DP)
under-tân|_0005 (vndertân, adj part -)
ge-hôr-sam|_0006 (adj -)
ge-louben|_0007 (m AS)
schouwen|_0008 (scowen, swv i)
ge-ëichenet|_0009 (geechinot, swv pV)
ge-zëichen<e>t|_0010 (gezeichenot, swv pV)
ver-zalt|_0011 (uerzalt, swv pV)
manig-valt|_0012 (manicfalt, adj ODP)
sage|-0013 (swv-a 3SGK)
sunnentage|_0014 (m AP)
ge-wisse|_0015 (gwise, adv)
mësse|_0016 (misse, f DS)

Die Ausgabedatei Y-REI.0 muß aus der regelmäßigen, ungestörten Abfolge kompletter Reimpaare bestehen. Da das Programm REIM nicht in der Lage ist, selbst zu prüfen, ob diese Bedingung erfüllt ist (dazu wäre ein hochkompliziertes besonderes Programm nötig), müssen Störungen vom Bearbeiter ermittelt und bereinigt werden. Solche Störungen können entstehen

1. durch den Ausfall von Versen oder Reimwörtern in der Handschrift. In diesem Fall muß auch die Zeile mit dem isolierten Reimwort aus Y-REI.0 getilgt werden; die Reimanalyse eines Reimworts ohne Reimpartner wäre ja auch sinnlos.

2. durch das Auftreten von Dreireimen: Das Programm REIM ist (gegenwärtig noch) nicht in der Lage, mit Dreireimen fertigzuwerden. In diesem Fall muß die zweite Zeile eines Dreireims in Y-REI.0 verdoppelt werden, so daß formal zwei komplette Reimpaare entstehen. Die Stellenangabe der duplizierten Zeile muß durch einen * markiert werden, damit der Dreireim später erkennbar bleibt!

Zum Auffinden solcher Störungen lädt man Y-REI.0 im WordPerfect-Editor und läßt den Bildschirm mit der Plustaste aus dem Ziffernblock ("Num" muß deaktiviert sein!) jeweils um eine Bildschirmseite weiterlaufen; die beiden untersten Zeilen müssen dann immer ein Reimpaar bilden. Wenn dies nicht mehr der Fall ist, muß sich die Störung innerhalb der vorausgehenden 24 Zeilen befinden.

16.2 Die Reimanalyse

Zur weiteren Analyse werden die beiden Reimwörter eines jeden Reimpaars aufgegliedert in Reimwortanfang, Reimvokal, Reimkonsonanz und Endsilbe.

<u>Programm:</u>	REIM, Option 2
<u>Eingabe-Datei:</u>	Y-REI.0
<u>Ausgabe-Datei:</u>	Y-REI.1

Die letzten drei Komponenten werden mit den Entsprechungen des Reimpartners verglichen und, wenn sich unterscheiden, in der Ausgabedatei Y-REI.1 mit einem Asterisk * markiert. Nicht vorhandene Reimkonstituenten werden durch ^B vertreten.

Beispiel:

ge-n|ë|s|en|_0001 (stv5 i)
w|ë|s|en|_0002 (stv5 i)
m|i|nn*|en|_0003 (mjnnen, swv i)
wërlt-d|i|ng*|en|_0004 (werlt#dingen, n DP)
under-t|â*|n*|^B|_0005 (vndertân, adj part -)
ge-hôr-s|a*|m*|^B|_0006 (adj -)
ge-l|ou|b*|en|_0007 (m AS)
sch|ou|w*|en|_0008 (scowen, swv i)
ge-|èi|ch|enet|_0009 (geechinot, swv pV)
ge-z|èi|ch|enet|_0010 (gezeichnenot, swv pV)
ver-z|a|lt|^B|_0011 (uerzalt, swv pV)
manig-v|a|lt|^B|_0012 (manicfalt, adj ODP)
s|a|g|e|_0013 (swv_a 3SGK)
sunnent|a|g|e|_0014 (m AP)
ge-w|i*|ss|e|_0015 (gwissee, adv)
m|ë*|ss|e|_0016 (misse, f DS)
|è*|nst|e|_0017 (f AS)
k|ü*|nst|e|_0018 (cunste, fu AP)

16.3 Das Anlegen der drei Reimindices

Aus der Datei Y-REI.1 werden nun drei separate Indices für die Reimvokale, die Reimkonsonanten und die Reimendsilben erstellt.

<u>Programm:</u>	REIM, Option 3
<u>Eingabe-Datei:</u>	Y-REI.1
<u>Ausgabe-Datei 1:</u>	Y-REI.V (= Reimvokale)
<u>Ausgabe-Datei 2:</u>	Y-REI.K (= Reimkonsonanten)
<u>Ausgabe-Datei 3:</u>	Y-REI.E (= Reimendsilben)

Diese drei Reimindices werden nach den Reimkonstituenten sortiert.

<u>Programm:</u>	SORTA; Feldbegrenzer: ALT+179
<u>Eingabe-Datei:</u>	Y-REI.V (oder Y-REI.K, Y-REI.E)
<u>Ausgabe-Datei:</u>	Y-REI.VS (oder Y-REI.KS, Y-REI.ES)
<u>Sortierreihenfolge:</u>	2,4,3,5 (bei Y-REI.KS und Y-REI.ES: 2,4,1,3)

Die Sortierreihenfolge 2,4,3,5 bedeutet, daß Y-REI.V primär nach dem Reimvokal des ersten Reimworts, sekundär nach dem Reimvokal des zweiten Reimworts, tertiär nach der Folgekonsonanz des ersten Reimworts und viertens nach der Folgekonsonanz des zweiten Reimworts sortiert wird.

Zuletzt müssen die sortierten Reimindices für den Ausdruck formatiert werden.

<u>Programm:</u>	REIM, Option 4
<u>Eingabe-Datei:</u>	Y-REI.VS (oder Y-REI.KS, Y-REI.ES)
<u>Ausgabe-Datei:</u>	Y-REI.V (oder Y-REI.K, Y-REI.E)

Die Ausgabe-Datei Y-REI.V (bzw. Y-REI.K, Y-REI.E) wird mit WordPerfect (nicht mit dem Editor!) geöffnet und mit dem Makro REIM für den Druck formatiert. U. a. wird dabei die Codierung der jeweiligen Reimkonstituenten durch geschweifte Klammern in Steuerzeichen für Fettdruck umgesetzt. Die mit dem Makro REIM bearbeiteten Dateien werden können nun ausgedruckt werden.

Beispiel:

a : a

haben (swv Inf) : knaben (m AS) 299

ge_schach (gescah, stv5 3S Pt I) : sprach (sprah, stv4 3S Pt I) 157

ge_schach (gescah, stv5 3S Pt I) : un_ge_mach (mn AS) 475

sach (sah, stv5 3S Pt I) : ge_schach (gescach, stv5 3S Pt I) 237

schaden (scaden, m DP) : er_laden (irladen, stv6 Pz Pt) 415

zins_haft (adj) : kraft (craft, fu DS) 535

kraft (craft, fu NS) : be_haf<t>t (behaft, swv Pz Pt) 367

a : andere Vokale

naht (f(u) DS) : er_dâht (irdaht, swv Pz Pt) 249

wart (stv3b 3S Pt I) : ge_kârt (swv Pz Pt) 351

wart (stv3b 3S Pt I) : ge_lêr<e>t (gelart, swv Pz Pt) 170[2]

â : â

Pilâ (pyla, EN NS) : dâ (adv) 281

ge_nâde (gnade, f AS) : un_ge_nâde (f DS) 109

mâge (m AP) : wâge (f AS) 377

lâgen (stv5b 3P Pt I) : phlâgen (plagen, stv5 3P Pt I) 183

er_dâhte (irdahte, swv 3S Pt I) : brâhte (anv 3S Pt I) 433

17 Bearbeitung der Lemmatisierungswörterbücher

Zur Verbesserung der automatischen Vorlemmatisierung müssen die Lemmatisierungswörterbücher durch Einspeisung neuer Einträge kontinuierlich verbessert werden. (Die Lemmatisierungswörterbücher befinden sich auf dem Zentralrechner unter C:\lem-wb, während dieses Verzeichnis auf den anderen Rechnern nur LRG und LRG-best enthält.) Im Zuge der Vorlemmatisierung können einzelne Einträge ergänzt werden (dazu siehe 8.5); es ist aber auch möglich, die Lemmatisierungswörterbücher durch alle in ihnen noch nicht enthaltenen Einträge eines neu bearbeiteten Textes zu ergänzen. Die Lemmatisierungswörterbücher sollen nur von Mitarbeitern mit ausreichender Erfahrung oder ggf. zusammen mit diesen bearbeitet werden.

Ausgangspunkt einer solchen Ergänzung ist einerseits ein komplett korrigierter Index Y-NF.K, andererseits die unmittelbare Vorstufe eines Lemmatisierungswörterbuchs der Form LW-Z.Kn mit n = 1,2,3, ...

Zunächst wird mit dem Programm LEM-2, Option 1, die erste Vorstufe eines Lemmatisierungswörterbuches erzeugt. Sie enthält ein Feld mit der normalisierten Wortform und ein Feld mit der maximalen grammatischen Bestimmung.

Programm: LEM-2, Option 1

Eingabe-Datei: Y-NF.K

Ausgabe-Datei: Y-LW.1

Anschließend ist eine alphabetische Sortierung von Y-LW.1 nötig:

Programm: SORTA

Eingabe-Datei: Y-LW.1

Ausgabe-Datei: Y-LW.1S

Es folgt die Zählung und Zusammenfassung identischer Zeilen in der Datei Y-LW.1S

Programm: LW-ERG, Option 1

Eingabe-Datei: Y-LW.1S

Ausgabe-Datei: Y-LW.V

Ein wichtiger Zwischenschritt gilt der Aufspürung von Fehlern und Inkonsistenzen in den Ansätzen der Lemmata und in den grammatischen Bestimmungen. Diesem Zweck dient das interaktive Programm LW-KOR (= LW-ERG, Option 2); es sucht aufeinanderfolgende Zeilen, die bei

identischer normalisierter Wortform Abweichungen in den Ansätzen von Lemma, Kategorie und Bestimmung aufweisen, die möglicherweise auf einer Inkonsistenz des Ansatzes beruhen. Relativ häufig treten etwa Abweichungen in den Genusangaben oder in den Angaben zu obligatorischem und fakultativem Pluralumlaut auf. LW-KOR zeigt dann beide Zeilen auf dem Bildschirm und fragt, ob sie unter einer Zeile zusammengefaßt werden sollen, und wenn ja, unter welcher.

Wenn keine der angebotenen Optionen zufriedenstellend ist, kann die erste betreffende Zeile von LW-KOR durch einen Asterisk * markiert und der Fall später in der Ausgabedatei manuell korrigiert werden. Solche markierten Zeilen werden zusätzlich in eine Hilfsdatei aufgenommen, ebenso auch alle Zeilen, die durch LW-KOR mit einer anderen Zeile zusammengefaßt worden sind. Die Hilfsdatei erlaubt daher erstens, alle mit LW-KOR getroffenen Entscheidungen noch einmal nachzuvollziehen und gegebenenfalls zu korrigieren; außerdem werden auf diese Weise Fehler in dem zugrundeliegenden Normalformindex Y-NF.K festgehalten, die durch LW-KOR zwar in der Ausgabedatei Y-LW.K, aber damit natürlich noch nicht in der gleichfalls betroffenen Datei Y-NF.K korrigiert werden können. Die Fehler im Normalformindex Y-NF.K müssen daher anschließend manuell verbessert werden.

<u>Programm:</u>	LW-ERG, Option 2 (LW-KOR)
<u>Eingabe-Datei:</u>	Y-LW.V
<u>Ausgabe-Datei:</u>	Y-LW.K
<u>Hilfsdatei:</u>	Y-LW.H

Entscheidungen bei der Anwendung von LW-KOR müssen mit größter Sorgfalt getroffen werden, da Fehlentscheidungen in das erweiterte Lemmatisierungswörterbuch eingehen und dadurch multipliziert werden können!

Die korrigierte Datei Y-LW.K wird sodann mit der zur aktuellen Version des Lemmatisierungswörterbuchs gehörigen Vorstufe LW-Z.Kn zusammengeführt:

<u>Programm:</u>	LW-ERG, Option 3
<u>Eingabe-Datei 1:</u>	LW-Z.Kn
<u>Eingabe-Datei 2:</u>	Y-LW.K
<u>Ausgabe-Datei 1:</u>	LW-Z.K(n+1)
<u>Ausgabe-Datei 2:</u>	Y-LW.ERW

In die Ausgabedatei Y-LW.ERW werden alle aus Y-LW.K nach LW-Z.K(n+1) übernommenen Einträge geschrieben. Wenn das Programm in Y-LW.K und LW-Z.Kn Zeilen mit identischer Wortform und identischem Lemma, aber unterschiedlicher grammatischer Bestimmung findet, muß der Bearbeiter interaktiv entscheiden, welche der beiden Zeilen nach LW-Z.K(n+1) übernommen werden soll.

Ogleich auf diese Weise schon ein Teil der zwischen LW-Z.Kn und Y-LW.K bestehenden Abweichungen in den Lemmaansätzen ermittelt und behoben werden, können auch noch durch die Zusammenführung der beiden Dateien Inkonsistenzen oder Inkonsistenzen entstehen und / oder erkannt werden. Deshalb sollte auf die Ausgabe-Datei LW-Z.K(n+1) noch einmal das Korrekturprogramm LW-KOR angewendet werden:

<u>Programm:</u>	LW-ERG, Option 2 (LW-KOR)
<u>Eingabe-Datei:</u>	LW-Z.K(n+1)
<u>Ausgabe-Datei:</u>	LW-Z.Km
<u>Hilfsdatei:</u>	LW-Z.Hm

Anschließend muß LW-Z.Km wieder in LW-Z.K(n+1) umbenannt werden. Es sei noch einmal daran erinnert, daß die Entscheidungen bei der Anwendung von LW-KOR mit größter Sorgfalt getroffen werden müssen, da Fehlentscheidungen in das erweiterte Lemmatisierungswörterbuch eingehen und dadurch weiter verbreitet werden!

Im letzten Schritt wird aus LW-Z.K(n+1) die neue erweiterte Fassung von LW-Z erzeugt:

<u>Programm:</u>	LW-ERG, Option 4
------------------	------------------

Eingabe-Datei: LW-Z.K(n+1)

Ausgabe-Datei: LW-Z

Es bleibt nun noch die Erweiterung des Ergänzungswörterbuchs LW-Z.EG(n+1), das bei der Vorlemmatisierung durch das Programm LEM-3 neben dem Lemmatisierungswörterbuch LW-Z herangezogen wird. Das Ergänzungswörterbuch wird mit Hilfe des Programms LW-ERG, Optionen 5 und 6, erweitert.

Zunächst wird aus LW-Z.K(n+1) die dem Ergänzungswörterbuch zugrundeliegende Datei LW-Z.E(n+1) erzeugt:

Programm: LW-ERG, Option 5

Eingabe-Datei: LW-Z.K(n+1)

Ausgabe-Datei: LW-Z.E(n+1)

Dann wird LW-Z.E(n+1) mit SORTA zu LW-Z.ES(n+1) alphabetisch sortiert:

Programm: SORTA

Eingabe-Datei: LW-Z.E(n+1)

Ausgabe-Datei: LW-Z.ES(n+1)

Schließlich wird aus LW-Z.ES(n+1) das neue Ergänzungswörterbuch LW-Z.EG erzeugt:

Programm: LW-ERG, Option 6

Eingabe-Datei: LW-Z.ES(n+1)

Ausgabe-Datei: LW-Z.EG

Achtung: LW-Z.EG hat ein fixed-length-Format und ist daher weder mit WP-Editor noch gar mit WordPerfect editierbar! Ein Versuch, das Ergänzungswörterbuch mit solchen Programmen zu laden, führt unweigerlich zum Systemabsturz!

18 Die Projektprogramme

18.1 Grundsätzliches

Der Start der in SPITBOL geschriebenen Projektprogramme ist auf folgende Weisen möglich:

1. Durch die Eingabe von "r[BLANK][Programmname (ohne Endung 'spx')]" direkt hinter dem DOS-Prompt.
2. In Windows über "Start" – "Ausführen" – Eingabe von "r[BLANK][Programmname]".
3. In Windows **XP** über die Eingabeaufforderung (Start – Alle Programme – Zubehör - Eingabeaufforderung): Eingabe von "cd\[Enter] cd[Blank]spx[Enter] r[Blank][Programmname]"
4. Sofern das betreffende Programm kompiliert ist (erkennbar an der Endung "exe"), kann es auch im Windows-Explorer oder "Arbeitsplatz" durch einen Doppelklick auf die Programmdatei gestartet werden.

Die Kompilation eines SPITBOL-Programmes (d. h. die Erzeugung einer ausführbaren Programmdatei) erfolgt im DOS mit dem Befehl exe oder exe-w⁹¹ und dem Programmnamen, z. B. c:\>exe-w such-nf. Es erscheint der Startbildschirm des betreffenden Programms; damit ist die Kompilierung beendet, und das gerade gestartete Programm kann mit dem entsprechenden Befehl direkt wieder beendet werden. Als Ergebnis der Kompilierung befindet sich im Verzeichnis c:\spx eine ausführbare Datei, z. B. such-nf.exe.

Folgende Besonderheiten sind bei den Projektprogrammen zu beachten:

- Die Beendigung / der Abbruch eines Programmes ist in der Regel mit [F7] möglich.
- Die Programmprozeduren werden in der Regel mit ESC gestartet.
- Bei der Auswahl von Eingabe- und Ausgabedateien können über ALT+V die zuletzt benutzten Dateien eingefügt werden.

⁹¹ Mit exe erfolgt die Kompilation für Win98, mit exe-w für WinXP (diese Kompilationen sind meist auch unter Win98 lauffähig). Im Verzeichnis c:\spx befinden sich dazu zwei Batch-Dateien exe.bat und exe-w.bat.

- Mit "beliebiger Taste" ist teilweise keine "beliebige" Taste, sondern ENTER gemeint.
- Der Wechsel zwischen einzelnen Eingabefeldern ist durch [TAB] möglich.
- [BACK-SHIFT] funktioniert in der Regel nicht; der Cursor ist vor die zu löschende Stelle zu setzen und dann [Entf] zu nutzen.

18.2 A-ZU-H

Das Programm A-ZU-H ermöglicht, Textdateien mit ausgabenorientiertem Ordnungsteil in Textdateien mit handschriftenorientiertem Ordnungsteil zu überführen und umgekehrt. Notwendige Voraussetzung ist, daß die betreffende Textdatei neben dem primären Ordnungsteil eine sekundäre Zählung im laufenden Text in geschweiften Klammern enthält. Folgende Arbeitsschritte sind zur vollständigen Überführung des einen Typs Textdatei in den anderen Typ notwendig:

Programm: A-ZU-H

Eingabe-Datei: X.TXT (oder X.TXH)

Ausgabe-Datei: X.TXH (oder X.TXT)

Die Ausgabedatei muß mit dem WordPerfect-Editor weiterbearbeitet werden. Zunächst ist der nun einfache [TAB] zwischen Ordnungsteil und Text durch [TAB] und nachfolgenden [BLANK] zu ersetzen. Anschließend wird der kurze Ordnungsteil am Zeilenanfang durch den vollständigen Ordnungsteil mit Quellensigle ersetzt werden. Dazu steuert man den Cursor in die Zeile nach dem Header und unmittelbar vor Beginn des Textes und läßt dann ohne weitere Bestätigung die Ersetzung von "[FNZ]" durch "[FNZ]Ordnungsteil" durchführen. Zuletzt muß noch der komplette Ordnungsteil in geschweiften Klammern im laufenden Text reduziert werden. Dazu läßt man die Ersetzung von "{Ordnungsteil}" durch "{" durchführen.

18.3 BASIS-S

18.3.1 Voraussetzungen

BASIS-S benötigt folgende Dateien:

1. Die Gesamt-Lemmaliste LRG-GES, die sich im Ordner C:\LEM-WB\ befinden muß. Die Liste enthält (so gut wie) alle Lemmata von Lexer und Findebuch. Ggf. muß zunächst der Ordner C:\LEM-WB eingerichtet und die Datei LRG-GES dorthin kopiert werden.
2. Die Korpuslemmaliste Korp-LST, die die Lemmata aller bisher lemmatisierten Quellen und die Siglen der Korpusquellen, in denen das jeweilige Lemma belegt ist, enthält. Erzeugt wird Korp-LST wird mit dem gleichnamigen Programm Korp-LST (Option 3). Eingabedatei ist dabei die Datei IND-LIST in Verzeichnis C:\SPX:
3. IND-LIST enthält die Dateinamen aller bisher erstellten NF-Indices, die zugehörigen Quellensiglen und die Namen der Textdateien. Die NF-Indices müssen sich in Unterverzeichnissen des Ordners C:\PP\P an der in IND-LIST angegebenen Stelle befinden. Diese Informationen benötigt BASIS-S, für das IND-LIST daher gleichfalls verfügbar sein muß.
4. Eine aus Korp-LST gezogene Teilliste von WBK, die nach ihrer Wortbildung zusammengehören. Diese WBK-Teilliste kann mit Hilfe des Programms WBK-TEIL (Option 1) und/oder mit dem Editor-Makro SUV erstellt werden. Für eine systematische und vollständige Aufgliederung empfiehlt sich eine mehrfache Anwendung von WBK-TEIL, um etwa Präfixbildungen, Suffixbildungen und Komposita auseinanderzusortieren.

Bei Komposita trennt WBK-TEIL die mutmaßlichen Kompositionsglieder durch einen senkrechten Strich | und erleichtert damit die Bearbeitung von Komposita durch BASIS-S. Daher sollte eine Komposita-Teilliste mit WBK-TEIL erstellt und möglichst danach noch manuell auf korrekte Trennung (Positionierung des Strichs) überprüft werden. Zusammenbildungen / Wortgruppen-derivate und Zusammenrückungen / Wortgruppenkonversionen können algorithmisch nicht befriedigend von Komposita getrennt werden. Komposita im Sinne von WBK-TEIL schließen daher Zusammenbildungen und Zusammenrückungen ein. Bei der Analyse mit BASIS-S sollten sie je-

doch durch Anwahl des betreffenden Wortbildungstyps (Zusammenbildung bzw. Konversion) von den Komposita unterschieden werden.

18.3.2 Aktualisierung dieser Dateien

Nach der Fertigstellung neuer lemmatisierter Indizes (NF-Indizes) müssen ihre Dateinamen zunächst in die Datei IND-LIST (siehe oben unter 3) eingefügt werden. Dann wird die aktualisierte KORP-LST mit dem Programm KORP-LST und der aktualisierten Datei IND-LIST neu erzeugt (siehe oben unter 2). Vorher muß die alte Datei KORP-LST in KORP-LST.A umbenannt werden.

Durch Vergleich von KORP-LST mit KORP-LST.A ist zu ermitteln, welche Lemmata neu hinzugekommen sind:

<u>Programm:</u>	WBK-TEIL, Option 3
<u>Eingabe-Datei 1:</u>	KORP-LIST
<u>Eingabe-Datei 2:</u>	KORP-LIST.A
<u>Ausgabe-Datei:</u>	KORP-LIST.NEU

KORP-LST.NEU muß dann in derselben Weise in Teillisten zerlegt werden wie zuvor die alte KORP-LST (siehe oben unter 4). Die neuen Teillisten können dann mit BASIS-S bearbeitet werden; die Ergebnisdateien werden an die schon vorhandenen angehängt.

18.3.3 Arbeitsweise von BASIS-S

Nach dem Start von BASIS-S muß zunächst bestimmt werden, ob eine neue WBK-Teilliste oder eine Ergebnisdatei von BASIS-S bearbeitet werden soll. Dann sind die Namen der Eingabedateien und der Ausgabedatei anzugeben; dabei muß die erste Eingabedatei eine WBK-Teilliste, die zweite die aktuelle KORP-LST sein. Die Festlegung bzw. Auswahl der Dateinamen kann auch mit Hilfe eines kleinen Dateimanagers erfolgen, der mit ALT+D aufgerufen wird. Mit ALT+V können die Dateinamen des letzten Programmlaufs von BASIS-S eingefügt werden. Die Eingabe aller Dateinamen ist mit ESC abzuschließen.

Anschließend muß angegeben werden, ob die WBK-Teilliste (vornehmlich) Suffigierungen, Präfigierungen oder Komposita enthält. Dazu erscheint folgender Bildschirm:

Festlegung des Wortbildungstyps
1 – Suffigierung
2 – Präfixbildung
3 – Kompositum
Bitte wählen (1-3 + ENTER):

Danach liest BASIS-S die erste Zeile der WBK-Teilliste ein und bearbeitet sie folgendermaßen: Anhand der Quellenbezeichnungen des Quellenfeldes wird ein Zahlenschlüssel erstellt, der genauen Aufschluß darüber gibt, wo und wie die Wortbildungskonstruktion im Korpus belegt ist. Dieser **“Distributionsschlüssel”** enthält für jedes der 25 Korpussegmente (I-0, II-1, II-2 usw.) eine Stelle; sie erhält den Wert

- 0, wenn die WBK im Segment nicht vorkommt;
- 1, wenn die WBK nur in einem Verstext des Segments belegt ist;
- 2, wenn die WBK nur in einem Prosatext des Segments belegt ist;
- 3, wenn sie sowohl in einem Prosa- als auch in einem Verstext des Segments belegt ist.

Im nächsten Arbeitsschritt zerlegt BASIS-S die WBK nach Maßgabe des zuvor angegebenen Wortbildungstyps in ihre mutmaßlichen unmittelbaren **Konstituenten** (UK). Das Basismorphem wird in das B(asis)-Feld, das Affix in das P(räfix)- bzw. S(uffix)-Feld geschrieben. Bei Komposita wird die mutmaßliche Basis des Erstgliedes des Kompositums in das zweite B(asis)-Feld, die des Zweitgliedes in das erste B(asis)-Feld geschrieben.

Anschließend sucht BASIS-S die mutmaßliche Basis zunächst in der alphabetisch sortierten Gesamtlemmaliste. Die Suche wird im Erfolgsfall beim ersten mit der mutmaßlichen Basis identi-

schen Lemma beendet, ansonsten beim ersten Lemma abgebrochen, das lexikalisch größer als die mutmaßliche Basis ist. Im Erfolgsfall wird die Kategorie des gefundenen Lemmas in das Kategorienfeld der Basis übertragen, das bei gescheiterter Suche mit einem Fragezeichen ? gefüllt wird. Die Suche wird beim ersten Versuch oft scheitern, weil sich das Basismorphem von der Basis lautlich unterscheidet, und zwar meist durch das Fehlen oder Vorhandensein der Endung -e oder -(e) und / oder durch Umlaut oder Umlautlosigkeit des Wurzelvokals. Nach gescheiterter Suche verändert BASIS-S daher die mutmaßliche Basis in entsprechender Weise und unternimmt dann eine erneute Suche.

Nach erfolgreicher Suche in der Gesamtlemmaliste sucht BASIS-S die mutmaßliche Basis in der Korpuslemmaliste KORP-LST. Wird sie dort gefunden, so berechnet BASIS-S auch für die Basis den Distributionsschlüssel und schreibt ihn in das Feld D(istributionsschlüssel).

Sodann schaltet BASIS-S in den **interaktiven Modus** um. Es zeigt seine bisherigen Ergebnisse auf einem Arbeitsbildschirm und wartet auf die Reaktion des Benutzers. Der Benutzer kann dann

a) mit STRG+H einen Hilfebildschirm aufrufen; er informiert über die Abkürzungen des Arbeitsbildschirms und über Besonderheiten, die bei der Eingabe und beim Löschen von Text zu berücksichtigen sind;

b) mit ESC die Ergebnisse bestätigen und die Analyse der nächsten WBK veranlassen;

c) die Einträge im Feld B(asis) und K(ategorie) manuell verändern und dann mit STRG+S eine erneute Suche in der Gesamtlemmaliste oder mit STRG+K eine Suche im Korpus und Berechnung des Distributionsschlüssels veranlassen;

d) mit STRG+L in das Feld L(emmaliste) gehen, mit Pfeil-nach-oben/ Pfeil-nach-unten nach einem geeigneten Basislemma suchen und dies gegebenenfalls mit ENTER als Basis übernehmen, mit Bild-nach-oben ins B(asis)-Feld ohne Übernahme als Basis zurückspringen;

e) mit STRG+D eine graphische Darstellung der Distribution von WBK und Basis im Korpus aufrufen;

f) mit STRG+T in das T(yp)-Feld gehen und einen anderen Wortbildungstyp wählen; es gelten dieselben Tastenfunktionen wie für das L(emmalist)-Feld. Beim Springen in das T-Feld verändert BASIS-S "Suffigierung" in "Präfigierung" und umgekehrt;

g) nach Änderung des Wortbildungstyps mit STRG+R eine Reanalyse der WBK veranlassen; BASIS-S zerlegt dann die WBK nach Maßgabe des neu gewählten Wortbildungstyps und wiederholt anschließend die Auswertungsschritte;

h) mit STRG+M die Markierung ändern; es gelten dieselben Tastenfunktionen wie für das L(emmalist)-Feld. Der Default-Wert ist ? = "problematisch"; die Alternativwerte sind: - = "unproblematisch", ! = "auffällig";

i) mit STRG+N in das N(otiz)-Feld springen und dort beliebige Anmerkungen eingeben. Das N(otiz)-Feld besteht aus drei separaten Zeilen (ohne automatischen Zeilenumbruch), zwischen denen man sich mit den Tasten Pfeil-nach-oben/Pfeil-nach-unten bewegt. Rücksprung mit Bild-nach-oben;

j) mit STRG+W veranlassen, dass die Belege der WBK im Korpus mit Kontext angezeigt werden. Der Kontext umfasst je nach Zeilenlänge der Textdatei 5–10 Zeilen. Zur besseren Lesbarkeit wird der angezeigte Kontext vereinfacht (Tilgung von hier unwesentlichen Präeditierungszeichen und der Shift-Klammern etc.).

Bei **Doppelmotivation** hat man es mit alternativen Basen und in der Regel auch mit verschiedenen Wortbildungstypen zu tun. Im Rahmen von BASIS-S ist damit folgendermaßen zu verfahren:

1. Der Benutzer muß entscheiden, ob die von BASIS-S angebotene Analyse und Motivationsmöglichkeit tatsächlich die primäre und wahrscheinlichste ist. Wenn dies nicht der Fall ist, muß er mit den entsprechenden Schritten die Alternative als 1. Basis anwählen.

2. Für eine eventuelle zweite Analyse und Motivationsmöglichkeit stehen in BASIS-S die Felder der "2. Basis" bereit. Am besten geht man so vor: Man geht in das B(asis)-Feld der 2. Basis, ändert dann mit STRG+T den Wortbildungstyp, veranlaßt mit STRG+R einer Reanalyse der WBK,

sucht mit STRG+S die zweite Basis in der Gesamtlemmaliste und gegebenenfalls mit STRG+K auch im Korpus.

3. Im Übrigen stehen für die Bearbeitung der 2. Basis dieselben Möglichkeiten zur Verfügung wie für die 1. Basis.

18.3.4 Die Ergebnisdatei von BASIS-S

Schließt der Benutzer die Bearbeitung einer WBK mit Esc ab, so berechnet BASIS-S zunächst für alle Korpusbelege den ihnen zukommenden **Basisrang** und hängt ihn in eckigen Klammern an die Quellenbezeichnung des WBK-Belegs an. Im Falle von zwei Basen (Doppelmotivation, Kompositum) müssen für jeden WBK-Beleg zwangsläufig auch zwei Basisränge berechnet werden. Sie erscheinen beide, durch den Slash / getrennt, in der der Quellenbezeichnung beigefügten eckigen Klammer.

Das **Bearbeitungsergebnis** speichert BASIS-S schließlich in einem Datensatz ab, der folgendermaßen strukturiert ist:

- (1) [Markierung][WBK-Lemma]#[Kategorie][TAB][Distributionsschlüssel der WBK]
- (2) [TAB]Q: Quellenfeld der WBK mit Basisrängen in eckigen Klammern
- (3) [TAB]Wortbildungstyp: 1. Basis + P:Präfix S:Suffix [TAB] +/-Distributionsschlüssel der 1. Basis
- (4) [TAB]Q1: Quellenfeld der 1. Basis
- (5) [TAB]Wortbildungstyp: 2. Basis + P:Präfix S:Suffix [TAB] +/-Distributionsschlüssel der 2. Basis
- (6) [TAB]Q2: Quellenfeld der 2. Basis
- (7) [TAB]N: Notizen (die drei Zeilen des Notiz-Feldes werden, durch | getrennt, aneinander gehängt)

18.4 GRAPHO

Siehe 15.

18.5 IND-NORM

Das Programm IND-NORM überprüft die Zeilen von NF-Indizes auf eine korrekte Feldstruktur und korrekte Feldeingangsmarkierungen. Das Grapho-Feld wird derzeit noch nicht berücksichtigt. Eingabedatei ist eine Dateiliste vom Typ der IND-LIST; in der Ausgabedatei werden die fehlerhaften Zeilen ausgegeben.

18.6 KORP-LST

Siehe 18.3.1.

18.7 KORP-STA

Das Programm KORP-STA erstellt eine Type-Token-Statistik für ausgewählte Wortartenkategorien, bezogen auf die Korpustexte und das (lemmatisierte) Gesamtkorpus. Eingabedateien sind die aktuellen Versionen der Indexliste IND-LIST und der Lemmaliste KORP-LST (mit Belegzahlen!), die ggf. mit dem Programm KORP-LST erstellt werden muß.

18.8 LEM

Siehe 8.1, 8.4, , 10.1, 10.2, 10.4, 12.

18.9 LEM-AEND

Siehe 14.

18.10 MORPHO

Das Programm MORPHO ergänzt die grammatische Bestimmung der Pronomina um zusätzliche Angaben und markiert Relativpronomina.

18.11 NF-KTX

Das Programm NF-KTX bringt die Felder eines Normalformindex in eine für die Eingabe der KTX-Korrekturen besser geeignete Reihenfolge (Option 1: "NF-Reihenfolge umstellen") oder stellt die ursprüngliche Reihenfolge der Felder des Normalformindexes wieder her (Option 2: "NF-Reihenfolge wiederherstellen"). Im Dateinamen der umgestellten Datei, d.h. der Ausgabe-Datei von NF-KTX (1), sollte NF durch NK ersetzt werden.

18.12 NFO-WOK

Siehe 10.8.

18.13 NFORM

Siehe 10.6.

18.14 NFORM-TX

NFORM-TX erzeugt aus einem NFO- oder LEM-Index eine Textdatei. Nach Wahl trennt NFORM-TX Wortformen mit handschriftlichem Zeilenumbruch wieder auf und stellt den abgetrennten Teil an de Anfang der Folgezeile.

18.15 NONS

Das Programm bietet eine Reihe nützlicher Tools:

- 1 Ersetzen mehrerer Einzelzeichen oder Strings
- 2 Codieren/Rückcodieren von Umlaut- und Akzentvokalen
- 3 Teilen einer Datei in bis zu 9 Teildateien
- 4 Umstellen und/oder Auswählen von Datensatzfeldern
- 5 Tilgen identischer Zeilen
- 6 Überführung einer Datei in eine Ausgabe-Datei im SPITBOL-fixed-length-Format

18.16 PRUEF-NF

Siehe 10.7.

18.17 SORTA

Das Programm SORTA sortiert zeilenförmige Datensätze, deren Felder in jeweils gleicher Weise voneinander getrennt sind (etwa durch [TAB], Komma, Semikolon). Die Sortierfolge der Felder kann durch den Benutzer bestimmt werden; das Programm analysiert den ersten Datensatz und ermöglicht dann, die Sortierfolge der Felder durch die Angabe der kommagetrennten Feldnummern zu bestimmen. Bei rückläufiger Sortierung muß ein R, bei absteigender Sortierung ein A vor die betreffende Feldnummer gesetzt werden.

18.18 SUCH-NF

Das Programm SUCH-NF sucht in einzelnen oder mehreren NFO-Indices nach Zeilen, die einen – sehr komplex formulierbaren – Suchbegriff enthalten und schreibt diese Zeilen in eine Ausga-

be-Datei. Es ist auch möglich, nach einer Kombination von bis zu neun verschiedenen Suchbegriffen zu suchen. Auf diese Weise ist es möglich, rasch Belege für ein bestimmtes und sehr genau bestimmtes Phänomen (Lemmata, Wortbildungsmuster, Wortarten, bestimmte grammatische Formen, auch in beliebiger Kombination) aus den Indices zu extrahieren und zur gezielten Auswertung bereitzustellen.

Der Startbildschirm verlangt die Entscheidung, ob ein **einzelner NFO-Index (Option 1)** oder **mehrere Indices zugleich (Option 2)** durchsucht werden sollen.

Die **Eingabedatei** ist bei Option 1 direkt der gewünschte Index (mit genauem Pfad und Dateinamen), bei Option 2 eine Datei, in der die zu durchsuchenden Indices aufgelistet sind. Dabei ist darauf zu achten, daß diese Datei nach Art der Datei IND-LIST (die natürlich verwendet werden kann) die Pfade und Dateinamen zeilenweise und jeweils mit einem [TAB] abschließend nennt. Als **Ausgabedatei** kann ein beliebiger Pfad und Dateiname vorgegeben werden. Die Eingaben werden mit [ESC] abgeschlossen.

Der folgende Bildschirm erlaubt die Festlegung verschiedener Optionen, die den Suchvorgang und die Ausgabedatei betreffen. Vor allem muß hier festgelegt werden, für wieviele aufeinander folgende Zeilen die Abfrage gelten soll (1 bis 9; Standard: 1). Im Ordnungsteil können die Raum- und die Textsortenangaben durch TAB voneinander getrennt werden (Standard: N – nur in wenigen Ausnahmefällen ist diese Option sinnvoll). In der Regel wird der Ausgabedatei eine ausführliche Statistik angehängt (Standard: J). Schließlich können ausgesternte Zeilen berücksichtigt oder übergangen werden (Standard: N – ausgesternte Zeilen übergehen).

Mit [ESC] gelangt man schließlich zu der Maske zur **Formulierung der Suchkriterien**. Für folgende Felder des NFO-Indexes können Suchkriterien angegeben werden: Zeile, Lemma, Kategorie, Flexion, Normalform, Wortform, Stelle (= Ordnungsteil), Status, Kommentar. Es ist möglich, Suchkriterien in einem, mehreren oder allen Feldern anzugeben und damit ggf. einen sehr komplexen Suchbegriff zu formulieren. Im Einzelnen sind dabei folgende Konventionen und Festlegungen zu beachten:

Tastaturbelegung:

ü = [STRG+u]

â = [ALT+1]

é = [ALT+2]

ë = [STRG+e]

< = [ALT+8]

> = [ALT+9]

· = [STRG+p]

¨ = [STRG+l]

Backspace = [Shift+Backspace]

SNOBOL-Syntax:

Suchbegriffe: stets in einfache oder doppelte Anführungszeichen setzen⁹²

„oder“: | (Tastatur: [ALT+0])

Anfang eines Feldes: pos(0) (Syntax: Suchbegriff, BLANK, pos(0))

Ende eines Feldes: rpos(0)

beliebige Zeichen „a“, „b“ oder „c“: any(“abc”)

nicht „a“, „b“ oder „c“: notany(“abc”)

„nicht“: NICHT

Start einer Schleife: W [am Anfang des Feldes ZEILE] (wenn der definierte Suchbegriff gefunden wird, wird die Suche erneut begonnen; wenn etwa solange gesucht werden soll, bis ein Verb gefunden wird, lautet der Suchbegriff NICHT „swv“ | „stv“ | „anv“ – solange kein Verb gefunden wird, ist der Suchbegriff erfüllt, und das Programm geht in die Schleife und sucht weiter)

Vordefinierte Begriffe im Feld „Lemma“:

⁹² Dies gilt nur für Strings, also Wörter oder Buchstabenfolgen, nach deren Vorkommen gesucht wird, nicht aber die unten genannten vordefinierten Begriffe!

jedes Personalpronomen *ër, ich, dû, wir, ir*: pers.pron
 Vordefinierte Begriffe im Feld "Kategorie":
 jedes Verb: verb
 jedes Substantiv: substantiv
 jedes Adjektiv oder Partizip: adj
 umlautfähiges Substantiv: umlaut
 Vordefinierte Begriffe in den Feldern "Lemma" und "Normalform":⁹³
 beliebiger Vokal *aëiouäöüèâêîôûæ* (außer Schwa): v0 oder vokal0
 ein beliebiger Vokal: v oder vokal
 beliebiger Vokal einschl. Schwa: v_e0 oder vokal_e0
 ein beliebiger Vokal einschl. Schwa: v_e oder vokal_e
 beliebiger Kurzvokal *aëiouäöüè*: kv0 oder kurzvokal0
 ein beliebiger Kurzvokal: kv oder kurzvokal
 ein beliebiger Langvokal: langvokal
 ein beliebiger Diphthong: diph
 beliebiger Konsonant: konsonant0
 ein beliebiger Konsonant: konsonant
 beliebiger Obstruent: obstruent0
 ein beliebiger Obstruent: obstruent
 ein beliebiger Nasal: nasal
 ein beliebiger Liquid: liquid
 beliebiger Dental: d0 oder dental0
 ein beliebiger Dental: d oder dental
 ein beliebiger Glide: glide
 ein beliebiger Labial: labial
 beliebiges e oder é ("Schwa") = any('eé'): E
 beliebiges e oder é ohne Klammern, in Rund- oder Spitzklammern = E | (E) | <E>: E.k
 Vordefinierte Begriffe im Feld "Wortform":
 beliebige Interpunktion: int

Der **Suchvorgang** startet mit [ESC]. Nach Abschluß der Suche bietet das Programm die Möglichkeit, entweder das **Ergebnis anzuzeigen (Option 1)**, das **Programm erneut zu starten (Option 2)** oder das **Programm zu beenden (F7)**.

18.18.1 Beispiele

18.18.1.1 Einfache Suche

Suche nach einem Lemma mit bestimmten Bestandteilen:

Lemma<1>: 'mèister'

Gefunden werden alle Indexzeilen, deren Lemmata die Buchstabenfolge mèister beinhalten (etwa: *mèister, mèister(e)n, mèister-schaft, mèister-lich* etc.

Suche nach einem genau bestimmten Lemma:

Lemma<1>: pos(0) 'mèister' rpos(0)

Gefunden werden alle Indexzeilen, die genau die Buchstabenfolge mèister beinhalten und sonst nichts (mit pos(0) wird verlangt, daß der Suchbegriff am Anfang des Feldes beginnt, mit rpos(0) wird festgelegt, daß der Suchbegriff auch am Ende des Feldes stehen muß).

Suche nach einer bestimmten Wortart:

Kategorie<1>:'swv'

⁹³ Die Bezeichnungen für diese vordefinierten Begriffe werden nicht in Anführungszeichen gesetzt!

Gefunden werden alle Indexzeilen mit der Buchstabenfolge swv im Kategorienfeld, das heißt, alle schwachen (swv) und alle anomalen schwachen Verben (swv-a).

Kategorie<1>: 'v'

Gefunden werden alle Indexzeilen mit dem Buchstaben v im Kategorienfeld, das heißt, alle Verben (stv, swv, anv, swv-a), aber auch Adverbien (adv).

Eine genauere Definition muß entweder durch exakte Angabe des Suchbegriffs (stv3b) oder durch die Zusätze pos(0) und rpos(0) erfolgen (etwa: pos(0) 'm' rpos(0) findet alle Wörter, die im Kategorienfeld ausschließlich als Maskulina ohne Umlaut bestimmt sind; pos(0) 'f' hingegen findet alle Wörter, deren Kategorienbeschreibung mit f beginnt, aber weitere Bestandteile haben kann, aber nicht haben muß, etwa fn und f<u>).

Suche nach einer bestimmten grammatischen Bestimmung:

Flexion<1>: '3SGI'

Gefunden werden alle Indexzeilen mit der grammatischen Bestimmung 3. Pers. Sg. Präs. Ind.

Flexion<1>: '3SG'

Gefunden werden alle Indexzeilen mit der grammatischen Bestimmung 3. Pers. Sg. Präs. Ind. oder Konj.

Flexion<1>: '3S'

Gefunden werden alle Indexzeilen mit der grammatischen Bestimmung 3. Pers. Sg. Präs. oder Prät., Ind. oder Konj.

Flexion<1>: '3'

Gefunden werden alle Indexzeilen mit der grammatischen Bestimmung 3. Pers. Sg. oder Pl., Präs. oder Prät., Ind. oder Konj.

Flexion<1>: 'K'

Gefunden werden alle Indexzeilen mit der grammatischen Bestimmung Konj.

Sonstige Suchen:

Suchbegriffe lassen sich für alle Felder der Indices formulieren; es ist beispielsweise auch möglich, nach bestimmten Buchstabenfolgen im Wortformenfeld oder im Normalformenfeld zu suchen.

18.18.1.2 Kombinierte Suchbegriffe

Zur genaueren Suche können die Suchkriterien für mehrere Indexfelder zugleich festgelegt werden. Auch dafür einige Beispiele:

Suche nach einer bestimmten Wortart mit einer bestimmten grammatischen Bestimmung:

Kategorie<1>: 'stv2'

Flexion<1>: '3SGI'

Gefunden werden alle Formen für die 3. Pers. Sg. Präs. Ind. von starken Verben der Klasse II.

Kategorie<1>: 'u'

Flexion<1>: 'P'

Gefunden werden alle Pluralformen von umlautfähigen Substantiven.

Kategorie<1>: 'stv'

Flexion<1>: '2SVI' | 'VK'

Gefunden werden alle Formen für die 2. Pers. Sg. Prät. Ind. oder (|) den Konj. Prät. der starken Verben.

Suche nach lautlichen Erscheinungen:

Lemma<1>: 'oè'

Kategorie<1>:'swv'
Flexion<1>: 'V'
Normalform<1>: 'ô'

Gefunden werden alle schwachen Verben (swv) mit dem Wurzelvokal oè, die im Präteritum (V) Rückumlaut (ô in der Normalform) aufweisen.

Lemma<1>: 'è'
Kategorie<1>:'swv'
Flexion<1>: 'V'
Normalform<1>: 'a' | 'â'

Gefunden werden alle schwachen Verben (swv) mit dem Wurzelvokal è, die im Präteritum (V) Rückumlaut (a oder â [wegen *dènken* – *dâhte*] in der Normalform) aufweisen.

18.18.1.3Zweizeilige Suchbegriffe

Die Anzahl der Zeilen, für die ein Suchbegriff formuliert wird, ist im Startbildschirm festzulegen.

Suche nach morphosyntaktischen Phänomenen:

Lemma<1>: 'dër'
Kategorie<1>:'art'
Kategorie<2>:'adj'
Flexion<2>: 'w'

Gesucht wird nach schwach flektierten Adjektiven, denen der bestimmte Artikel *dër* unmittelbar vorausgeht.

18.18.1.4Mehrzeilige und komplexe Suchbegriffe

Abfragen können sich auf bis zu neun aufeinander folgende Indexzeilen beziehen; die Anzahl der Zeilen, für die ein Suchbegriff formuliert wird, ist im Startbildschirm festzulegen.

Suche nach morphosyntaktischen Phänomenen:

Lemma<1>: 'dër'
Kategorie<1>:'art'
Kategorie<2>:'adj'
Kategorie<3>:substantiv

Gesucht wird die unmittelbare Folge aus Artikel, Adjektiv und Substantiv.

Kategorie<1>:'präp'
Kategorie<2>:'art'
Kategorie<3>:substantiv
Flexion<3>: 'D'

Gesucht wird die unmittelbare Folge aus einer den Dativ regierenden Präposition, einem Artikel und einem Substantiv (im Dativ).

18.18.1.5Suche mit Schleifen

Beispiel: Gesucht wird nach Perfektkonstruktionen vom Typus "Form von *haben*, gefolgt von einem Partizip, in beliebigem Abstand, aber innerhalb einer syntaktischen Einheit". Dazu muß ein Suchbegriff für drei Zeilen formuliert werden:

Suchbegriff Zeile 1:

Lemma<1>: pos(0) 'haben' rpos(0)
Kategorie<1>:'swv-a'
Flexion<1>: pos(0) any('123')
Wortform<1>: NICHT int

Zur Erläuterung: Gesucht wird nach Indexzeilen mit dem Eintrag haben im Lemmafeld; die Zeile beginnt (pos(0)) und endet (rpos(0)) mit haben; jedwede Komposita sind folglich ausgeschlossen. Zusätzlich ist die Kategorie mit swv-a definiert. Als Flexionsform werden beliebige Formen der 1., 2. und 3. Person gesucht (pos(0) any('123')); ausgeschlossen sind damit der Infinitiv sowie die Partizipien. Die letzte Bedingung wird durch NICHT int im Wortformenfeld definiert: Es darf im Wortformenfeld kein Interpunktionszeichen stehen (da andernfalls keine Perfektkonstruktion mit nachfolgendem Partizip vorläge).

Suchbegriff Zeile 2:

Zeile<2>: W

Flexion<2>: NICHT 'pV' rpos(0)

Wortform<2>: NICHT s.int

Zur Erläuterung: Wenn in Zeile 2 kein Partizip Präteritum (NICHT 'pV' rpos(0)) und keine Interpunktion (NICHT s.int) folgt, wird die Suche fortgesetzt (W), und zwar solange, bis der Suchbegriff für Zeile 3 gefunden wird. Wenn hingegen ein Interpunktionszeichen gefunden wird, ist der negativ formulierte Suchbegriff für Zeile 2 nicht mehr erfüllt, und die Suche wird abgebrochen.

Suchbegriff Zeile 3:

Flexion<3>: 'pV' rpos(0)

Zur Erläuterung: Gesucht wird nach einem beliebigen Partizip Präteritum ohne weitere grammatische Bestimmung (daher rpos(0)).

18.19 TX-PRUEF

Das Programm TX-PRUEF überprüft Textdateien auf Standardgemäßheit; besonders die fehlerträchtigen Shift-Klammern werden daraufhin geprüft, ob sie auch wieder geschlossen werden.

18.20 WBK-TEIL

Siehe 18.3.1.

18.21 ZAHL

Das Programm ZAHL dient der Erstellung, Ergänzung oder Überprüfung der Zählung von Textdateien oder Indices.

18.21.1 Die Erstellung / Ergänzung einer (fortlaufenden) Numerierung

Damit das Programm ZAHL eine fortlaufende Numerierung erstellen kann, müssen zwei Bedingungen erfüllt sein: Zum einen muß in jeder Zeile dem Text ein TAB vorausgehen; zum anderen muß in gewissen Abständen eine Zeilenzahl am Zeilenanfang stehen (und vom Text durch einen TAB getrennt sein).

Nach dem Start des Programms ist eine Auswahl unter folgenden drei Optionen möglich:

1. Zeilen-Numerierung
2. Seiten- und Zeilen-Numerierung
3. Vers- und Wortzahl mit Nullen auffüllen

Die Option 1 (Zeilennumerierung) gilt sowohl für eine einfache Zeilenzählung von Prosatexten wie auch eine durchgehende Verszählung. Die Option 2 (Seiten- und Zeilen-Numerierung) wird auch angewandt bei Texten wie dem "Parzival" Wolframs von Eschenbach, bei dem seit Lachmann die Verse jeweils in Dreißigereinheiten gezählt werden. Mit der Auswahl von Option 3 ist es möglich, die Verszahl links mit führenden Nullen aufzufüllen, um sicherzustellen, daß eine Verszahlen aus gleichvielen Ziffern bestehen.

Nach der Benennung von Eingabe- und Ausgabedatei erscheint die Frage: "Sollen die Zeilenzahlen links mit Nullen aufgefüllt werden? (0 = nein; sonst Zahl zwischen 2 und 9 eingeben)". Anzugeben ist hier de facto, wieviele Stellen die Zählung insgesamt haben soll. Sollte die fortlaufende Zeilenzählung gestört sein – etwa infolge von Plus- oder Minusversen –, markiert das Programm die fraglichen Stellen in der Ausgabedatei mit dem Hinweis

* Störung in der Zeilenzählung.

Diese Stellen sind in der Ausgabedatei aufzusuchen und zu überprüfen, ob die Störung textbedingt, also "richtig", ist, oder ob hier ein Fehler in der Datei vorliegt. In diesem Falle ist die Störung manuell zu korrigieren. Sollte eine Neuzählung der gesamten restlichen Datei notwendig sein, löscht man am besten ab der Störung die gesamte fehlerhafte Zählung:

1. [Alt+F2] (= suchen und ersetzen)
2. "Mit Bestätigung?" "n"
3. Suche: "[FNZ]{*}[TAB]"
4. Ersetzen durch: "[FNZ][TAB]"
5. [F2]

Nach dem Speichern der Datei muß dann das Programm ZAHL ein weiteres Mal darauf angewendet werden, um die restliche Zählung zu ergänzen. Dabei ist darauf zu achten, daß keine weiteren Numerierungsfehler auftreten, da durch den Ersetzungsvorgang auch die sporadisch gesetzten Zeilenzahlen verlorengegangen sind, die dem Programm als Anhaltspunkt bei der Erzeugung der korrekten Zählung dienen. In diesem Falle ist also ein genauer Vergleich der Zählung mit der Vorlage oder der ursprünglichen Eingabedatei vonnöten.

18.21.2 Die Überprüfung einer (fortlaufenden) Numerierung

Neben der Erzeugung einer (fortlaufenden) Numerierung ist mit ZAHL auch die Überprüfung einer vorhandenen Numerierung möglich. Hierzu bietet ZAHL zwei Optionen:

4. Zeilen-Numerierung einer Textdatei
5. Ordnungsteil eines vertikal sortierten Indexes

Eine eventuelle Störung in der Zählung wird von dem Programm mit dem Hinweis

* Störung in der Zeilenzählung

markiert. Diese Stellen sind in der Ausgabedatei aufzusuchen und zu überprüfen, ob die Störung textbedingt, also "richtig", ist, oder ob hier ein Fehler in der Datei vorliegt, der dann zu korrigieren ist.

Sollten umfangreichere Störungen in der Zählung aufgetreten sein, empfiehlt es sich, die Eingabedatei entsprechend zu korrigieren und dann das Programm ZAHL erneut anzuwenden; alternativ ist es möglich, nur die fehlerhafte Zählung durch "Suchen und Ersetzen" zu tilgen und anschließend durch ZAHL neu zu erzeugen.

18.21.3 Vers- und Wortzahl mit Nullen auffüllen

Die dritte Programmoption ermöglicht es, den Ordnungsteil in seiner Länge zu vereinheitlichen.

19 Der WordPerfect-Editor

19.1 Grundlegende Funktionen

Datei öffnen:

1. [F5], [F5]
2. Verzeichnis auswählen mit [★], [✳], [✳], [✳]
3. in das ausgewählte Verzeichnis wechseln mit [Enter], [Enter]
4. Datei auswählen mit [★], [✳], [✳], [✳]

5. Datei laden mit [1]

Datei schließen:

1. [F7]
2. auf die Frage "Datei sichern?" antworten mit [J] (oder [N])
3. den vorgegebenen Pfad und Dateinamen mit [Enter] bestätigen (oder vorher ändern)
4. auf die Frage "überschreiben?" antworten mit [J] (oder [N])
5. auf die Frage "Editor verlassen?" antworten mit [J] oder [N]

Datei speichern:

1. [F10]
2. den vorgegebenen Pfad und Dateinamen mit [Enter] bestätigen (oder vorher ändern)
3. auf die Frage "überschreiben?" antworten mit [J] (oder [N])

Suchen:

1. [F2]
2. Suchbegriff eingeben
3. [F2]

Rückwärts suchen:

1. Shift+[F2]
2. Suchbegriff eingeben
3. [F2]

Suchen und Ersetzen

1. Alt+[F2]
2. Ersetzen mit Bestätigung [J], sonst [N]
3. Suchbegriff eingeben
4. [F2]
5. Ersetzungsbegriff eingeben
6. [F2] startet das Suchen und Ersetzen; ggf. Bestätigung notwendig

19.2 Tastaturlayouts

Zahlreiche Makros für häufig wiederkehrende Aufgaben (etwa Öffnen des Lemmaregisters, Einfügen von grammatischen Bestimmungen) sind über Tastaturbelegungen aufzurufen. Je nach Dateiformat und spezieller Aufgabe unterscheiden sich einige dieser Tastaturmakros in ihrer Programmierung; so gibt es etwa für die Vorlemmatisierungskorrektur, für die Ergänzung des Lemmatisierungswörterbuch und für die KTX-Korrektur am Bildschirm sehr ähnliche Makros, die immer mit Alt+Y aufgerufen werden sollen.

Um dies zu gewährleisten, muß mit dem Makro TASTATUR je nach Erfordernis zunächst das jeweils benötigte Tastaturlayout aktiviert werden:

1. ALT + F10
2. Eingabe "tastatur"
3. [ENTER]
4. Auswahl des Tastaturlayouts: 1 = Vorlemmatisierungskorrektur; 2 = Ergänzung des Lemmatisierungswörterbuchs; 3 = KTX-Korrektur am Bildschirm; 4 = Indexkorrektur und –auswertung; 5 = Texteingabe

19.2.1 Tastaturlayout "Vorlemmatisierungskorrektur"

Navigation

ALT+N	nächste Zeile mit ... \Wortform
ALT+V	vorige Zeile mit ... \Wortform

ALT+M	Vorschub um 20 Zeilen, Cursor bleibt in aktueller Zeile
ALT+P	Seitenvorschub: unterste Zeile des aktuellen Bildschirms wird oberste Zeile
ALT+✱	letzte Zeile mit Korrektureintrag
ALT+✱	nächste Zeile mit Korrektureintrag
ALT+A	Option 1/p: ermittelt die aktuelle Position und blendet eine entsprechende Meldung in die Statuszeile ein; danach können beliebige Arbeitsschritte einschließlich Makro-Anwendungen folgen. Option 2/z: kehrt zur Ausgangsposition zurück. Die Meldung in der Statuszeile kann nur mit Alt+A, 2 wieder beseitigt werden!

Lemma und flexionsmorphologische Bestimmung

ALT+L	Lemma im Lemmaregister suchen und gegebenenfalls übertragen
ALT+Y	Lemma im Lemmaregister suchen und gegebenenfalls übertragen
ALT+B	flexionsmorphologische Bestimmung suchen und gegebenenfalls übertragen

Textdatei (in Datei 3)

ALT+S	aus der VL-K-Datei heraus entsprechende Zeile der Textdatei suchen
ALT+K	aktuelle Zeile der Textdatei in den Header kopieren und mit F7 zurückkehren

Lexer und Lemmaregister

ALT+W	im Wörterbuch (Lexer) nachschauen
ALT+F	im Findebuch nachschauen
ALT+R	im Lemmaregister nachschauen

19.2.2 Tastaturlayout “Ergänzung des Lemmatisierungswörterbuchs”

Ergänzen des Lemmatisierungswörterbuchs

ALT+C	Kopieren aller Lemmatisierungslöcher in eine neue Datei (Datei 2!)
ALT+M	nächster Block von n (Default i = 5) identischen Löchern
ALT+G	Ändern der Zahl n für Makro ALT-M
ALT+L	Erzeugen eines neuen Lemmaeintrags für das Lemmatisierungswörterbuch in Datei 4 mit Hilfe des Lemmaregisters
ALT+Y	dasselbe
ALT+B	Flexionsmorphologische Bestimmung suchen und gegebenenfalls übertragen

“Löcherstopfen”

ALT+N	nächstes Lemmatisierungsloch
ALT+D	überträgt das Lemma der aktuellen Zeile auf die nächstfolgende Zeile
ALT+H	überträgt das passende Lemma aus einer vorausgehenden oder folgenden Zeile in die aktuelle Zeile

Lexer und Lemmaregister

ALT+W	im Wörterbuch (Lexer) nachschauen
ALT+F	im Findebuch nachschauen
ALT+R	im Lemmaregister nachschauen

19.2.3 Tastaturlayout “KTX-Korrektur am Bildschirm”

Navigation

ALT+N	Cursor an den Anfang des nächsten Vierzeilenblock der KTX-Datei
ALT+V	Cursor an den Anfang des vorigen Vierzeilenblock der KTX-Datei
ALT+M	partieller Seitenvorschub in der KTX-Datei: der Vierzeilenblock, in dem sich der Cursor gegenwärtig befindet, wird oberster Viererblock der nächsten Bildschirmseite; der Cursor befindet sich am Anfang des nun untersten Viererblocks
ALT+U	Bildschirmseite um einen Vierzeilenblock nach unten; Cursor steht am Anfang dieses Viererblocks

ALT+S	sucht die aktuelle Stelle der KTX-Datei im NF-Index in Datei 2
ALT+T	sucht die aktuelle Stelle der KTX- oder NF-Datei in der Textdatei in Datei 3

Lemma [und flexionsmorphologische Bestimmung]

ALT+Y	Lemma im Lemmaregister suchen und gegebenenfalls übertragen (die Registerdatei wird in Dat 4 geöffnet)
-------	--

Lexer und Lemmaregister

ALT+W	im Wörterbuch (Lexer) nachschauen
ALT+F	im Findebuch nachschauen
ALT+R	im Lemmaregister nachschauen

19.2.4 Tastaturlayout “Indexkorrektur und -auswertung”

Aufruf der Indexdateien

ALT+P	Sucht zu einer Belegliste des Programms SUCH-NF in Datei 1 die jeweiligen Textstellen in Datei 2. Die jeweils geöffnete Textdatei wird beim Verlassen nicht gespeichert
ALT+Q	Sucht zu einer Belegliste des Programms SUCH-NF in Datei 1 die jeweiligen Indexzeilen in Datei 2. Nach dem Starten des Makros kann der Bearbeiter entscheiden, ob die jeweils geöffneten Indexdateien beim Schließen auf Nachfrage, immer oder nie gespeichert werden sollen.
ALT+G	Sucht zu einer Belegliste des Programms SUCH-NF in Datei 1 die jeweiligen Textstellen in Datei 2. Die jeweils geöffnete Textdatei wird beim Verlassen nicht gespeichert.

Lexer und Lemmaregister

ALT+W	im Wörterbuch (Lexer) nachschauen
-------	-----------------------------------

19.2.5 Texteingabe

Dieses Tastaturlayout wird derzeit nicht benötigt und ist noch nicht erstellt worden.

19.3 Personalisierte Makros

Mit personalisierten Makros ist es möglich, die für die Vorlemmatisierungskorrektur oder die KTX-Korrektur am Bildschirm notwendigen Dateien zugleich zu starten.

Vorlage für das personalisierte Makro für die Vorlemmatisierungskorrektur ist das Makro NAME. Im Quellcode sind die entsprechenden Verzeichnis- und Dateinamen einzutragen; anschließend ist das Makro unter anderem Namen – sinnvollerweise unter dem Namen desjenigen, für den es eingerichtet ist – zu speichern.

Das personalisierte Makro für die KTX-Korrektur am Bildschirm wird mit dem Makro KTX-NEU erzeugt. Dieses Makro ist zu starten, nachdem zuerst die jeweilige Datei Y-KTX in Datei1, der Normalformindex Y-NF.0 in Datei 2 und die Textdatei X.TXT in Datei 3 geladen worden ist. Das Makro KTX-NEU erzeugt nun ein Makro KTX, das für folgende Arbeitssitzungen genau diese drei Dateien öffnen wird. Um versehentliches Überschreiben zu verhindern, sollte das Makro KTX unter anderem Namen – sinnvollerweise auch wieder unter dem Namen desjenigen, für den es eingerichtet ist – gespeichert werden.

19.4 Übersicht über die Editor-Makros

ALTA	siehe POS-MF und POSPAUSE
ALTB	siehe BEST-ERG und BEST-KAT
ALTC	siehe LOCH-KOP
ALTD	siehe LEM-DUP
ALTE	siehe O-EDITOR

ALTF	siehe FINDBUCH
ALTG	siehe N-AENDER; SUCH-SIN
ALTH	siehe LEM-HOL
ALTI	siehe INFO
ALTK	siehe ZIHEADER
ATL	siehe LEM-ERG und VL-K-KOR
ALTM	siehe LOECHER, WEITER-M und KTX-WEIT
ALTN	siehe WEITER-N und KTX-NVB
ALTO	siehe ORDNER
ALTOHLP	Fügt vor der aktuellen Zeile eine Leerzeile ein. Der Cursor wird anschließend in die Leerzeile gesteuert.
ALTP	siehe WEITER-P; SUCH-STT
ALTQ	siehe SUCH-NFI
ALTR	siehe LRG-SUCH
ALTS	siehe SU-VLK-T und KTX-SNF
ALTT	siehe KTX-STEL, STELLE-T, SU-VL-T
ALTU	siehe KTX-NEXT
ALTV	siehe WEITER-V und KTX-VVB
ALTW	siehe LEXER
ALTY	siehe LEM-ERG, KTX-LEM, KTX-LEMK
ALTZ	siehe WZAHL-KO
BEST-V2	Flexionsmorphologische Bestimmung suchen und gegebenenfalls übertragen. Aufruf über ALT+B (zu benutzen beim „Löcherstopfen“ in X-V2)
BEST-VK	Flexionsmorphologische Bestimmung suchen und gegebenenfalls übertragen. Aufruf über ALT+B (zu benutzen bei der VK-Korrektur von X-VK)
EN-AEND	Änderung der grammatischen Kategorie EN in PN, ON oder BN
FARBE	Festlegung der Bildschirmfarben
FINDBUCH	im Findebuch nachschauen. Aufruf über ALT+F
GRAPHO	Hilfe bei der Kontrolle und Beseitigung der vom Programm GRAPHO eingefügten Markierungen
INFO	Aufruf verschiedener Informationsdateien zu den wichtigsten Makros und zu einzelnen Arbeitsschritten der Vorlemmatisierung. Aufruf über ALT+I
KTX	Individuelles Makro, mit dem die drei zur KTX-Korrektur am Bildschirm notwendigen Datei geladen werden. Das Makro ist zunächst mit dem Makro MTX-NEU zu erzeugen. Es sollte ggf. individuell umbenannt werden.
KTX-LEM	Sucht bei der KTX-Korrektur am Bildschirm im Lemmaregister (in Datei 4) nach einem Lemma und überträgt es ggf.; Aufruf über ALT+Y
KTX-LEMK	Sucht bei der KTX-Korrektur am Bildschirm nach einer passenden Kategorie und grammatischen Bestimmung und überträgt sie ggf. Aufruf über ALT+Y
KTX-MDAT	von KTX-NEU benötigt; darf nicht verändert werden und ist daher schreibgeschützt
KTX-NEU	Zu benutzen bei der Bildschirmkorrektur von KTX-Dateien: Zunächst ist die Datei Y-KTX in Datei 1 zu laden, der Normalformindex Y-NF.0 in Datei 2, die Textdatei X.TXT in Datei 3. Anschließend kann mit dem Makro KTX-NEU ein neues Makro KTX erzeugt werden; bei späteren Arbeitssitzungen können die drei für die KTX-Korrektur benötigten Dateien direkt geladen werden.
KTX-NEXT	Blättert bei der KTX-Korrektur am Bildschirm die Bildschirmseite um einen Vierzeilenblock nach unten; Cursor steht am Anfang dieses Viererblocks. Aufruf über ALT+U
KTX-NVB	Steuert bei der KTX-Korrektur am Bildschirm den Cursor an den Anfang des nächsten Vierzeilenblock der KTX-Datei. Aufruf über ALT+N

KTX-SNF	Sucht bei der KTX-Korrektur am Bildschirm die aktuelle Stelle der KTX-Datei im Normalformindex in Datei 2. Aufruf über ALT+S
KTX-STEL	Sucht bei der KTX-Korrektur am Bildschirm die aktuelle Stelle der KTX-Datei in der Textdatei in Datei 3. Aufruf über ALT+T
KTX-VVB	Steuert bei der KTX-Korrektur am Bildschirm den Cursor an den Anfang des vorigen Vierzeilenblock der KTX-Datei. Aufruf über ALT+V
KTX-WEIT	partieller Seitenvorschub bei der KTX-Korrektur am Bildschirm: in der KTX-Datei wird der Vierzeilenblock, in dem sich der Cursor gegenwärtig, oberster Viererblock der nächsten Bildschirmseite; der Cursor befindet sich am Anfang des nun untersten Viererblocks. Aufruf über ALT+M
LEM-DUP	überträgt das Lemma der aktuellen Zeile auf die nächstfolgende Zeile. Aufruf über ALT+D
LEM-ERG	Erzeugen eines neuen Lemmaeintrags für das Lemmatisierungswörterbuch in Dat 4 mit Hilfe des Lemmaregisters. Zu Benutzen bei der Ergänzung des Lemmatisierungswörterbuchs. Aufruf über ALT-L oder ALT-Y
LEM-HOL	überträgt das passende Lemma aus einer vorausgehenden oder folgenden Zeile in die aktuelle Zeile
LEXER	Suche nach einem bestimmten Lemma im elektronischen Lexer. Aufruf über ALT+W
LIST-KOR	Bearbeitung von Ergebnis- oder Fehlerlisten aus NF-Indices: In Datei 1 wird eine Datei mit (kompletten) Belegzeilen aus verschiedenen NF-Indices geladen, in Datei 3 die Datei IND-LIST. Das Makro ermittelt dann in der Datei IND-LIST den Index, aus dem die aktuelle Zeile in Datei 1 stammt, lädt ihn in Datei 2 (wenn er dort nicht schon geladen ist) und sucht die betreffende Zeile. Wenn ein Index neu in Datei 2 geladen wird, sichert das Makro die dort befindliche Datei, bevor es den Bildschirm für den neuen Index leert.
LOCH-KOP	Kopieren aller Lemmatisierungslöcher in eine neue Datei (Dat 2!). Zu Benutzen bei der Ergänzung des Lemmatisierungswörterbuchs. Aufruf über ALT-C
LOECHER	nächster Block von n (Default i = 5) identischen Löchern. Zu Benutzen bei der Ergänzung des Lemmatisierungswörterbuchs. Aufruf über ALT-M
LRG-SUCH	Suche im Lemmaregister. Aufruf über ALT+R
N-AENDER	Ändern der Zahl n für Makro LOECHER. Zu Benutzen bei der Ergänzung des Lemmatisierungswörterbuchs. Aufruf über ALT-G
NAME	Aufruf von Arbeitsverzeichnis, Textdatei und VL.K-Index des jeweiligen Bearbeiters (dieses Makro muß mit Pfaden und Dateinamen angepaßt und unter anderem Namen gespeichert werden)
O-EDITOR	direkter Aufruf des Verzeichnisses c:\editor im Dateimanager
ORDNER	direkter Aufruf der am häufigsten gebrauchten Verzeichnisse. Aufruf über ALT+O
POS-MF	Option 1/p: ermittelt die aktuelle Position u. blendet eine entsprechende Meldung in die Statuszeile ein; danach können beliebige Arbeitsschritte einschließlich Makro-Anwendungen folgen. Option 2/z: kehrt zur Ausgangsposition zurück. Aufruf über ALT+A; die Meldung in der Statuszeile kann nur mit Alt+A, 2 wieder beseitigt werden
POSPAUSE	ähnlich wie POS-MF
SICHERN	ermöglicht die Festlegung, ob bei der Bearbeitung von Indices mit dem Makro SUCH-NFI die geöffneten Indices auf Nachfrage, immer oder nie gespeichert werden sollen.
STELLE-T	Sucht bei der Vorlemmatisierungskorrektur die aktuelle Stelle in der Textdatei in Datei 3. Aufruf über ALT+T
SU	kopiert Zeilen, auf die ein zuvor definierter Suchbegriff zutrifft, ohne Bestätigung von Datei 1 in Datei 2 (mit Bestätigung siehe SU-B)

SU-B	kopiert Zeilen, auf die ein zuvor definierter Suchbegriff zutrifft, nach Bestätigung von Datei 1 in Datei 2 (ohne Bestätigung siehe SU)
SU-VL-T	zeigt bei der Vorlemmatisierungskorrektur die entsprechende Zeile der Textdatei an. Bei erstmaligem Aufruf des Makros muß der Pfad und der Dateiname der Textdatei angegeben werden (Aufruf im Tastaturlayout "Vorlemmatisierungskorrektur" über ALT+T)
SU-VLK-T	ruft aus der VL-K-Datei heraus die entsprechende Zeile der Textdatei, wenn diese in Datei 3 geladen ist; zu benutzen bei der Vorlemmatisierungskorrektur; Aufruf über ALT+S
SUCH-NFI	sucht zu einer Belegliste, die das Programm SUCH-NF erzeugt hat (in Datei 1), die entsprechenden Stellen in den betreffenden Indices (in Datei 2). Nach dem Starten des Makros kann der Bearbeiter entscheiden, ob die jeweils geöffneten Indexdateien beim Schließen auf Nachfrage, immer oder nie gespeichert werden sollen (Änderung der ausgewählten Speicher-Option über das Makro SICHERN). Aufruf im Tastaturlayout "Indexkorrektur und -auswertung" über ALT+Q.
SUCH-SIN	sucht zu einer Belegliste, die das Programm SUCH-NF erzeugt hat (in Datei 1), die entsprechenden Stellen in den betreffenden Indices (in Datei 2). Die jeweils geöffnete Indexdatei wird beim Verlassen <u>nicht gespeichert</u> (siehe alternativ SUCH-NFI). Aufruf im Tastaturlayout "Indexkorrektur und -auswertung" über ALT+G.
SUCH-STI	sucht zu einer Belegliste, die das Programm SUCH-NF erzeugt hat (in Datei 1), die entsprechenden Stellen in den betreffenden Indices (in Datei 2). Die jeweils geöffnete Indexdatei wird beim Verlassen <u>nicht gespeichert</u> (vgl. SUCH-SIN; siehe alternativ SUCH-NFI).
SUCH-STT	sucht zu einer Belegliste, die das Programm SUCH-NF erzeugt hat (in Datei 1), die entsprechenden Stellen in den betreffenden Textdateien (in Datei 2). Die jeweils geöffnete Textdatei wird beim Verlassen nicht gespeichert! Aufruf im Tastaturlayout "Indexkorrektur und -auswertung" über ALT+P.
SUV	verschiebt Zeilen, auf die ein zuvor definierter Suchbegriff zutrifft, ohne Bestätigung von Datei 1 in Datei 2 (mit Bestätigung siehe SUV-B)
SUV-B	verschiebt Zeilen, auf die ein zuvor definierter Suchbegriff zutrifft, nach Bestätigung von Datei 1 in Datei 2 (ohne Bestätigung siehe SUV)
SZ-INFO	Information über die aktuelle Belegung der Tasten ALT+0 bis ALT+9
TASTATUR	ermöglicht die Auswahl und Aktivierung von fünf verschiedenen Tastaturlayouts: 1 = "Vorlemmatisierungskorrektur"; 2 = "Ergänzung des Lemmatisierungswörterbuchs"; 3 = "KTX-Korrektur am Bildschirm"; 4 = "Indexkorrektur u. -auswertung"; 5 = "Texteingabe"
VL-K-KOR	Erzeugen eines neuen Lemmaeintrags für das Lemmatisierungswörterbuch in Dat 4 mit Hilfe des Lemmaregisters. Zu Benutzen bei der Ergänzung des Lemmatisierungswörterbuchs. Aufruf über ALT+L oder ALT+Y
WEITER-M	Seitenvorschub um 20 Zeilen, Cursor bleibt in aktueller Zeile. Zu benutzen beim Löcherstopfen und der Vorlemmatisierungskorrektur. Aufruf über ALT+M
WEITER-N	nächste Zeile mit einer zu bestimmenden Wortform; zu benutzen bei der Vorlemmatisierungskorrektur; Aufruf über ALT+N
WEITER-P	Seitenvorschub: die unterste Zeile des aktuellen Bildschirms wird oberste Zeile. Aufruf über ALT+P
WEITER-V	vorige Zeile mit einer zu bestimmenden Wortform; zu benutzen bei der Vorlemmatisierungskorrektur; Aufruf über ALT+V
WZAHL-KO	Korrektur der Wortzählung in Indices. Keinesfalls anzuwenden beim "Löcherstopfen"! Aufruf über ALT+Z
ZAHL	zählt die Zeilen, auf die ein zuvor definierter Suchbegriff zutrifft

ZIHEADER	aktuelle Zeile der Textdatei in den Header kopieren; Rückkehr mit F7. Aufruf über ALT+K
?	Cursor wird in die letzte Zeile mit einem Korrektureintrag gesteuert; zu benutzen bei der Vorlemmatisierungskorrektur; Aufruf über Alt+[Pfeil nach oben]
?	Cursor wird in die nächste Zeile mit einem Korrektureintrag gesteuert; zu benutzen bei der Vorlemmatisierungskorrektur; Aufruf über Alt+[Pfeil nach oben]

20 Literatur

Die im folgenden genannte Literatur ist eine mehr oder wenige zufällige Auswahl und keineswegs als auch nur halbwegs systematische oder vollständige Bibliographie zu verstehen!

- Klein, Thomas: Rheinische und westfälische Sprachgeschichte bis 1300. In: Jürgen Macha u. a. (Hgg.): Rheinische und westfälische Sprachgeschichte. Köln, Weimar, Wien 2000, 3–48.
- Klein, Thomas: Die mittelfränkische e-Synkope und -Apokope und die Metrik des "Rheinischen Marienlobs". In: Kulturen – Sprachen – Übergänge. Festschrift für H. L. Cox zum 65. Geburtstag. Hg. von Gunter Hirschfelder, Dorothea Schell, Adelheid Schruttka-Rechtenstamm. Köln, Weimar, Wien 2000, S. 153–171.
- Klein, Thomas: *he, her, er – de, der*. Zu den *r*-Pronomina im Mitteldeutschen. In: Sprache und Literatur des Mittelalters in den *nideren landen*. Gedenkschrift für Hartmut Beckers. Hg. von Volker Honemann u. a. Köln, Weimar, Wien 1999, S. 141–155 (Niederdeutsche Studien 44).
- Klein, Thomas: Vom lemmatisierten Index zur Grammatik. In: Maschinelle Verarbeitung altdeutscher Texte. V. Kolloquium Würzburg 1997. Hg. von W. Wegstein u. Norbert Richard Wolf. ?–?.
- Klein, Thomas: Vers und Syntax in frühmittelhochdeutscher Dichtung. Wege zur Datengewinnung und -auswertung. In: Deutsche Sprache in Raum und Zeit. Festschrift für Peter Wiesinger. Hg. von P. Ernst, F. Patocka. Wien 1998, S. 537–568.
- Klein, Thomas: Zur Frage der Korpusbildung und zur computerunterstützten grammatischen Auswertung mittelhochdeutscher Quellen. In: Klaus-Peter Wegera (Hg.): Mittelhochdeutsche Grammatik als Aufgabe. ZfdPh 110 (1991), Sonderheft, S. 3–23.
- Solms, Hans-Joachim u. Wegera, Klaus-Peter: Morphologie des Frühneuhochdeutschen. In: Sprachgeschichte. Ein Handbuch ... Hg. von Werner Besch u. a. Berlin, New York, 2. Aufl. 2000, S. ?–?.
- Solms, Hans-Joachim: Das System der Präfixverben in der frühesten Überlieferung des Hartmannschen 'Gregorius' (Hs. A aus dem Alemannischen des 13. Jahrhunderts). In: Werner Besch (Hg.): Deutsche Sprachgeschichte. Grundlagen, Methoden, Perspektiven. Festschrift für Johannes Erben zum 65. Geburtstag. Bern, New York, Frankfurt a. M., S.115–128.
- Solms, Hans-Joachim: Zur westmitteldeutschen "Wechselflexion" bei den mhd. starken Verben der Klassen III–V. In: ders., Sheila Watts, Jonathan West (Hgg.): Das Germanische Verb. Tübingen ? (Linguistische Arbeiten ?), S. ?–?.
- Solms, Hans-Joachim: Historische Wortbildung. In: Sprachgeschichte. Ein Handbuch ... Hg. von Werner Besch u. a., Berlin, New York, 2. Aufl. 1998, 1. Teilband, S. 596–610.
- Solms, Hans-Joachim: Zur Wortbildung der Verben in Hartmann von Aues 'Iwein' (Hs.B) und 'Gregorius' (Hs.A): Das Präfix *ge-* im System der verbalen Präfigierung. Zugleich ein Beitrag zur Diskussion historischer Wortbildung. In: ZfdPh 110 (1991), Sonderheft, S.110–140.
- Wegera, Klaus-Peter: "Gen, oder wie Herr Gottsched will, chen." Zur Geschichte eines Diminutivsuffixes. In: Wortschatz und Orthographie in Geschichte und Gegenwart. Festschrift für Horst Haider Munske. Tübingen 2000, S. 43–58.
- Wegera, Klaus-Peter: Mittelhochdeutsche Wortbildung. Zum Stand ihrer Erforschung, dargestellt am Beispiel der Diminutive. In: Tagungsbericht der Fachtagung "Historische Wortbildung des Deutschen" vom 12. bis 14. Oktober 2000, S. ?–?.

- Wegera, Klaus-Peter: Grundlagenprobleme einer neuen mittelhochdeutschen Grammatik. In: Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Hg. von Werner Besch u. a. Bd. 2, 2. vollst. neu bearb. u. erw. Aufl. Berlin, New York 2000 (Handbücher zur Sprach- und Kommunikationswissenschaft 2), S. 1304-1320.
- Wegera, Klaus-Peter: Rez. von: Hermann Paul, Mittelhochdeutsche Grammatik. 23. Aufl. neu bearb. von Peter Wiehl und Siegfried Grosse. Tübingen 1989. In: PBB 113 (1991), S. 275–279.
- Wegera, Klaus-Peter: Mittelhochdeutsche Grammatik und Sprachgeschichte. In: Deutsche Sprachgeschichte. Grundlagen, Methoden, Perspektiven. In: Festschrift für Johannes Erben. Hg. von Werner Besch. Frankfurt a. M. u. a. 1990, S. 103–113.
- Wegera, Klaus-Peter: Wortbildung des Frühneuhochdeutschen. In: Sprachgeschichte. Ein Handbuch ... Hg. von Werner Besch u. a. 2. Halbbd. Berlin 1985, S. 1348–1356; in der 2. Aufl. zusammen mit Heinz-Peter Prell, 2. Halbbd., Berlin 2000, S. ?–?.
- Mittelhochdeutsche Grammatik als Aufgabe. Hg. von Klaus-Peter Wegera (ZfdPh 110 [1991] Sonderheft).

21 Anhang

21.1 Die Projektarbeitsstellen

21.1.1 Bochum

Ruhr-Universität Bochum
Germanistisches Institut
Lehrstuhl Altgermanistik 1
Universitätsstr. 150
GB 4/132
44780 Bochum

Leiter:

Prof. Dr. Klaus-Peter Wegera
E-Mail: Klaus-Peter.Wegera@ruhr-uni-bochum.de
Homepage: <http://www.ruhr-uni-bochum.de/wegera/>
Telefon: 0234-3222574
Fax: 0234-3214254

21.1.2 Bonn

Rheinische Friedrich-Wilhelms-Universität
Germanistisches Seminar
Ältere Abteilung
Lehrstuhl Prof. Dr. Th. Klein
Am Hof 1 d
53113 Bonn

Leiter:

Prof. Dr. Thomas Klein
E-Mail: thomas.klein@uni-bonn.de
Telefon: 0228-737712 (Sekretariat 0228-733974)
Fax: 0228-73-4517

21.1.3 Halle

Martin-Luther-Universität Halle-Wittenberg
Germanistisches Institut, Altgermanistik

Herweghstr. 96

06099 Halle/S.

Telefon Sekretariat (Frau Drosdziok): 0345-5523599

Telefon Mitarbeiter: 0345-5523596

Telefon Hilfskräfte: 0345-5523616

Fax: 0345-5527067

Projekthomepage: http://www.germanistik.uni-halle.de/altgermanistik/ab_ag_for2.htm

Leiter:

Prof. Dr. Hans-Joachim Solms

E-Mail: solms@germanistik.uni-halle.de

21.1.4 Jena

Friedrich-Schiller-Universität Jena

Philosophische Fakultät

Institut für Germanistische Sprachwissenschaft

D-07740 Jena

Fürstengraben 30

D-07743 Jena

Leiter:

Prof. Dr. Eckhard Meineke

E-Mail: eckhard.meineke@uni-jena.de

Homepage: <http://www2.uni-jena.de/philosophie/germsprach/Sprachgesch.html>

Telefon: 03641-944340 (Sekretariat: 03641-944310)

Fax: 03641-944342

21.1.5 Oslo

Universitet i Oslo

Germanistisk Institutt

Postboks 1003 – Blindern

NO-0315 Oslo

Niels Treschows hus

7. etg, rom NT 623

Niels Henrik Abels vei 36

NO-0851 Oslo

Leiter:

Professor Dr. Heinz-Peter Prell

E-Mail: h.p.prell@ilos.uio.no

Homepage: <http://www.hf.uio.no/ilos/om-instituttet/ansatte/vit/hpprell.xml>

Telefon: 0047-22- 856903

Fax: 0047-22-856887

21.2 Das Korpus

21.2.1 Das Hauptkorpus

NB: In eckigen Klammern werden jeweils die Texttypen genannt, denen die einzelnen Texte zugeordnet werden können.

0 Oberdeutsch							
	Prosa 1	Prosa 2	Prosa 3	Vers 1	Vers 2	Vers 3	Vers 4

I	1050-1150	Williram, Auslegung des Hohenlieds [BK]	Wiener Notker [BK]	Bamberger Glaube u. Beichte [Th]	Ezzolied [Th G]	Merigarto [S]	Rheinauer Paulus [G]	
II	1150-1200	siehe 1 Bairisch, 2 Schwäbisch / Alemannisch-Bairisch, 3 Alemannisch						
III	1200-1250				Wolfram von Eschenbach: Parzival [E]	Hartmann von Aue: Iwein [E]	Gottfried von Straßburg: Tristan [E]	Nibelungenlied [E]
IV	1250-1300	siehe 1 Bairisch, 2 Schwäbisch / Alemannisch-Bairisch, 3 Alemannisch						
V	1300-1350							

1 Bairisch						
		Prosa 1	Prosa 2	Vers 1	Vers 2	Urkunden 1
I	1050-1150	siehe 0 Oberdeutsch				
II	1150-1200	<i>Physiologus</i> [S Th]	Windberger Psalter [Bi]	Kaiserchronik [H E S]	<i>Heinrichs Litanei</i> [G]	
III	1200-1250	Millstätter Predigten [Pr]	<i>St. Pauler Predigten</i> [Pr]	Bruder Wernher: Marienleben [E L]	<i>Stricker: Karl d. Gr. D</i> [E] [vorher: <i>Die Hochzeit</i>] [Th]	
IV	1250-1300	Buch der Könige [H Bi]	<i>Bartholomäus</i> [S]	Dietrichs Flucht [E]	<i>Ulrich von Liechtenstein: Frauen dienst</i> [E]	
V	1300-1350	Ruprecht von Freising: Rechtsbuch [R]	<i>Oberaltaicher Evangelistar</i> [Bi]	Johann von Würzburg: Wilhelm von Österreich [E]	<i>Rabenschlacht W</i> [E] [vorher: <i>Maria Magdalena</i>] [E L]	Landshut [U]

2 Schwäbisch / Alemannisch-Bairisch						
		Prosa 1	Prosa 2	Vers 1	Vers 2	Urkunden 1
I	1050-1150	siehe 0 Oberdeutsch				
II	1150-1200	<i>Speculum ecclesiae</i> [Pr]	<i>Wien-Münchener Evangelien</i> [Bi]	Deutung der Meßgebräuche [Th]		
III	1200-1250	<i>Zwifaltener Benediktinerregel</i> [R Th]	Hoffmannsche Predigten [Pr]	[<i>Albert v. Augsburg: Ulrichsleben?</i>] [E L]		
IV	1250-1300	David von Augsburg: Traktate [Th]	<i>Augsburger Stadtbuch</i> [R]	Winsbeke und Winsbekin [S?]		<i>Augsburg</i> [U]
V	1300-1350	Baumgarten [Th My]	[<i>Augsburger Franziskanerregel?</i>] [R Th]	Ulrich von Türrheim: Rennewart [E]	<i>Hartwig von dem Hage: Margaretha</i> [E L]	Augsburg [U]

3 Alemannisch						
---------------	--	--	--	--	--	--

		Prosa 1	Prosa 2	Vers 1	Vers 2	Urkunden 1
I	1050-1150	siehe 0 Oberdeutsch				
II	1150-1200	Züricher Predigten [Pr]	<i>Gebete und Benediktionen aus Muri</i> [G]	Linzer Entechrist [Th L?]	Scoph von dem lone [Th]	
III	1200-1250	St. Trudperter Hohelied [BK]	<i>Lucidarius</i> [S]	Frauenfelder "Flo-re" [E]	Gottfried von Straßburg: Tristan (f) [E]	
IV	1250-1300	Schwabenspiegel [R]	<i>Schwarzwälder Predigten</i> [Pr]	Rudolf von Ems: Weltchronik [H E S]	<i>Rudolf von Ems: Wilhelm</i> [E]	Freiburg [U]
V	1300-1350	<i>Summa theologica</i> [Th] [vorher: St. Georgener Predigten] [Pr]	Nikolaus von Straßburg: Predigten [Pr]	Rappoltsteiner "Parzival" [E]	<i>Hugo von Langenstein: Martina</i> [L E]	Freiburg [U]

4 Westmitteldeutsch						
		Prosa 1	Vers 1	Vers 2	Urkunden 1	
I	1050-1150					
II	1150-1200	Trierer Psalmen [Bi]	Arnsteiner Marienlied [G]	Mittelfrk. Reimbibel [Bi]		
III	1200-1250	Vatikanische Gebete [G]	Rheinisches Marienlob [G L Th]	Tundalus [L]		
IV	1250-1300	siehe 4a Mittelfränkisch / 4b Rheinfränkisch-Hessisch				
V	1300-1350					

4a Mittelfränkisch						
		Prosa 1	Prosa 2	Vers 1	Vers 2	Urkunden 1
I	1050-1150	siehe 4 Westmitteldeutsch				
II	1150-1200					
III	1200-1250					
IV	1250-1300	Die Lilie (Prosa) [Th My]	[vorher: Amtleut- buch von St. Brigi- den] [R]	Die Lilie (Vers) [Th My]	<i>Karl und Galie</i> [E]	<i>Köln</i> [U]
V	1300-1350	Tauler: Predigten [Pr My]	<i>Buch der Minne</i> [Th My]	Yolande von Vi- anden [L E]	<i>Schlacht bei Göll- heim</i> [E H]	Köln [U]

4b Rheinfränkisch-Hessisch							
		Prosa 1	Prosa 2	Vers 1	Vers 2	Vers 3	Urkunden 1
I	1050-1150						
II	1150-1200	siehe 4 Westmitteldeutsch					

III	1200-1250						
IV	1250-1300	Salomons Haus [Th My]	[vorher: Mitteldeutsche Predigten] [Pr L]	Rhfrk. Marien Himmelfahrt [E L]			
V	1300-1350	Oxfordener Benediktinerregel [R Th]	Hermann von Fritzlar: Heiligenleben [L]	Leben der hl. Elisabeth [L E]	Die Erlösung [Th Bi]	Rheinfrk. Reimpredigten [Th Pr]	Mainz [U]

5 Ostmitteldeutsch / Hessisch-Thüringisch						
		Prosa 1	Prosa 2	Vers 1	Vers 2	Urkunden 1
I	1050-1150					
II	1150-1200	Frankfurter Predigtfragmente [Pr L]		Trierer Ägidius [E L]		
III	1200-1250	Mitteldeutsche Predigten J [Pr L]		Graf Rudolf [E]		
IV	1250-1300	Mühlhäuser Reichrechtbuch [R]	Jenaer Martyrologium [L]	Athis und Prophetias [E] [vorher: Passional] [E L]		
V	1300-1350	Berliner Evangelistar [Bi]	Matthias Beheim: Evangelienbuch [Bi]	Landgraf Ludwigs Kreuzfahrt [H E]	Heinrich von Freiberg: Tristan [E]	Jena-Weida [U]

6 Ostfränkisch							
		Prosa 1	Prosa 2	Prosa 3	Vers 1	Vers 2	Urkunden 1
I	1050-1150						
II	1150-1200						
III	1200-1250						
IV	1250-1300						
V	1300-1350	Christine Ebner: Von der gnaden überlast [Th My]	Satzungsbuch Nürnberg [R]	Würzburger Polizeisätze [R]	Hugo von Trimberg: Der Renner [E S]	Lupold Hornburg: Reden [Pr]	Nürnberg [U]

21.2.2 Das Erweiterungskorpus

NB: Diese Übersicht ist derzeit (Juli 2004) nicht auf dem neuesten Stand!

0 Oberdeutsch						
	Prosa 4 / 5	Prosa 6 / 7	Prosa 8 / 9	Prosa 10 / 11	Prosa 12 / 13	Prosa 14 / V. 5

I	1050-1150	Williram: Auslegung des Hohelieds (diverse Hss.)	Otlohs Gebet	Wessobrunner Predigtslg. B	St. Galler Glaube und Beichte I	St. Galler Glaube und Beichte III	Rhfrk. Interlinear-version der Psalmen
		Alkuins Traktat	Wessobrunner Glaube und Beichte	Älterer Physiologus	St. Galler Glaube und Beichte II	Arzneibuch Ippokrates	
II	1150-1200	siehe 1 Bairisch, 2 Schwäbisch / Alemannisch-Bairisch, 3 Alemannisch					
III	1200-1250						V. 5: Wirnt von Gravenberg: Wigalois
IV	1250-1300	siehe 1 Bairisch, 2 Schwäbisch / Alemannisch-Bairisch, 3 Alemannisch					
V	1300-1350						

1 Bairisch							
		Prosa 3 / 4	Prosa 5 / 6	Prosa 8 / 9	Vers 3 / 4	Vers 5 / 6	Urkunden
I	1050-1150	siehe 0 Oberdeutsch					
II	1150-1200	Kleintexte Sammelhs. Innsbruck 652	Lamprechter Gebete		Rolandslid	Wiener Genesis	
		Münchener Predigtfragmente			Frau Ava: Leben Jesu	Priester Adelbrecht	
III	1200-1250	Lilienfelder Heilsgeschichte	Grieshabersche Predigten I	Rothsche Predigtslg.	Der Stricker: Karl der Große		
		St. Lambrecht Gebete	Millstätter Psalter	Leysersche Predigten			
IV	1250-1300	Raitenhaslach Ben.-Regel	Admonter Ben.-Regel		Gundacker von Judenburg: Christi Hort	Thomasin von Zerklare: Der Welsche Gast	?
		Mettener Predigtslg. II	Regensburger Schwabenspiegel		Die Warnung	Neidhart R	
V	1300-1350	Stundenbuch ÖNB 2745	Gebetbuch cgm 101	Leipziger Predigten E	Rabenschlacht	Bruder Philipp: Marienleben	2: Niederbayern; 3: Brixen 4: Regensburg
		Bair. Predigtslg.	Gebetbuch cgm 73		Albrecht: Jüngerer Titurel		

2 Schwäbisch / Alemannisch-Bairisch						
		Prosa 3	Prosa 4	Prosa 5	Vers 3	Urkunden
I	1050-1150	siehe 0 Oberdeutsch				

II	1150-1200	Schwäbische Trauformel				
III	1200-1250	Benediktbeurer Ratschläge und Gebete			Albert von Augsburg: Ulrichsleben	
IV	1250-1300	Priester Konrad: Predigtbuch (W1)	Augsburger Klarsenregel	Schwabenspiegel	Rudolf von Ems: Barlaam und Josaphat	
V	1300-1350	<i>Augsburger Achtbuch?</i>	<i>Augsburger Schwabenspiegel?</i>	Augsburger Drittor-densregel		

3 Alemannisch						
		Prosa 3 / 4	Prosa 5 / 6	Vers 3 / 4	Vers 5 / 6	Urkunden
I	1050-1150	siehe 0 Oberdeutsch				
II	1150-1200	Rheinauer Gebete		Cantilena de Conversione S. Pauli		
		Engelberger Gebete				
III	1200-1250	Baseler Predigtfragment		Reinhart Fuchs		
		Alem. Predigtbruchstücke		Vorauer Novelle		
IV	1250-1300	Hugo von St. Victor: Expositio	Hugo von Konstanz: Predigten	Liederhandschrift D		?
		Engelberger Ben.-Regel				
V	1300-1350	<i>Statutenbuch Hagenau?</i>	Klosterneuburger Evangelienwerk	Speculum humanae salvationis	Sigenot	2: Freiburg II 3: St. Blasien I
		<i>Schwabenspiegel?</i>	Faszikel I aus Zürich, ZB, Rh 99b	Leben der hl. Caecilia		4: St. Blasien II 5: Ravensburg 6: St. Gallen 7: Straßburg 8: Zürich

4 Westmitteldeutsch						
		Prosa 2 / 3	Prosa 4 / Vers 3	Vers 4 / 5	Vers 6 / 7	Urkunden
I	1050-1150					
II	1150-1200	Berliner Ägidius	Uppsalaer Sündenklage			
		Hildegart- Gebetbuch	Pfaffe Lambrecht: Tobias			
III	1200-1250	Md. Predigten H ₂		Von Christi Geburt	Friedberger Christ	
			Idsteiner Sprüche der Väter	Albanus	Herzog Ernst A	
IV	1250-1300	siehe 4a Mittelfränkisch / 4b Rheinfränkisch-Hessisch				

V	1300-1350	
---	-----------	--

4a Mittelfränkisch						
		Prosa 3	Prosa 4	Vers 3 / 4	Vers 5 / 6	Urkunden
I	1050-1150	siehe 4 Westmitteldeutsch				
II	1150-1200					
III	1200-1250					
IV	1250-1300	<i>Deutschordensstatuten?</i>	<i>Heilige Regel?</i>	<i>Freidank?</i> <i>Karl und Galie B?</i>		
V	1300-1350	<i>Deutschordensregeln u. statuten?</i>	Aachener Ratsentscheid	Rhein. Merlin Gottfried Hagen: Reimchronik D	Morant und Galie Turnier, Ritterfahrt, Ritterpreis	2: Kölner Eidbücher

4b Rheinfränkisch-Hessisch							
		Prosa 3	Prosa 4	Prosa 5	Vers 4	Vers 5	Urkunden
I	1050-1150	siehe 4 Westmitteldeutsch					
II	1150-1200						
III	1200-1250						
IV	1250-1300	Md. Predigten Ka			Freidank A	Der Sünden Widerstreit	2: Mainz II
V	1300-1350	<i>Nikolaus von Landau: Predigten?</i>	Darmstädter Legendar	<i>Klagschrift der Gesellschaft ...?</i>	Frankfurter Dirigierrolle	St. Galler Passionsspiel	

5 Ostmitteldeutsch						
		Prosa 3	Prosa 4	Prosa 5	Vers 3	Urkunden
I	1050-1150					
II	1150-1200					
III	1200-1250	Sedulius: Carmen paschale	Md. Predigten W		Kaiserchronik H	
IV	1250-1300	Arzneibuch Bartholomäus; Macer			Wolfram von Eschenbach: Willehalm	?
V	1300-1350	Leipziger Predigtsg. A	<i>Zwickauer Rechtsbuch?</i>	<i>Freiberger Rechtsbuch?</i>	Tilo von Kulm: Von siben ingesigeln	2: Erfurt I 3: Erfurt II 4: Weida 5: Frauensee

6 Ostfränkisch						
		Prosa 4	Prosa 5	Vers 3	Vers 4	Urkunden
I	1050-1150					
II	1150-1200					
III	1200-1250					
IV	1250-1300					
V	1300-1350	<i>Buoch von guoter spise?</i>	Nürnberger Stadt- buch			2: Nürnberg II

21.3 Header

21.3.1 Der Header für Textdateien

+H

Korpus-Sigle:

Dateiname:

Datum:

Film-Nummer:

Textart:

Kurztitel:

Bibliothek/Archiv:

Signatur:

Auswahl:

Umfang in Wortformen:

Datierung:

Lokalisierung:

Edition:

Literatur:

Sonstiges:

vorhandener Text / verfügbarer Text:

Text eingegeben nach der Handschrift:

Datum:

Bearbeiter:

Text eingegeben nach der Edition:

Datum:

Bearbeiter:

Text kollationiert:

Datum:

Bearbeiter:

Text präeditiert:

Datum:

Bearbeiter:

Text überprüft:

Datum:

Bearbeiter:

Sonstiges:

Fehler/Änderungen: ⁹⁴

⁹⁴ Hier werden alle Änderungen am Text dokumentiert, die ab dem oben angegebenen Datei-Datum vorgenommen wurden/werden sollen.

bereits korrigiert / geändert:
noch zu korrigieren / ändern:
@H

21.3.2 Der Header für Indices

+H
Textsigle:
Kurtztitel:

Verwendete Textdatei:
Datum der Datei:

Harmonisierung mit Textdatei:
Datum der Datei:
Statusfeld geändert:

Lemmatisierung
Bearbeiter:
Datum:

KTX_Text
Korrektur 1:
Datum:
Korrektur 2:
Datum:
Korrekturen eingetragen:
Datum:

Fehler in diesem Index
Bereits korrigiert:
Noch zu korrigieren:

@H

21.4 Annotierungsfahrplan: Von der Textdatei zum N-Index

(siehe ... = Verweise auf das Korpushandbuch)

0. ggf. Dateikonvertierung Word → TXT (siehe 8.1)

1. Erzeugung von X-W.txt (siehe 8.2)

<u>Programm:</u>	LEM, Option 1
<u>Eingabe-Datei:</u>	X-T.TXT
<u>Ausgabe-Datei:</u>	X-W.TXT

2. Erzeugung von X-V1.txt (siehe 8.4)

<u>Programm:</u>	LEM-2, Option 2
<u>Eingabe-Datei:</u>	X-W.txt
<u>Ausgabe-Datei:</u>	X-V1.txt

Ergänzung von X-W um eine normalisierte Vergleichsform, nach der von LEM-3 im Lemmatisierungswörterbuch gesucht wird.

3. Alphabetische Sortierung von X-V1 zu X-V1S (siehe 8.4)

<u>Programm:</u>	SORTA
------------------	-------

Eingabe-Datei: X-V1.txt
Ausgabe-Datei: X-V1S.txt

4. Vorlemmatisierung von X-V1S zu X-V2 (siehe 8.4)

Programm: LEM-3
Eingabe-Datei: X-V1S.txt
Ausgabe-Datei: X-V2.txt

[5. Verbesserung der automatischen Vorlemmatisierung (siehe 8.5)]

6. "Löcherstopfen" (siehe 8.6)

Wortformen, denen aus dem Lemmatisierungswörterbuch kein Annotationsangebot durch LEM-3 zugewiesen werden könnte, müssen händisch lemmatisiert werden. Dafür benutzte Makros: Alt+L (= LEM-V2.edm) und Alt+B (=BEST-V2.edm).

!!! Vorher müssen durch Anwendung des Makros TASTATUR, Option 1, die richtigen Alt+X-Makros installiert werden. !!!

7. Vertikale Sortierung von X-V2 zu X-VZ (siehe 8.7)

Programm: SORTA
Eingabe-Datei: X-V2.txt
Ausgabe-Datei: X-VZ.txt
Sortierreihenfolge: 2

8. Verbesserung von X-VZ zu X-V (siehe 8.7)

Programm: LEM-4
Eingabe-Datei: X-VZ.txt
Ausgabe-Datei: X-V.txt

LEM-4 versucht aufgrund von Merkmalen der mutmaßlichen syntaktischen Umgebung, bei mehreren angebotenen Lemmata das richtige Lemma an die erste Position zu rücken, und entsprechend auch im Flexionsangebot die richtige Flexionsbestimmung an die erste Stelle zu setzen.

In der Ausgabedatei doppelt LEM-4 das Angebot: voran steht nun ein „Trema“-Bereich, in dem die Lemmata mit einem Trema ` (teils auch als • dargestellt) beginnen; durch einen Tab und % abgetrennt folgt der „Gravis“-Bereich, in dem die Lemmata mit einem Gravis ` beginnen. Der Gravis-Bereich ist zunächst ein genaues Spiegelbild des Trema-Bereichs, nur dass ihm die Flexionsangebote fehlen, z. B.

```
{1}`ge-mach #mn ^NS/AS/NP/AP/DS {2}`ge-mach #adj ^-/0 % {1}`ge-mach @mn ^ {2}`ge-mach @adj  
^
```

Während das Trema-Lemma und seine Wortart belegspezifisch angepasst werden können, dürfen das Gravis-Lemma und seine Wortart grundsätzlich nicht verändert werden! Das Gravis-Lemma kann lediglich zusammen mit dem Trema-Lemma durch ein anderes Lemma des Lemmaregisters ersetzt werden.

Beispiel einer belegspezifischen Anpassung des #-Genus (Kontext: *daz gemacht*):

```
{1}`ge-mach #n ^NS/AS/NP/AP/DS {2}`ge-mach #adj ^-/0 % {1}`ge-mach @mn ^ {2}`ge-mach @adj  
^
```

9. Umformatierung von X-V zu X-VK (siehe 8.7)

Programm: LEM, Option 5
Eingabe-Datei: X-V.txt
Ausgabe-Datei: X-VK.txt

10. Korrektur von X-VK (siehe 9)

Prüfen, ob das richtige Lemma und – sofern es flektierbar ist – die richtige Flexionsbestimmung im Angebot an jeweils erster Stelle stehen. Wenn nicht, Korrektureinträge am Beginn der (ersten) Zeile des Datensatzes.

!!! Vorher müssen durch Anwendung des Makros TASTATUR, Option 1, die richtigen Alt+X-Makros installiert werden. !!!

Falls das richtige Lemma im Angebot nicht vorkommt, muss es mit dem Makro Alt+Y aus dem Lemmaregister geholt werden. Neue Lemmata dürfen im appellativischen Bereich nur in ganz besonderen Ausnahmefällen eingefügt werden, und wenn, dann genau übereinstimmend im „Gravis“- wie im „Trema“-Bereich.

11. Auswertung der Korrektureinträge (siehe 10.1)

Programm: LEM, Option 6
Eingabe-Datei: X-VK.txt
Ausgabe-Datei: X-LO.txt

12. Markierung der Pronominaladverbien (siehe 10.2)

Programm: LEM, Option 7
Eingabe-Datei: X-LO.txt
Text-Datei: X.TXT
Ausgabe-Datei 1: X-LD.txt
Ausgabe-Datei 2: X-DAR.txt

13. Einfügung der Zusätze bei Partikelverben (siehe 10.3)

Programm: LEM, Option 7
Eingabedatei: X-LD.txt
Ausgabedatei: X-LP.txt

14. Formale Überprüfung von X-LP (siehe 10.4)

Programm: LEM, Option 8
Eingabe-Datei: X-LP.txt
Ausgabe-Datei: X-L.txt

15. Überprüfung der Zählung und der Vollständigkeit von X-LEM (siehe 10.5)

Programm: ZAHL, Option 5
Eingabe-Datei: X-L.txt
Ausgabe-Datei: X-LF.txt (= Fehlerdatei; überprüfen und ggf. X-L.txt korrigieren)

16. Erstellung der um das Normalformfeld erweiterten Datei X-N0 (siehe 10.6)

Programm: NFORM
Eingabe-Datei: X-L.txt
Ausgabe-Datei: X-N0.txt

17. Überprüfung von X-N0 (siehe 10.7)

Programm: PRUEF-NF
Eingabe-Datei: X-N0.txt
Ausgabe-Datei: X-NP.txt; nach Abschluß der Arbeit umbenennen in X-N0.txt

19. Erzeugung des Kontrolltextes X-K (siehe 12)

Programm: LEM, Option 9
Eingabedatei: X-N0.txt
Ausgabedatei: X-K.txt

Von hier ab muss das Verfahren noch überarbeitet u.neu dargestellt werden

20. Korrektur von X-K am Bildschirm oder im Ausdruck (in diesem Fall siehe 20a, b, c)

20a. Drucken von X-K

Programm: WordPerfect; Makro KTX-D
Eingabedatei: X-K.txt
Ausgabedatei: X-KD.txt

20b. Korrektur von X-KD

20c. Eintragung der Korrekturen in X-N0

21. Umbenennung von X-N0 in X-N1

nicht mehr benötigte Dateien können gelöscht werden; aufzubewahren sind X.TXT, X-DAR, X-Nk

22. Abgleich von X-NF.K mit der Textdatei

Programm: NFO-WOK

Eingabe-Datei 1: X-NF.K

Eingabe-Datei 2: X-WOK (neu erzeugen!)

Ausgabe-Datei: X-NF.Kp; nach Abschluß der Arbeit umbenennen in X-NF.K

oder (seltener): Erzeugung einer korrigierten Textdatei X-TXK aus X-NF.K:

Programm: NFORM-TX, Menü: 1

Eingabedatei: X-NF.K

Ausgabedatei: X.TXK

23. Übertragung aller Kommentare in eckigen Klammern ins Kommentarfeld

Programm: IND-KLAM

Eingabedatei: X-NF.0 oder X-NF.K

Ausgabedatei: X-NF.0n oder X-NF.Kn; anschließend umbenennen in X-NF.0 oder X-NF.K

24. Überprüfung der Zählung und der Vollständigkeit von X-NF.K

Programm: ZAHL, Option 5

Eingabe-Datei: X-NF.K

Ausgabe-Datei: X-NF.KF (= Fehlerdatei; überprüfen und ggf. X-NF.K korrigieren)

21.5 Übersicht über die neue Benennung der Datei-Typen

Dateityp	Bearbeitungsstand	neuer Name	alter Name
T = Text nach der Handschrift A = Text der Ausgabe/Edition	0 = nicht korrigiert/kollationiert kein Zusatz = korrigiert/kollationiert: () E = präeditiert	X- T0 .txt, X- A0 .txt X- T .txt, X- A .txt X- TE .txt, X- AE .txt	X.TX0 X.TX1
W = WOK-Index		X- W .txt	X-WOK
V = vorlemmatisierter Index	1 = X-W, von Lem-2.spx (2) vorbereitet für die Vorlemmatisierung S = X-V1, alphabetisch sortiert 2 = X-VS, von Lem-3.spx vorlemmatisiert Z = X-V2, vertikal sortiert W = mit der Wörterbuchlemmatisierung durch VL-TRIER.spx verglichen kein Zusatz = X-VZ, mit Lem-4.spx verbessert K = X-V, für die Bildschirmkorrektur mit Lem.spx (5) formatiert	X- V1 .txt X- VS .txt X- V2 .txt X- VZ .txt X- VW .txt X- V .txt X- VK .txt	X-VL.1 X-VL.1s X-VL.2 X-VL.Z – X-VL X-VL.K
L = lemmatisierter Index	0 = Auswertung von X-VK mit Lem.spx (6) D = Bearbeitung der Pronominaladv. von X-L0 mit Lem.spx (7) P = Bearbeitung der Partikelverben von X-LD mit Lem.spx (4) kein Zusatz = X-LP nach Standardüberprüfung mit Lem.spx (8) F = Fehlerdatei bei Vollständigkeitsprüfung von X-L mit Zahl.spx (5)	X- L0 .txt X- LD .txt X- LP .txt X- L .txt X- LF .txt	X-LEM.0
N = Normalformindex	0 = noch nicht KTX-korrigiert 1 = einmal KTX-korrigiert kein Zusatz = doppelt KTX-korrigiert	X- N0 .txt X- N1 .txt X- N .txt	
G = Grapho-Index		X- G .txt	
K = KTX-Text		X- K .txt	

