

Digitalisierung von Primärquellen für die TextGrid-Umgebung: Modellfall Campe-Wörterbuch

Version: 12. Oktober 2009 - 1.0

Arbeitspaket AP 4

verantwortlicher Partner: Universität Würzburg

TextGrid

Modulare Plattform für verteilte und kooperative
wissenschaftliche Textdatenverarbeitung -
ein Community-Grid für die Geisteswissenschaften



Bundesministerium
für Bildung
und Forschung

Projekt: **TextGrid**

Teil des D-Grid Verbundes und der deutschen e-Science Initiative

BMBF Förderkennzeichen: 07TGO1 C

Laufzeit: Februar 2006 - Mai 2009

Dokumentstatus: Final

Verfügbarkeit: öffentlich

Autoren:

Werner Wegstein

Mirjam Blümm

Dietmar Seipel [Abschnitt 8]

Christian Schneiker [Abschnitt 8]

Einleitung	4
1. Beschreibung der Primärquellen	5
1.1 Joachim Heinrich Campe: Wörterbuch der Deutschen Sprache (CWDS).....	5
1.2 Joachim Heinrich Campe: Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke [CWEV]	7
2. Ziele der Digitalisierung	9
3. Bilddigitalisierung	9
3.1 Geräte und Digitalisierungsparameter	10
3.2 Qualitätskontrolle.....	11
3.3 Archivierung	11
4. Texterfassung	12
4.1 Ziele der Texterfassung.....	12
4.2 Verfahren der Texterfassung: OCR versus ‚double keying‘	14
4.3 Richtlinien für die manuelle Texterfassung	15
4.4 Qualitätskontrolle.....	16
5. Vorarbeiten für die Textkodierung	17
5.1 Anpassung der Textauszeichnungen	17
5.2 Erfassungskorrektur	18
5.3 Druckfehlerkorrektur	18
5.4 Anpassung der Fraktur-Typographie.....	19
5.6 Abkürzungen, Namen, Werktitel und Bibelstellen.....	20
5.5 Fehlerkorrekturen im elektronischen Text	20
6. Lexikographische Analyse und Kodierungskonzept.....	20
7. Definition des TEI-Schemas	22
8. Textkodierung	22
8.1 Implementierung in Prolog	23
8.1.1 Benutzerschnittstelle	24
8.1.2 Annotation einzelner Wörterbucheinträge	24
8.1.3 Transformationsregeln	26
8.2 Das Resultat als TEI-Datei.....	27
9. Präsentation.....	28
10. Schlussbemerkung	29

Einleitung

Ziel des Arbeitspakets 4 ist es, Testmaterialien mit stark strukturierten Textdaten und Testmaterialien bestehend aus Bilddaten mit hohen Anforderungen an Qualität und Metadatenverwaltung zu erstellen, an denen die Nutzbarkeit der TextGrid-Werkzeuge in der erforderlichen Komplexität erprobt und in Tests mit Anwendern unterschiedlicher Profile evaluiert werden kann. Die Wörterbücher von Joachim Heinrich Campe entsprechen diesen Anforderungen gleich in mehrfacher Hinsicht. Sie vertreten als Textsorte 'Wörterbuch' in idealer Weise den Typus stark strukturierter Textdaten: sie beschreiben zu jedem Stichwort metasprachlich die Bedeutung durch Paraphrasen oder Synonyme, sie veranschaulichen die Wortbedeutung jeweils auch objektsprachlich durch Kontextbeispiele, meist in Gestalt von Quellenzitaten (mit unterschiedlich genauer Referenz) und sie ergänzen die Gebrauchsbedingungen durch stilistische Bewertungen nach Sprachschichten und Stilfärbungen sowie zeitliche, räumliche und fachsprachliche Zuordnung. Ferner liefern sie Informationen zur grammatischen Form, ggf. auch der Aussprache, der Etymologie und der Wortgeschichte. Mit ihrem komplexen Nebeneinander von Fraktur-Grundschrift und Antiqua-Fonts sowie Sprachformen, die noch nicht orthographisch und grammatisch normiert sind, demonstrieren sie dabei anschaulich die Probleme der Digitalisierung von Texten aus dem Beginn des 19. Jahrhunderts. Technisch stellen sie durch die Einbeziehung griechischer und hebräischer Wortbelege besondere Anforderungen an den Zeichensatz. Auch überrascht es bei einem Wörterbuch aus der Frühzeit der hochdeutschen Lexikographie nicht, dass die Wortartikel noch keineswegs durchgängig einheitlich strukturiert sind, so dass sie dadurch eine Herausforderung an die Entwicklung eines einheitlichen Markup-Konzepts darstellen: zwar soll die strukturelle Vielfalt authentisch abgebildet werden, dies darf aber nicht auf Kosten der Validierbarkeit differenzierter Datenstrukturen gehen, d.h. valider TEI P5-Daten auf der Basis von XML Schema.

Als Modellfall für Bilddaten mit hohen Anforderungen an Qualität und Metadatenverwaltung eignen sich Campes Wörterbücher in gleicher Weise. Sie sind bisher noch nicht digitalisiert, die Digitalisate schließen somit eine Lücke bei den Wörterbüchern des 19. Jahrhunderts. Vom Volumen her sind sie mit ca. 6000 Seiten ausreichend umfangreich. Die verwendeten Schriftgrößen, Qualität und Farbe des Papiers, Druckfarbe und Sättigung sowie problematische Detailstrukturen der Frakturfonts erfordern eine Digitalisierung mittels hochauflösender Bilder in Farbe. Eine unkomprimierte Druckseite beansprucht als Bild dadurch etwa 100 MB an Speicherplatz. Insgesamt kommt so ein Datenvolumen von ca. 600 GB zusammen. Die Komplexität der Metadatenverwaltung wird bestimmt von der Granularität, in der die Daten verarbeitet und später auch repräsentiert werden sollen: Gesamtwerk, Einzelband, Buchstabenstrecke, Druckseite oder einzelner Wörterbucheintrag. Hinzu kommen noch die Textbestandteile des Wörterbuchs, die über den reinen Wörterbuchtext hinausgehen - Vorreden, Zuruf, Berichtigungen und Zusätze, Druckfehler und Verbesserungen, Ehrenrettung, Preisschrift sowie ein 'Verzeichniß der in diesem Wörterbuche, zum Ersatz fremder Ausdrücke, vorgeschlagenen neuen und, der Erneuerung würdig scheinenden, alten Wörter, welche zum Theil, entweder ganz, oder doch

einer hier nachgewiesenen Bedeutung nach, im Adelungischen Wörterbuche fehlen'. Auch im Bereich der Bilddaten wie der Metadaten erweisen sich Campes Wörterbücher somit als geeignetes Testmaterial.

1. Beschreibung der Primärquellen

Im Rahmen des Projekts wurden zwei Wörterbücher Joachim Heinrich Campes digitalisiert, die in sachlichem Zusammenhang stehen: das fünfbändige 'Wörterbuch der deutschen Sprache' [cwds], erschienen von 1807 - 1811 als Reaktion auf Adelungs vierbändiges 'Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart'¹ und das im Titel ausdrücklich als Ergänzungsband zu Adelungs und Campes Wörterbüchern bezeichnete 'Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke' [cwev], erschienen in zweiter Auflage 1813 auf der Basis früherer Vorarbeiten zwischen 1790 und 1801, die in die Materialsichtung einbezogen wurden.

1.1 Joachim Heinrich Campe: Wörterbuch der Deutschen Sprache (CWDS)

Wörterbuch der Deutschen Sprache. Veranaltet und herausgegeben von Joachim Heinrich Campe. Erster Theil. A - bis – E. Braunschweig 1807. In der Schulbuchhandlung. XXIV, 1023 S. [cwds1]

Strukturdaten:

- | | |
|----------------|--|
| a. S. I | Titelblatt |
| b. S. II | leer (ohne Paginierung) |
| S. III - XXIII | 1. Vorrede von J.H. Campe |
| S. XXIV | 2. Druckfehler und Berichtigungen (ohne Paginierung) |
| c. S. 1 - 356 | Buchstabenstrecke A |
| S. 356 - 661 | Buchstabenstrecke B |
| S. 661 - 663 | Buchstabenstrecke C |
| S. 663 - 810 | Buchstabenstrecke D |
| S. 810 - 1023 | Buchstabenstrecke E |

Wörterbuch der Deutschen Sprache. Veranaltet und herausgegeben von Joachim Heinrich Campe. Zweiter Theil. F - bis – K. Braunschweig 1808. In der Schulbuchhandlung. IV, 1118 S. [cwds2]

¹ Johann Christoph Adelung (1774 - 1786): Versuch eines vollständigen grammatisch-kritischen Wörterbuchs der hochdeutschen Mundart, mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen, Leipzig, 5 Bde.
Johann Christoph Adelung (1793 - 1801): Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart, mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen, Zweyte vermehrte und verbesserte Ausgabe Leipzig, 4 Bde.

Strukturdaten:

- a. S. I Titelblatt
- b. S. II leer (ohne Paginierung)
- S. III - IV 1. Vorrede von J.H. Campe
- c. S. 1 - 210 Buchstabenstrecke F
- S. 211 - 487 Buchstabenstrecke G
- S. 487 - 814 Buchstabenstrecke H
- S. 814 - 832 Buchstabenstrecke I
- S. 832 - 861 Buchstabenstrecke J
- S. 861 - 1104 Buchstabenstrecke K
- d. S. 1105 - 1116 1. Berichtende Beilage des Ausarbeiters.
- S. 1117 - 1118 2.1 Im ersten Theile ist noch Folgendes zu berichtigen:
(ohne Paginierung)
- S. 1118 2.2 Im zweiten Theile: (ohne Paginierung)
- 3. Erklärung der Zeichen *) und Kürzungen (als Streifen
am Rand angeklebt, ohne Paginierung)

Wörterbuch der Deutschen Sprache. Veranaltet und herausgegeben von Joachim Heinrich Campe. Dritter Theil. L - bis – R. Braunschweig 1809. In der Schulbuchhandlung. IV, 908 S.[cwds3]

Strukturdaten:

- a. S. I Titelblatt
- b. S. II leer (ohne Paginierung)
- S. III 1. Zuruf eines Todten / an die Deutsche Allgemeinde, /
Publicum genannt. von Herder.
- S. III - IV 2. Vorrede von J.H. Campe
- c. S. 1 - 176 Buchstabenstrecke L
- S. 176 - 385 Buchstabenstrecke M
- S. 385 - 529 Buchstabenstrecke N
- S. 529 - 573 Buchstabenstrecke O
- S. 573 - 715 Buchstabenstrecke P
- S. 715 - 729 Buchstabenstrecke Q
- S. 729 - 908 Buchstabenstrecke R
- d. S. 908 1. Druckfehler und Verbesserungen.
- 1.1 Im ersten Theile ist noch Folgendes zu berichtigen.
- 1.2 Im zweiten Theile.
- 1.3 Im dritten Theile.
- 2. Erklärung der Zeichen *) und Kürzungen (als Streifen
am Rand angeklebt, ohne Paginierung) = Band 2, d3.

Wörterbuch der Deutschen Sprache. Veranaltet und herausgegeben von Joachim Heinrich Campe. Vierter Theil. S - und – T. (Nebst einer Beilage) Braunschweig 1810. In der Schulbuchhandlung. IV, 945 S.[cwds4]

Strukturdaten:

- a. S. I Titelblatt
- b. S. II leer (ohne Paginierung)
- S. III - IV 1. Vorrede von J.H. Campe
- c. S. 1 - 757 Buchstabenstrecke S

S. 757 - 915	Buchstabenstrecke T
d. S. 916 - 940	1. Ehrenrettung / dieses / Wörterbuches der Deutschen Sprache./ Von / dem Ausarbeiter desselben / Theodor Bernd. / Eine Beilage.
S. 941	2. Berichtigungen *) und Zusätze. (ohne Paginierung)
S. 941	Zum vierten Theile.
S. 941	3. Außer der schon verschiedene Mahl gegebenen Berichtigun-/gen und Zusätzen zu den frühern Theilen dieses /Werkes sind noch folgende nachzuholen.
S. 941	3.1 Zum ersten Theile.
S. 942 - 943	3.2 Zum zweiten Theile.
S. 943 - 944	3.3 Zum dritten Theil.
S. 945	4. Erklärung der Zeichen *) und Kürzungen (als Streifen am Rand angeklebt, ohne Paginierung) = Band 2, d3.

Wörterbuch der Deutschen Sprache. Veranstatlet und herausgegeben von Joachim Heinrich Campe. Fünfter und letzter Theil. U - bis – Z. Braunschweig 1811. In der Schulbuchhandlung. IV, 979 S. [cwds5]

Strukturdaten:

a. S. I	Titelblatt
b. S. II	leer (ohne Paginierung)
S. III - IV	1. Vorrede von J.H. Campe
c. S. 1 - 254	Buchstabenstrecke U
S. 254 - 524	Buchstabenstrecke V
S. 524 - 803	Buchstabenstrecke W
S. 803 - 804	Buchstabenstrecke X
S. 804	Buchstabenstrecke Y
S. 804 - 977	Buchstabenstrecke Z
d. S. 978	1. Berichtigungen *) und Zusätze. (ohne Paginierung)
	Zum fünften Theile.
S. 978	2. Nachtrag zu den Berichtigungen der vier ersten Theile.
S. 978	2.1 Zum ersten Theile.
S. 979	2.2 Zum zweiten Theile.
S. 979	2.3 Zum dritten Theile.
S. 979	2.4 Zum vierten Theile.

Anmerkung: Die unpaginierten, eine Spalte breiten Benutzerhilfen mit dem Titel 'Erklärung der Zeichen und Kürzungen' wurden, soweit sie beim Binden noch vorhanden, unterschiedlich behandelt: teils zum Ausklappen an eine Seite angeklebt, teils als Blatt eingebunden. Sie wurden nur einmal, zusammen mit Band 2, digitalisiert.

1.2 Joachim Heinrich Campe: Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke [CWEV]

Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke. Ein Ergänzungsband zu Adelung's und Campe's Wörterbüchern. Neue Starkvermehrte und durchgängig verbesserte Ausgabe von Joachim Heinrich Campe.

Braunschweig 1813. In der Schulbuchhandlung. 673 S. [cwev]

Strukturdaten:

- | | |
|---------------|--|
| a. S. I | Titelblatt |
| b. S. II | 1. Widmung: 'Unsere Sprache an uns'. (Klopstock-Zitat) |
| S. III - XII | 2. Vorrede / zur ersten Ausgabe. Braunschweig, im Herbstmonde 1800. |
| S. XIII - XIV | 3. Vorrede / zur zweiten Ausgabe dieses Werks. von J. H. Campe |
| c. S. 1 - 70 | 1. Grundsätze, Regeln und Grenzen / der / Verdeutschung. - Eine von dem königlichen Gelehrtenverein zu Berlin gekrönte Preisschrift. |
| S. 71 - 139 | Buchstabenstrecke A |
| S. 139 - 161 | Buchstabenstrecke B |
| S. 161 - 244 | Buchstabenstrecke C |
| S. 244 - 276 | Buchstabenstrecke D |
| S. 276 - 307 | Buchstabenstrecke E |
| S. 307 - 330 | Buchstabenstrecke F |
| S. 330 - 345 | Buchstabenstrecke G |
| S. 345 - 360 | Buchstabenstrecke H |
| S. 360 - 390 | Buchstabenstrecke I |
| S. 390 - 391 | Buchstabenstrecke K |
| S. 391 - 404 | Buchstabenstrecke L |
| S. 404 - 430 | Buchstabenstrecke M |
| S. 430 - 441 | Buchstabenstrecke N |
| S. 441 - 453 | Buchstabenstrecke O |
| S. 453 - 509 | Buchstabenstrecke P |
| S. 509 - 514 | Buchstabenstrecke Q |
| S. 514 - 542 | Buchstabenstrecke R |
| S. 542 - 579 | Buchstabenstrecke S |
| S. 579 - 597 | Buchstabenstrecke T |
| S. 597 - 600 | Buchstabenstrecke U |
| S. 601 - 609 | Buchstabenstrecke V |
| S. 609 - 610 | Buchstabenstrecke W |
| S. 610 | Buchstabenstrecke X |
| S. 610 | Buchstabenstrecke Y |
| S. 610 | Buchstabenstrecke Z |
| S: 611 - 612 | Druckberichtigungen. (ohne Paginierung) |
| S. 613 -673 | Verzeichniß / der in diesem Wörterbuche, zum Ersatz fremder Ausdrücke, vorgeschlagenen neuen und, der Erneuerung würdig / scheinenden, alten Wörter, welche zum Theil, entweder ganz, oder doch einer hier nachgewiesenen Be- / deutung nach, im Adelungischen Wörterbuche fehlen. |
| S. 674 | Impressum 'Gedruckt bei Friedrich Vieweg.' |

Campes Vorarbeiten und die Vorstufen zum CWEV wurden im Verlauf des Projekts ebenfalls in die Sichtung mit einbezogen und - soweit sie nicht vor Ort im Original zur Verfügung standen - als Digiwunschbuch an der SUB Göttingen digitalisiert:

Campe, Joachim Heinrich. Proben einiger Versuche von deutscher Sprachbereicherung. In:

Braunschweigisches Journal philosophischen, philologischen und pädagogischen Inhalts. Herausgegeben von E.Chr. Trapp. Elfte Stück, November 1790, S. 257-296.

Campe, Joachim Heinrich. Proben einiger Versuche von deutscher Sprachbereicherung. Von Joachim Heinrich Campe. Braunschweig, 1791.

Zweiter Versuch deutscher Sprachbereicherungen oder neue, stark vermehrte Ausgabe des ersten von Joachim Heinrich Campe. Braunschweig: in der Schulbuchhandlung 1792.

<http://resolver.sub.uni-goettingen.de/purl?PPN556283988>

Campe, Joachim Heinrich. Ueber die Reinigung und Bereicherung der Deutschen Sprache, dritter Versuch welcher den von dem königl. Preuß. Gelehrtenverein zu Berlin ausgesetzten Preis erhalten hat. Verb. und verm. Ausg. (nebst Nachtr. u. Berichtigungen). Braunschweig. Schulbuchh. 1794. (24, CCLXII, 352 S.)

Campe, Joachim Heinrich. Nachtrag und Berichtigung zum ausübenden Theile der Campischen Preisschrift über die Reinigung und Bereicherung der Deutschen Sprache. Braunschweig. Schulbuchh. 1794.

Joachim Heinrich Campe: Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedruckten fremden Ausdrücke. Ein Ergänzungsband zu Adelungs und Campes Wörterbüchern. In zwei Bänden, Braunschweig: in der Schulbuchhandlung 1801.[cwev1801]

Erster Band A – E. <http://resolver.sub.uni-goettingen.de/purl?PPN556270789> ,

Zweiter Band F – Z. <http://resolver.sub.uni-goettingen.de/purl?PPN556271092> .

2. Ziele der Digitalisierung

Die Parameter für die Bilddigitalisierung in Archivqualität wurden nach eingehender Analyse zweier Druckexemplare so ausgelegt, dass - im Hinblick auf die komplexe Mischung von Fraktur- und Antiqua-Fonts - die Digitalisate für die Texterfassung mit optischen Zeichenerkennungsprogrammen, z.B. FineReader XIX, bearbeitet werden können, dass sie aber auch für Ausdrücke in guter Qualität auf Laser- bzw. Tintenstrahldruckern bis zum Format DIN A3 geeignet sind.

Bei der Kodierung der digitalen Textdaten wurde darauf geachtet, dass das Datenformat die linearen Strukturen des Drucks berücksichtigt und damit eine buchorientierte Sicht auf Seite und Zeile des Wörterbuchs zulässt, damit zusammen gleichzeitig aber auch die Strukturformen der Wörterbucheinträge so abbildet, dass die Wörterbuchdaten sehr differenziert abgefragt, ohne manuellen Eingriff nach semantischen oder logischen Konzepten konvertiert und in weiteren Arbeitsschritten iterativ in ihrer Kodierungsgranularität noch verfeinert werden können.

3. Bilddigitalisierung

3.1 Geräte und Digitalisierungsparameter

Geräte und Arbeitsorganisation

Als Scanner für die Basisdigitalisierung diente ein Avison Flachbettscanner A3 mit Buchkante. Für durch die Qualitätskontrolle veranlasste Nach-Scans konnte der Buchscanner Zeutschel OS 10000 im Digitalisierungszentrum der Universitätsbibliothek Würzburg genutzt werden². Die Arbeiten wurden bandweise durch ein Zweierteam von Studierenden des Aufbaustudiengangs ‚Linguistische Informations- und Textverarbeitung‘ mit dem Schwerpunkt ‚Bildverarbeitung‘ ausgeführt.

Auflösung, Farbtiefe und Datenformate der Bilddateien

Den Zielsetzungen und der Druckanalyse entsprechend wurde für die Bilddigitalisierung der Archivdateien eine Auflösung von 600 dpi mit einer Farbtiefe von 24-bit (RGB-Farbe, 8-bit pro Farbkanal) im TIFF-Dateiformat gewählt. Die Bildgröße der unkomprimierten Archivdateien im TIFF-Format (Adobe Photoshop) liegt je nach Randgröße pro Seite bei etwa 100 MB. Die Arbeitsdateien mit gleicher Auflösung und Farbtiefe, jedoch im komprimierenden JPEG-Format mit maximaler Qualität (Adobe Photoshop Qualitätsstufe 12) benötigen je nach Seitenfüllung, Welligkeit und Fleckigkeit des Papiers zwischen 3 und 10 MB. Für die Präsentation, etwa im Internet, können die Bilder in zufriedenstellender Qualität auf eine Auflösung von 150 dpi reduziert werden. Bei mittlerer Qualität der JPEG-Kompression (Adobe Photoshop Qualitätsstufe 6) beanspruchen sie dann noch etwas mehr als 500 KB Speicherplatz.

Dateibezeichnung

Die Bezeichnung der Bilddateien setzt sich zusammen aus dem Kürzel „cwds“ für 'Campe: Wörterbuch der Deutschen Sprache' bzw. „cwev“ für 'Campe: Wörterbuch zur Erklärung und Verdeutschung', für „cwds“ zusätzlich einer Bandnummer, dem Unterstrich als Trennzeichen, einem Kennbuchstaben für die Bestandteile Titelblatt (a), Vorspann (b), Hauptteil (c) und Nachspann (d) sowie vier Ziffern für die Seitenzahl. Die Zählung beginnt mit 0001 für das Titelblatt (a) und den Hauptteil (c), unabhängig davon, ob die Seite eine Seitenzahl - in römischen oder arabischen Ziffern - trägt oder nicht, so dass Seitenzählung und Ziffer in der Dateibezeichnung übereinstimmen. Unbedruckte Seiten wurden nicht digitalisiert, aber bei der Zählung berücksichtigt. Die Digitalisate mit dem Beginn des ersten Bandes des ‚Wörterbuchs der Deutschen Sprache‘ sind beispielsweise in folgenden Bilddateien gespeichert:

cwds1_a00001.tif	Titelblatt
	S. II ist unbedruckt und wurde nicht digitalisiert
cwds1_b0003.tif	Vorrede von J.H. Campe auf den Seiten III – XXIII bis
cwds1_b0023.tif	

2 Überlegungen und Arbeitsschritte und zur Bildverarbeitung im Detail finden sich bei: Anne R. Kenney and Oya Y. Rieger: Moving Theory into Practice: Digital Imaging for Libraries and Archives. Mountain View, CA: Research Libraries Group, Inc., 2000; als Online-Tutorial unter: <http://www.library.cornell.edu/preservation/tutorial/> (Zugriff: 31.5.2009).

DFG Praxisregeln „Digitalisierung“ zu den Förderprogrammen der Wissenschaftlichen Literaturversorgungs- und Informationssysteme. Erarbeitet durch den Unterausschuss für Kulturelle Überlieferung am 12. und 13. Februar 2009, beschlossen durch den Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme am 20.03.2009. © April 2009, Deutsche Forschungsgemeinschaft, Geschäftsstelle: Kennedyallee 40, D-53175 Bonn. Online unter: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/praxisregeln_digitalisierung.pdf

cwds1_b0024.tif	‚Druckfehler und Berichtigungen‘ auf der nicht paginierten Seite XXIV
cwds1_c0001.tif	Wörterbuch, beginnend mit dem Buchstabenbereich A

3.2 Qualitätskontrolle

Zeitnah zur Bilddigitalisierung der einzelnen Bände wurden die Digitalisate separat auf Vollständigkeit der Seiten, die Übereinstimmung von Bild und Dateibezeichnung und visuell auf Vollständigkeit der Bildinformation überprüft. Im Vorfeld der Textfassung wurde in einem zweiten Kontrollschritt in Verbindung mit der Ausrichtung der Bilder und der Nachbearbeitung der Ränder, soweit erforderlich, die Qualität noch einmal im Detail auf nichtdigitalisierte Fehlstellen durch Falten, geknickte Seiten, Löcher und minderwertiges Papier mit der Erfassung behindernder, stark durchschlagender Druckfarbe geprüft. Durch diese Detailkontrolle wurden auch im Nachspann des ersten Bandes zwei vom Buchbinder falsch eingebundene Seiten mit Berichtigungen (ohne Paginierung) aus dem zweiten Band entdeckt und umgeordnet. Insgesamt wurden rund 200 Seiten aus einem Zweitexemplar des Wörterbuchs in der Universitätsbibliothek nachgescannt.

3.3 Archivierung

Zur kurz- bzw. mittelfristigen Sicherung wurden die 5.698 Bilddateien (Gesamtdatenvolumen 537 GB) auf zwei verschiedenen Medienträgern in Archivqualität zwischengespeichert. Die langfristige Speicherung der Archivdateien innerhalb von 'TextGrid' geschieht über die TextGrid-Middleware, die zusammen mit den Grid-Knoten und Datenbanken das TextGrid-Repository ('TextGrid-Rep') bildet. Die DatenGrid-Knoten von TextGrid werden dabei allesamt von einem zentralen TextGrid-crud-Service für die Operationen create-, read, update- und delete ('crud') bedient, der den Clients – beispielsweise dem 'TextGrid-Lab' – lesenden und schreibenden Zugriff auf die gespeicherten TextGrid-Objekte ermöglicht.

Der TextGrid-Report 3.6 'TextGrid-Installation eines Datenknotens'¹ beschreibt hierzu eingehend den technischen Hintergrund des Speicherverfahrens und erläutert in Kapitel 2 am Beispiel von Campe-Wörterbuch-Dateien die zum Einspielen großer Datenmengen erforderlichen Arbeitsschritte sowie in Kapitel 4 das Replica-Management im Grid. Die Funktionalität und Steuerung von TextGrid-crud werden in Abschnitt 7 des TextGrid Manual 3.5: Tool development im Detail beschrieben.

Ergänzend zur Nutzung in der TextGrid-Arbeitsumgebung hat das Digitalisierungszentrum der Universitätsbibliothek Würzburg³ die Bilddateien des Campe-Wörterbuchs mit DigiTool⁴ in stark komprimierter Form für den METS-Viewer aufbereitet und durch die Registrierung der Metadaten im Verbundkatalog dafür gesorgt, dass das Werk auch in der Bibliotheksumgebung auffindbar und nutzbar ist.

³ Dem Leiter des Digitalisierungszentrums, Herrn Dr. Hans-Günther Schmidt und Herrn Marco Dittrich danke ich sehr für die engagierte Unterstützung des TextGrid-Projekts ‚Campe-Wörterbuch‘.

⁴ <http://www.exlibrisgroup.com/de/category/DigiToolOverview> (Zugriff: 15.10.2009); zum Hintergrund: Uwe M. Borghoff / Peter Rödiger, Standards und Standardisierung in Kontext von Grid/eScience und Langzeitarchivierung, nestor-materialien 11, hg. vom nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland, 2009. Link: http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_11.pdf (Zugriff: 15.10.2009).

4. Texterfassung

4.1 Ziele der Texterfassung

Die typographische Analyse einer Wörterbuchseite – hier Seite 2 von Band 1 des ‚Wörterbuchs der Deutschen Sprache‘ von Campe – macht verständlich, warum sich Wörterbücher als Testmodell für stark strukturierte Textdaten besonders eignen.

Zunächst gilt es die linearen Strukturen von Seite, Spalte und Zeile zu erfassen: Seitenanfang, Kolumnentitel⁵ und Seitenzählung sowie die Spaltenaufteilung. Besonderes Augenmerk verdient dabei die Zeilenstruktur: Ausrückungen markieren die Stichwörter⁶, Einrückungen bzw. Leerraum in sehr unterschiedlicher Tiefe kennzeichnen Zitate.

Innerhalb der Zeile gilt es die verschiedenen Schriftfamilien auseinander zu halten, Fraktur⁷ für die deutschen Textbestandteile, Antiqua für fremdsprachige, insbesondere in lateinischer Sprache, die Laufweiten der Schrift, z.B. Sperrung für Namen (z.B. auf S. 2. *Schreiber* im Wortartikel ‚Der Aar‘), und die Abstufungen der Schriftgröße. Stichwörter zu Beginn eines Wortartikels sind in einer Schrift von ca. 10p gesetzt, der Text des Wortartikels in ca. 8p, das Stichwort wird innerhalb des Wortartikels oder innerhalb eines Zitats aber wiederum durch eine Schriftgröße von ca. 9p mehr oder weniger deutlich hervorgehoben. Schließlich müssen auch für die Erfassung von Sonderzeichen, auf S. 2 etwa zum Stichwort ‚Die Aalraupe‘, ‚Der Aar‘ oder ‚Aashaft‘ und für fremdsprachliche Alphabete, etwa Griechisch oder Hebräisch geeignete Repräsentationsformen festgelegt werden.

⁵ Kolumnentitel wurden bei der Texterfassung zwar berücksichtigt, als druckspezifische Kennzeichnung wurden sie in die elektronische Textedition aber nicht mit übernommen. Reklamanten und Kustoden, Ordnungshilfen für die korrekte Verarbeitung des Druckes wurden gar nicht erst erfasst.

⁶ Die Kennzeichnung von Ausrückungen stößt bei der manuellen Texterfassung durch nicht deutsch sprechende Datentypisten auf Verständnissgrenzen: bei zwei aufeinanderfolgenden einzeiligen Wörterbucheinträgen wird der zweite Eintrag nicht mehr als Ausrückungen erkannt und markiert. Beispiel auf S.2: ‚Die Aarweihe‘ nach ‚Der Aaron‘.

⁷ Zum geschichtlichen Hintergrund: Albert Kapr, *Fraktur, Form und Geschichte der gebrochenen Schriften*, Mit einem Beitrag von Hans Peter Willberg und allen Satzregeln zur Fraktur von Friedrich Forssman. Verlag Hermann Schmidt, Mainz 1993. Albert Kapr: *Schriftkunst. Geschichte, Anatomie und Schönheit der lateinischen Buchstaben*. Verlag der Kunst, Dresden 1971.

fer gestellt wird, woran sich Kale und Weißfische sehen und womit sie herausgezogen werden.

† Die Kalraupe, *Mz.* die —n, ein in süßem Wasser lebender Fisch, der eine glatte Haut wie der Kal, aber einen dicken Kopf und Bauch hat, (*Gadus lota* L.). In N. D. die Kalquappe, in D. D. der Kaubaal, der Kuppe, Rutte, Rutten, Trusch, Trusche und Trüsche.

Die Kalreufe, *Mz.* die —n, eine Reufe zum Kalfangen.

Die Kalfschlange, *Mz.* die —en, oder der Meeraal, eine Kalart, größer als der gemeine Kal, mit aschgrauem Rücken, weißem Bauche, weißlich getüpfelten Seitentlinien und silberfarbigen Ringen um die großen Augen. Er hat zwei kurze Bartfaden und die Kiemen sind mit kleinen spitzigen Zähnen besetzt (*Muraena congens* L.).

† Der Kalfschel, des —s, *Mz.* die —n, f. Kalfgabel.

Der Kalfstecher, des —s, *Mz.* w. d. *Gz.* 1) + So viel als Kalgabel, f. d. 2) Einer der damit Kale sticht oder fängt.

Der Kalfstreif, des —es, *Mz.* die —e, oder der Kalfstreifen, des —s, d. *Mz.* w. d. *Gz.* der schwarze Streif längs dem Rücken des Kaltes; uneigentlich, ein schwarzer Streif, welchen besonders fahle Pferde auf dem Rückgrath haben; der Kalfstreich.

Der Kalfstrich, des —es, *Mz.* die —e, f. Kalfstreif.

Der Kalfteich, des —es, *Mz.* die —e, ein Teich, in welchem Kale gehalten werden.

Das Kalthierchen, des —s, d. *Mz.* w. d. *Gz.*, bei einigen Naturforschern, das kleine aalsförmige Thierchen, welches sich in Eßig, Kleister, Sauerteige etc. erzeugt.

† Die Kalwate, *Mz.* die —n, ein gewisses Netz zum Kalfangen.

Das Kalwehr, des —es, *Mz.* die —e, eine Art Wehr von Pfählen in einem Flusse, vor welches man Netze, Reusen etc. stellt, um Fische und besonders Kale zu fangen.

Der Kalwels, des —es, *Mz.* die —e, f. Duntaal.

Der Kalwurm, des —es, *Mz.* die —würmer, überhaupt ein an Gestalt dem Kale ähnlicher Wurm, dergleichen die Eßig- und Kleisterwürmer sind.

* Der Kar, des —es, oder —en, *Mz.* die —e, oder —en, die alte Benennung aller großer Raubvögel, besonders aber des Adlers, die noch in N. D. üblich ist und bei Dichtern vorkommt.

Ein kühner Kar theilt mit gewalt'gen Schwingen
Die Lüfte, — — — — — Schreie über.
Bald werdet ihr im Meer der Haie, am Gestade
Der Karen Beute sein. — Ramler.

Der Karon, eine Pflanze, f. Arum.

Der Karweibe, des —n, oder die Karweibe, *Mz.* die —n, N. D. eine dem Kar oder Adler ähnliche Art Weihen.

Das Kas, des —es, *Mz.* die Kaser. 1) Von essen, ursprünglich Alles, was lebenden Geschöpfen zur Nahrung dient. In dieser Bedeutung wird von den Fischen die Nahrung und Lockspeise der großen Fische Kas, und von den Mäulern und Beckern der zusammengekehrte Abfall von Mehl (sonst Steinaas) Sau- und Schweinaas genannt. Da die toten Thier- und Menschenkörper andern Thieren auch zur Nahrung dienen, so wurde die Benennung Kas 2) auch auf toten Thier- und Menschenkörper übertragen.

— — — — — und fand mich — — — — —
— unter zerstückelten Ästern
Meiner Streitgenossen wieder. Schubart.
Und wälzt die Sklavensäfer fort. Wop.

Besonders und gewöhnlich aber wird es von toten, in Häulnis übergehenden und dann stinkenden Körpern der Thiere gebraucht. Hieraus entstand 3) die uneigentliche Bedeutung von Kas, nach welcher es überhaupt etwas Stinkendes, Schmutziges, Häßliches bedeutet. In sofern wird in einigen Gegenden jedes Geschwür Kas genannt, und in der Vöbelsprache jede Person, besonders weiblichen Geschlechts,

der man die höchste Verachtung bezeigen und die man mit einer niedrigen Benennung belegen will.

Die Kasblätter, *Mz.* die —n, eine böseartige Blatter, die schwarz und stinkend ist; die Kaspoche.

Kasen, v. intrs. Mittelw. der vergangenen Zeit, geasest. N. D. auf eine schmutzige und ekelhafte Weise in etwas herumwühlen, und überhaupt eine schmutzige Arbeit verrichten. Im Holsteinischen auch, unordentlich und verschwenderisch mit etwas zu Werke gehen. Die Kühe kasen im fetten Grase, sagt man, wenn sie mehr zertreten, als fressen. Das Kasen.

Die Kasfliege, *Mz.* die —n, eine Art großer Fliegen, die durch das Kas angelockt wird, (*Musca cadaverina* L.); die Fleischfliege, Schmeißfliege.

Kasfräßig, adj. u. adv. was gern Kas frisst, von Thieren gebräuchlich. S. Kasfressig.

Kasfressend, adj. was sich von Kase 2) nährt.

Kasfressig, adj. u. adv. was nicht bloß fleisches, sondern auch faules Fleisch (Kas) frisst, von Thieren gebräuchlich.

Der Kasgeier, des —s, d. *Mz.* w. d. *Gz.*, eine kleine Art Geier von der Größe eines Raben, die am Kopfe und Halse nur wenige weiße Dunen haben, in den Ländern am Mittelländischen Meere wohnen und sie vom Kase reinigen, wozu sie sich in großen Scharen versammeln (*Vultur percnopterus* L.); der Erdgeier.

Der Kasgeruch, des —es, *Mz.* die —gerüche, der unangenehme Geruch des Kases; dann, jeder unangenehme Geruch.

Der Kasgestank, des —es, ohne *Mz.* Gestank, den das Kas verbreitet; dann, ein jeder demselben ähnlicher, — überhaupt sehr unangenehmer, unaussprechlicher Geruch.

Kasgerig, adj. u. adv. nach Kas begierig.

Die Kasgrube, *Mz.* die —n, eine Grube, in welche das Kas geworfen wird; auch + Kasühle.

X Kashaft, adj. u. adv. dem Kase ähnlich, die Eigenschaft des Kases habend (cadaverös). Ein aashafter Geruch. Das edlere, von Menschen gebräuchliche Wort ist leichenhaft oder leichenartig.

X Kasig, adj. u. adv. 1) So viel als aashaft. 2) ≠ Uneigentlich. (S. Kas). 3) Häßlich, schmutzig. Kasig Wetter. In Preußen auch für, krank, matt, träge, faul.

Der Kasrüfer, des —s, d. *Mz.* w. d. *Gz.* eine Art Käfer, welche auf und von dem Kase leben, (*Silpha* L.).

Der Kaskopf, des —es, *Mz.* die —köpfe, eine gewisse Verzierung in der Baukunst, welche dem von der Haut entblößten Kopfe eines Thieres gleicht.

Die Kasrähne, *Mz.* die —n, eine Art Krähen (*Corvus corone* L.), die an andern Orten Nebelkrähe und Rabenkrähe heißt.

† Die Kasühle, *Mz.* die —n, f. Kasgrube.

Die Kasypflanze, *Mz.* die —n, der Name einer ausländischen Pflanzengattung von starkem aashaftem Geruche, der selbst die Kasfliegen anlockt, so daß sie ihre Eier auf dieselbe legen, (*Stapelia variegata* und *hirsuta* L.). Die Gärtner nennen sie unrichtig *Fritillaria crassa*.

Die Kaspoche, *Mz.* die —n, f. Kasblätter.

Der Kasrabe, des —n, *Mz.* die —n, eine Benennung des gemeinen Raben.

X Die Kasseite, *Mz.* die —n, die innenbige Seite der Felle, ebener die Fleischseite.

Der Kasvogel, des —s, *Mz.* die —vögel, in allgemeinen jeder Vogel, der sich von Kase nährt.

Kb, adv. u. praep. I. Als Umstandswort bezeichnet ab eine Trennung, Absonderung, und überhaupt eine Entfernung. Der Deckel ist ab. Drei Schritte vom Wege ab. Hut ab! Auf und ab, ab und zu, ab und an, für uns gefähr, bald mehr bald weniger. Der Preis ist verschieden, einen Thaler auf und ab. Auch bloß ab in Rechnungen und Schauspielen, wo es bedeutet, siehe ab, tritt ab. Gewöhnlich wird es von einer Bewegung von der in Rede stehenden Sache hinweg, oder von einer Bewegung nach unten

All diese zuverlässigen typographischen Kennzeichnungen sind später sehr gut für die elektronische Kodierung des Wörterbuchs zu nutzen und müssen daher in jedem Fall bei der Textfassung in geeigneter Weise gekennzeichnet um mit erfasst werden.

4.2 Verfahren der Texterfassung: OCR versus ‚double keying‘

Optische Zeichenerkennung (Optical Character Recognition, OCR) ist als automatisches Verfahren zur Konvertierung elektronischer Bildinformation in elektronische Textinformation schon seit einiger Zeit ausgereift und bietet sich für die Digitalisierung umfangreicher Texte förmlich an. Parallel zur stetig steigenden Leistung der Scanner zur Erzeugung elektronischer Abbilder der Textseiten wurde die Leistungsfähigkeit der Verarbeitungssoftware zur Zeichenerkennung so verfeinert, dass unter Idealbedingungen, wie sie moderne Drucke bieten, gute OCR-Programme nahezu fehlerfreie Resultate erzeugen.

Dies gilt freilich nicht für die Zeit um 1800, da Drucke aus diesem Zeitraum den Anforderungen für einen erfolgreichen Einsatz von OCR-Programmen meist nicht entsprechen: zerbrochene oder ineinanderlaufende Buchstaben, welliges Papier, das schon beim Scan-Vorgang Verzerrungen erzeugt, Verschmutzungen, wechselnder Sättigungsgrad der Druckfarbe, von der Rückseite durchscheinende Buchstaben, von problematischen Schriftschnitten und den Buchstabenformen der Frakturschrift ganz zu schweigen: Die Liste der Störfaktoren, die den Erkennungsvorgang beeinträchtigen, ließe sich leicht fortsetzen. Sie machen die zuverlässige Unterscheidung von Buchstaben, die ohnehin nur in minimalen Details von einander abweichen, sehr schwierig: großes I, die Ziffer 1 und der Kleinbuchstabe l oder bei der Campe-Frakturschrift z.B. s und f, u und n, m und w, B und V, E und G. Ein mit der Sprache und den Schriftformen vertrauter Leser liest über solche schwierigen Druckpassagen dank der Unterstützung durch den Kontext mühelos hinweg, praktisch ohne es zu merken. Ein OCR-Programm kann aber nur die optische Information auswerten, und wo diese gestört oder verzerrt ist, liefert es im günstigsten Fall ein Rückweisungszeichen, im ungünstigen kommt es zu einer Fehlzuordnung, weil eben nicht mehr erkannt werden kann, als im Druckbild vorhanden ist. Korrekturwörterbücher als Hilfe scheiden bei historischen Texten vor 1900 meist aus, da zu der Zeit, in der die Texte entstanden, eine orthographische Normierung wie sie heute üblich ist, noch nicht existierte. Eine Nachbearbeitung des mit OCR-Programmen erfassten Textes ist also in jedem Fall nötig, wenn als Ziel angestrebt wird, möglichst fehlerfreie digitale Texte mit Referenzqualität zu erstellen.

Die Entscheidung, welches Erfassungsverfahren effektiver und wirtschaftlicher ist – den Text zweimal unabhängig abschreiben zu lassen, möglichst von Datentypisten, die die Sprache nicht sprechen und damit bei historischen Texten keine Interferenzfehler durch den Einfluss der Gegenwartssprache produzieren und anschließend durch Vergleich der beiden Textfassungen die Abschreibefehler zu minimieren, das Verfahren des ‚Double Keying‘ oder der Einsatz eines OCR-Programms – hängt beim OCR-Einsatz ganz wesentlich nicht allein von dem zusätzlichen Korrekturaufwand hinsichtlich der Genauigkeit der Zeichenerkennung ab, sondern auch davon, ob und wie genau das Programm Formatierungsinformationen und zusätzliche typographische Kennzeichnungen in der Zielformatierung wiedergeben kann.

Im Vorfeld der Texterfassung wurde die Zeichenerkennungsleistung der OCR-Programme Fine Reader XIX und FineReader 9 nach eingehendem Training an Digitalisaten der Drucke von Campes Wörterbüchern untersucht. Bei FineReader XIX schwankte die Fehlerrate (ohne Formatierungsfehler) zwischen Werten von 100 bis 120 Fehlern pro Seite, bei Finereader 9 (ebenfalls ohne Layoutabweichungen) zwischen 50 und 100 Fehlern. Mit dem günstigsten Wert von 50 Fehlern pro Seite bezogen auf einen Seitendurchschnitt von rund 7430 Zeichen, erreicht FineReader 9 eine Genauigkeit von 99,327 %, FineReader XIX kommt mit 100 Fehlern pro Seite auf den Wert von 98,654 %. Die DFG-Praxisregeln ‚Digitalisierung‘ vom April 2009 klassifizieren eine Erfassung mit

einer Genauigkeit von weniger als 99,5% als wertlos⁸. Retrospektiv findet Stanislava Grigorova im direkten Vergleich mit einer Stichprobe über die ersten fünf Wörterbuchseiten von Campes ‚Verdeutschungswörterbuch‘ (S. 71 – 75), insgesamt 37.153 Zeichen in 602 Zeilen, in der manuellen Texterfassung durch das ‚Double Keying‘-Verfahren insgesamt drei Erfassungsfehler: auffätzig statt aufsätzig (S. 71 Z. 36), besinden statt befinden (S. 72 Z.24) und Vernis statt Bernis (S. 73 Z. 44)⁹. Manuelle Texterfassung erreicht hier eine Genauigkeit von 99,991%. Das bedeutet, eine Fehlerrate von weniger als einem Fehler auf 10.000 Zeichen und entspricht ziemlich genau dem üblichen Wert für qualitativ hochwertige digitalisierte Texte¹⁰.

Die Kosten dafür liegen bei Doppelerfassung mit Textvergleich und Fehlerkorrektur von kompliziertem Satz in Frakturschrift mit Einfügungen in Antiqua-Schrift, ergänzt um zusätzliche Markierungen bei 0,59 € pro 1.000 Anschläge und sind damit deutlich günstiger als Fehlersuche und -korrektur über OCR. Allerdings wird auch die Leistung der OCR-Verfahren durch neuartige Techniken für eine zuverlässige Fehleridentdeckung und -korrektur permanent weiterentwickelt. Es lohnt daher in jedem Fall, auch künftig die unterschiedlichen Verfahren an den zu digitalisierenden Materialien zu testen, bevor eine Entscheidung über das Verfahren der Texterfassung getroffen wird.

4.3 Richtlinien für die manuelle Texterfassung

Basis für die Ausschreibung der Texterfassung ist die Richtlinie ‚Buchstabenmuster und Kodierungsvorgaben für die Texterfassung‘, die an Abbildungen sämtlicher zu erfassender Zeichen in Frakturschrift und Antiqua aus den Wörterbüchern die dafür geforderte Zeichenkodierung festlegt und alle typographischen Besonderheiten beschreibt, die mit zu erfassen sind. Die Beschreibung ist als Anhang 1 beigelegt.

Bei der Auftragsvergabe an eine Firma in China wurden noch kleinere Modifikationen vereinbart, auf die das Erfassungsteam für Frakturschrift eingestellt war: die Kodierung von Seitenzahl und Kolumnentiteln in einer Zeile am Beginn der Seite, die Markierung von Ausrückungen, Leerraum und Zahl der Spalten, die Kennzeichnungen für Schriftgrößen und Auszeichnungen, sowie vermutete Druckfehler {?K}, sichere Druckfehler {!D} und nicht lesbaren Zeichen {??}.

Erst nach den ersten Texterfassungsproben zeigte sich, dass die Richtlinien für nicht deutsch sprechende Datentypistinnen, die sich bei der Texterfassung allein auf die Form der Buchstaben konzentrieren können und die durch Wortbedeutung und Sinnzusammenhang nicht unterstützt, aber auch nicht abgelenkt werden, noch präzisiert werden mussten¹¹. Problematisch für die Erfassung ist –

⁸ DFG-Praxisregeln ‚Digitalisierung‘ zu den Förderprogrammen der Wissenschaftlichen Literaturversorgungs- und Informationssysteme. Deutsche Forschungsgemeinschaft Bonn, April 2009, [Link: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/praxisregeln_digitalisierung.pdf Stand: 15.10.2009]. S. 11 „Unterhalb einer Genauigkeit von 99,5 % ist bei manueller Erfassung ein Ergebnis im Grunde wertlos. Anders ausgedrückt, bei 99% wäre jeder 100ste Buchstabe(!) falsch, pro Zeile also ca. ein Fehler.“

⁹ Einen ausführlichen Vergleich von OCR und ‚double-keying‘ enthält Kapitel III ‚Texterfassung‘ der Dissertation von Stanislava Grigorova, Studien zur Digitalisierung von Campes ‚Verdeutschungswörterbuch‘, Phil. Diss. Universität Würzburg, Juni 2009 (im Druck).

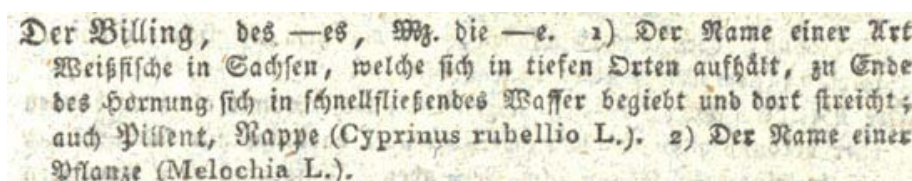
¹⁰ In einer älteren Studie zu Arbeiten für ARTFL berichten Olsen/McLean über Kosten und Nutzen des Einsatzes von OCR im Vergleich zu manueller Doppelerfassung bei der Erstellung hochwertiger Korpustexte mit dem Qualitätsanspruch „Less than one error every ten pages of running text“ Mark Olsen / Alice Music Mc Lean, Optical Character Scanning: A Discussion of Efficiency and Politics, in: Computers and the Humanities 27 (1993) S. 124.

¹¹ Vgl. Hanr-Werner Bartz/ Thomas Burch/ Ruth Christmann/ Kurt Gärtner/ Vera Hildenbrand/ Thomas Schares/ Klaudia Wegge, Wie das Deutsche in den Computer kam, in: Der digitale Grimm, Begleitbuch. Frankfurt/Main 2004, S. 73 - 90, hier S. 75.

unabhängig davon, auf welche Weise sie geschieht – die Positionierung zusätzlicher Kennzeichnungen im Wörterbuchtext:



\$0535.102 <negEZ><W+2>Billigen,</W+2> <A>v. trs. 1) Für billig erkennen, erklären. Solche Bedrückun=
\$0535.103 gen sind nicht zu <W+1>billigen.</W+1> Das Verfahren kann man unmöglich <W+1>bil=
\$0535.104 ligen. 2)</W+1> Bewilligen, Genehmigen. Die Ältern wollten diese Hei=
\$0535.105 rath nicht <W+1>billigen</W+1>. <W+1>Das Billigen. Die Billigung</W+1>.</negEZ>



\$0535.117 <negEZ><W+2>Der Billing,</W+2> <W+1>des -- es, Mz. die -- e. 1)</W+1> Der Name einer Art
535.118 Weißfische in Sachsen, welche sich in tiefen Orten aufhält, zu Ende
\$0535.119 des Hornung sich in schnellfließendes Wasser begiebt und dort streicht;
\$0535.120 auch <W+1>Pillent, Rappe (</W+1><A>Cyprinus rubellio L.). 2) Der Name einer
\$0535.121 Pflanze (<A>Melochia L.).</negEZ>

Richtet man die Kennzeichen an den Buchstaben aus und ignoriert dabei Interpunktion und Ordnungselemente, wird die Erkennung von Strukturmerkmalen dadurch erschwert. Im Stichwort 'Billigen' wird an der markierten Stelle die Kennzeichnung für 'Ende der Antiqua-Schrift' statt nach „trs.“ erst nach „1)“ eingesetzt. Dadurch wird die Gliederung des Wortartikels in Formbeschreibung (Wortklasse) und Bedeutungsbeschreibung gestört. Ähnlich wirkt sich die Verschiebung der Schriftgrößenmarkierung „</W+1>“ nach „2)“ auf die Trennung der Belegzitate von der Bedeutungsbeschreibung aus. Ferner sollten Satzzeichen nicht in die Schriftgrößenmarkierung im Stichwort ‚billigen‘ einbezogen werden: „<W+2>Billigen</W+2>“,“. Dies gilt in gleicher Weise für die Markierung der Antiqua-Schrift im Beispiel ‚Der Billing‘.

Unter Berücksichtigung von Ordnungselementen und Interpunktion – Worttrennung am Zeilen-, Spalten- oder Seitenende kann später ausgeblendet werden – sollten die zusätzlichen Kennzeichen immer möglichst eng am Text mitgeführt werden und – sofern mehr als eine Kennzeichnung anzubringen ist, sollte die Reihenfolge am Anfang wie am Ende immer in der gleichen Weise von innen nach außen erfolgen. Verschränkte Überschneidungen, womöglich noch über die einzelne Zeile hinaus, behindern nachfolgende Verarbeitungsschritte.

4.4 Qualitätskontrolle

Nach Erfassung der Seiten 1 – 500 von Band 1 des ‚Wörterbuchs der Deutschen Sprache‘ wurde ein Qualitätstest durchgeführt, in dem auf jeder Seite in der linken Spalte die Zeile 10, in der rechten Spalte die Zeile 110 mit dem Druck verglichen wurde. Bei einer durchschnittlichen Zeilenlänge von 77 Zeichen (einschließlich Leerzeichen und Textmarkierungen, jedoch ohne die Seiten-/Zeilenzählung) und 1000 überprüften Zeilen entspricht dies einer Genauigkeit von 99,996%, anders ausgedrückt: deutlich weniger als ein Fehler auf 10.000 Zeichen. Für jeden Band werden alle Korrekturen und Änderungen an dem Wortlaut des Wörterbuchtextes zur Qualitätssicherung, ggf. auch für nachträglich nötige Korrekturen an den Basisdaten, in eine Log-Datei im XML-Format eingetragen mit Angabe der Textreferenz, der vollständigen Originalzeile, der korrigierten Textzeile, den Namenskürzeln der für

die Änderung Verantwortlichen, dem Datum der Korrektur und der Ursache des Eintrags:

```
<change>
<source xml:id="0616.75_1">
  <![CDATA[^$0616.75 <shi>Nur dort, wo Klodion in goldnen Waffen <W+1>brennet</W+1>. ^#s+Alringer^#s-.]]>
</source>
<substitution xml:id="0616.75_2" who="#mb" date="2008-04-23">
  <![CDATA[^$0616.75 <shi>Nur dort, wo Klodion in goldnen Waffen <W+1>brennet</W+1>. ^#s+Alxinger^#s-.]]>
  <reason>Erfassungsfehler</reason>
</substitution>
</change>
```

Nach dem vorläufigen Abschluss der Arbeiten an den Wörterbuchtexten kann damit empirisch der Korrekturaufwand des Erfassungsverfahrens bestimmt und die Fehlerrate sowie die Genauigkeit des Textes ermittelt und weiter verbessert werden. Gleichzeitig bieten sich die Daten dann auch für den Test mit anderen Erfassungsverfahren an.

5. Vorarbeiten für die Textkodierung

Der angelieferte Datenerfassungstext wird unverändert archiviert und anschließend in vier Verarbeitungsschritten für die elektronische Textkodierung aufbereitet. In einem ersten Schritt werden die Interpunktionszeichen und Ordnungselemente, die bei der Texterfassung in die typographische Kennzeichnung mit einbezogen wurden, frei gestellt. Anschließend wird eine Erfassungskorrektur durchgeführt. Die bei der Erfassung als vermuteter Druckfehler, sicherer Druckfehler oder nicht lesbarer Buchstabe gekennzeichneten Textstellen werden am Drucktext überprüft und, sofern nötig, korrigiert. In dem dritten Korrekturschritt Druckkorrektur werden die in den verschiedenen Bänden meist im Nachspann unter ‚Druckfehler und Berichtigungen‘ mitgeteilten Verbesserungen in den Text eingetragen.

5.1 Anpassung der Textauszeichnungen

Bei der manuellen Erfassung in China wurden schließende Interpunktionszeichen, wie z.B. Klammern, Punkte, Kommata mit in die Auszeichnung der Typografie gezogen. Dies betrifft die Markierung von Antiqua-Schrift (<A>), von Größenunterschieden der Frakturschrift (<W+1> und <W+2>) sowie von gesperrtem Text (#s-). Für die elektronische Textkodierung muss die Kodierung eng um die so gekennzeichneten Textelemente gesetzt werden. Interpunktionszeichen und Ordnungselemente werden damit nicht markiert.

Folgende Arbeitsschritte wurden durchgeführt: eine manuelle Bereinigung der Kodierung sofern Zeilenumbrüche mitbetroffen waren; die Umsetzung der Kodierung mit Hilfe einer Prozedur für die Fallgruppen mit </W+2> (Hervorhebung des Stichworts im Wörterbucheintrag), </W+1> (Hervorhebung des Stichworts z.B. in Kontextbelegen und Zitaten innerhalb des Wörterbucheintrags), Kennzeichnung der Antiqua-Schrift, z.B. in lateinischen Wörtern und #s- (Kennzeichnung von gesperrter Schrift, z.B. in Namen, Werktiteln oder der Referenz von Bibelzitaten).

Von dieser Auszeichnungsbereinigung ausgenommen sind Abkürzungspunkte, die zusammen mit den Buchstabenkürzeln eine Einheit bilden und deshalb in der elektronischen Kodierung in einem Element zusammengefasst werden; ferner typographisch ausgezeichnete Mehrworteinheiten z.B. in Aufzählungen oder Phrasen, die zwar durch Interpunktionszeichen getrennt sind, die in der elektronischen Kodierung jedoch ebenfalls zusammen mit einem Element markiert werden. Schließlich wurden die bei der Erfassung mit in die Anfangsmarkierung <W+2> gezogenen

diasystematischen Angaben (z.B. * für ‚Alte aber brauchbare und zum Theil auch von Neuern gebrauchte Wörter‘ oder ** für ‚Veraltete Wörter‘), meist in Form von Sonderzeichen, und Indizes vor Stichwörtern, z. B. bei der Nummerierung von Homonymen vor die Anfangskennzeichnung <W+2> des Stichworts umgesetzt und mit Ausnahme von ;**‘ in Kombinationen von Sonderzeichen oder in der Kombination Index und Sonderzeichen einheitlich durch ein Leerzeichen getrennt.

5.2 Erfassungskorrektur

Mit dem zweiten Korrekturschritt werden die Korrekturmarkierungen, die von den Datentypistinnen bei der Texterfassung oder beim Vergleich der beiden Fassungen automatisch in den Text eingefügt wurden, an dem Drucktext überprüft und bereinigt. Es handelt sich um die Kennzeichnung nicht lesbarer oder nicht eindeutiger Buchstaben {??}, um Wortformen bei denen ein Druckfehler vermutet wird {?K} und Buchstaben, die sicher Druckfehler sind {!D}, z.B. Zwiebfische. In Fallzahlen sind das im ersten Band des Wörterbuchs 104 nicht eindeutig lesbare Stellen {??}, 26 Druckfehler {!D} und 30 vermutete Druckfehler {?K}. Sofern der Wortlaut des gedruckten Textes an den gekennzeichneten Stellen geändert werden muss, wird die Änderung in der Log-Datei des entsprechenden Bandes festgehalten. Der Text nach Ausführung der Erfassungskorrekturen wird archiviert.

5.3 Druckfehlerkorrektur

In dem Arbeitsschritt Druckfehlerkorrekturen werden die im Vorspann- bzw. Nachspanntext mitgelieferten ‚Druckfehler und Berichtigungen‘ im Wörterbuchtext korrigiert. Die Korrekturen werden für jeden Band in der Log-Datei verzeichnet.

Für den ersten Band ist das in Band 1 S. XXIV ‚2. Druckfehler und Berichtigungen‘ (ohne Paginierung), in Band 2 S. 1117 – 1118 ‚2.1 Im ersten Theile ist noch Folgendes zu berichtigen:‘ (ohne Paginierung), in Band 3 S. 908 ‚1. Druckfehler und Verbesserungen. 1.1 Im ersten Theile ist noch Folgendes zu berichtigen‘, in Band 4 S. 941 ‚3. Außer der schon verschiedene Mahl gegebenen Berichtigun-/gen und Zusätzen zu den frühern Theilen dieses /Werkes sind noch folgende nachzuholen. 3.1 Zum ersten Theile.‘ und in Band 5 S. 978 ‚2. Nachtrag zu den Berichtigungen der vier ersten Theile. 2.1 Zum ersten Theile.‘

Für den zweiten Band ist das in Band 2 S. 1118 ‚2.2 Im zweiten Theile‘ (ohne Paginierung), in Band 3 S. 908 ‚1. Druckfehler und Verbesserungen. ... 1.2 Im zweiten Theile.‘, in Band 4 S. 941 ‚3. Außer der schon verschiedene Mahl gegebenen Berichtigun-/gen und Zusätzen zu den frühern Theilen dieses /Werkes sind noch folgende nachzuholen. ... S. 942 - 943 3.2 Zum zweiten Theile.‘ und in Band 5 S. 978 ‚2. Nachtrag zu den Berichtigungen der vier ersten Theile. ... S. 979 2.2 Zum zweiten Theile.‘

Für den dritten Band ist das in Band 3 S. 908 ‚1. Druckfehler und Verbesserungen. ... 1.3 Im dritten Theile.‘, in Band 4 S. 941 ‚3. Außer der schon verschiedene Mahl gegebenen Berichtigun-/gen und Zusätzen zu den frühern Theilen dieses /Werkes sind noch folgende nachzuholen. ... S. 943 – 944 3.3 Zum dritten Theil.‘ und in Band 5 S. 978 ‚2. Nachtrag zu den Berichtigungen der vier ersten Theile. S. 979 2.3 Zum dritten Theile.‘

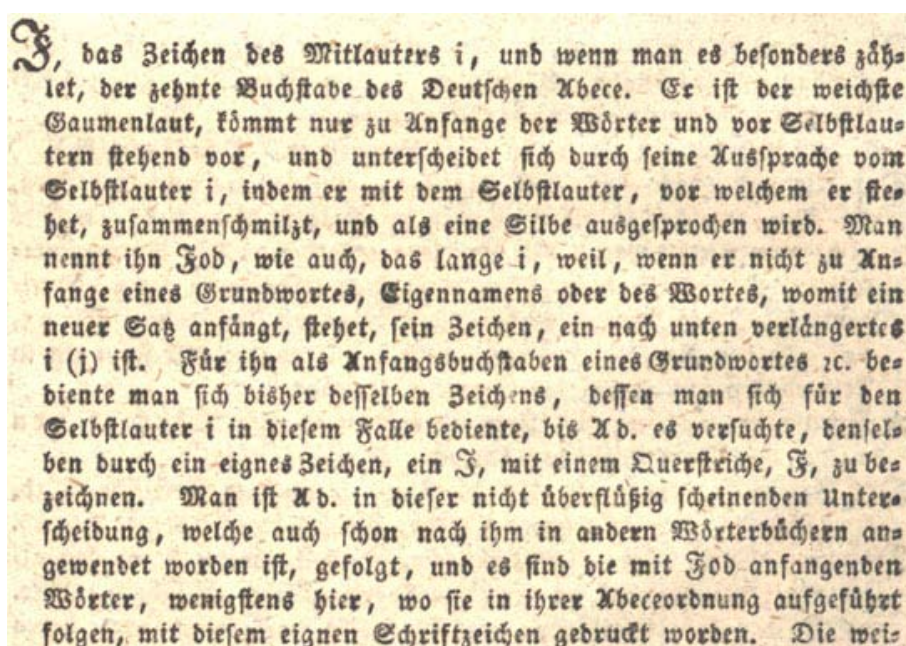
Für den vierten Band in Band 4 S. 941 ‚2. Berichtigungen *) und Zusätze. (ohne Paginierung), S. 941 Zum vierten Theile.‘ und in Band 5 S. 978 ‚2. Nachtrag zu den Berichtigungen der vier ersten Theile. S. 979 2.4 Zum vierten Theile.‘

Für den fünften Band in Band 5 S. 978 1. Berichtigungen *) und Zusätze. (ohne Paginierung) Zum fünften Theile’.

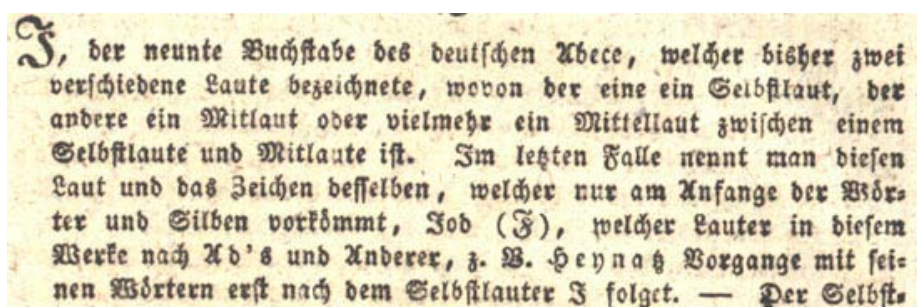
Für das ‚Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke’ werden Korrekturen auf den Seiten S: 611 - 612 unter ‚Druckberichtigungen.’ (ohne Paginierung) mitgeteilt.

5.4 Anpassung der Fraktur-Typographie

Einige typographische Besonderheiten der Frakturschrift wurden für die Digitalisierung an die heute übliche elektronische Kodierung deutschsprachiger Texte angepasst. Dazu gehört zunächst die Differenzierung des Fraktur-Großbuchstabens J in die Buchstaben I und J. Im ‚Wörterbuch der Deutschen Sprache’ verwendet Campe innerhalb des Wörterbuchtextes J unterschiedslos für I bzw. J. Lediglich im Stichwortbereich J wird der Großbuchstabe, eine Anregung Adelungs aufgreifend, durch einen zusätzlichen Querstrich im Vergleich zum Großbuchstaben I gekennzeichnet.



Wörterbuch der Deutschen Sprache. Zweiter Theil. F - bis – K. Braunschweig 1808, S. 833



Wörterbuch der Deutschen Sprache. Zweiter Theil. F - bis – K. Braunschweig 1808, S. 814

Die Umsetzung erfolgt regelgerecht durch eine Prozedur, die J vor Konsonant in I konvertiert und vor

Vokal als J belässt. In der Abkürzung ‚J.‘ in Namen wurde die volle Namensform ermittelt und entsprechend verfahren. Die Konvertierung wird pauschal in der Log-Datei festgehalten, Einzelfälle werden nicht verzeichnet.

Die je nach Umgebung und Vorkommen unterschiedliche typographische Realisierung von s als s bzw. f wurde zu s vereinheitlicht. Ebenso werden Ligaturen in Einzelbuchstaben getrennt.

5.6 Abkürzungen, Namen, Werktitel und Bibelstellen

Im Vorfeld der Textkodierung wird eine Liste mit Abkürzungen erstellt, die dazu genutzt wird, über eine Prozedur alle Abkürzungen einschließlich der Abkürzungspunkte mit der Markierung ‚<abbr> ... </abbr>‘ einzuklammern, die dazu beiträgt, Satzendepunkte und damit Beschreibungsstrukturen im Wörterbuch besser eingrenzen zu können. Diesem Ziel dient auch die Vorabmarkierung von Namen, Werktiteln und Referenzen auf Bibelstellen, insbesondere in Verbindung mit Textzitaten zur Verdeutlichung von Wortbedeutungen. Damit kann auch gleichzeitig dafür gesorgt werden, dass Schreibvarianten und Kurzformen für die gleiche Person oder das gleiche Werk richtig zugeordnet werden.

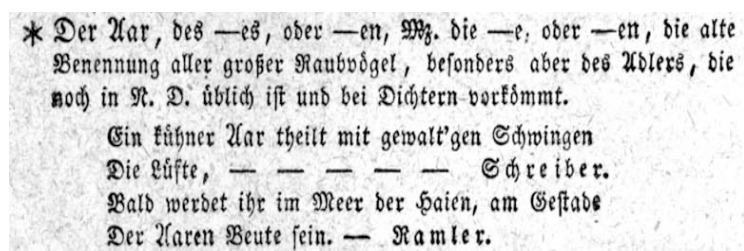
5.5 Fehlerkorrekturen im elektronischen Text

Mit der Anpassung der Textauszeichnungen an die intendierten Markierungsstrukturen, der Korrektur von Fehlern, die bei der Texterfassung festgestellt werden und der Korrektur der Fehler, die nach dem Druck der Bände noch von den Autoren selbst ermittelt wurden, liegt für jeden Band ein Basistext für die weitere Arbeit vor, der zum einen planmäßig in den fehleranfälligen Buchstabenkombinationen der Frakturschrift auf weitere Fehler überprüft wird und in dem auch alle bei den weiteren Arbeitsschritten entdeckten Fehler korrigiert werden. Diese Arbeit an der weiteren Minimierung von Fehlern im elektronischen Text ist derzeit noch nicht abgeschlossen.

Alle Eingriffe in den Wortlaut des ursprünglichen Drucktextes werden dabei für jeden Band über die Log-Datei verwaltet, die damit die verschiedenen Schichten von Textkorrekturen und Eingriffen in den Text für jeden einzelnen Fall dokumentiert.

6. Lexikographische Analyse und Kodierungskonzept

Die lexikographische Analyse eines einfachen Wörterbucheintrags, des Substantivs ‚Der Aar‘, soll das Kodierungskonzept verdeutlichen, das in einem ersten Schritt für die Digitalisierung der Wörterbücher zugrunde gelegt wird.



Joachim Heinrich Campe, Wörterbuch der Deutschen Sprache. Erster Theil. A - bis - E. Braunschweig 1807, S. 2

Der Wörterbucheintrag beginnt mit einer diasystematischen Kennzeichnung der zeitlichen Einordnung,

also einer Gebrauchbedingung: * steht für ‚Alte aber brauchbare und zum Theil auch von Neuern gebrauchte Wörter‘. Das Lemma ‚Der Aar‘ beginnt mit dem Artikel, der gleichzeitig die Wortklasse ‚Substantiv‘ und das Genus ‚maskulin‘ festlegt. Es folgen Angaben zur Flexion: die Doppelformen für den Genitiv Singular ‚-es oder -en‘, und den Nominativ Plural ‚-e oder -en‘, zeigen an, dass der Wechsel des Flexionsparadigmas – in der Grimmschen Nomenklatur von ursprünglich ‚schwach‘, nach ‚stark‘ um 1800 noch nicht abgeschlossen ist. Auf die Formen folgt die Bedeutungsbeschreibung der Bedeutung, kombiniert mit Hinweisen auf räumliche Geltung (‚in N. D.‘ steht für ‚Niederdeutsch‘) und Stilschicht (‚bei Dichtern‘) ‚dichterisch‘. Zwei Kontextbeispiele aus Gedichten von Schreiber und Ramler schließen den Wortartikel ab.

Lemmazeile mit grammatischen	* Der Aar,
Kennzeichen -----	des —es, oder —en, Pl. die —e, oder —en,
Formbeschreibung -----	
Bedeutungsbeschreibung -----	die alte
mit	Benennung aller großer Raubvögel, besonders aber des Adlers, die
Gebrauchsbedingungen -----	noch in N. D. üblich ist und bei Dichtern vorkommt.
Kontextbeispiele -----	Ein kühner Aar theilt mit gewalt'gen Schwingen
aus	Die Lüfte, — — — — — Schreiber.
poetischen	Bald werdet ihr im Meer der Haien, am Gestade
Texten -----	Der Aaren Beute sein. — Ramler.

Am Eintrag ‚Der Aar‘ lässt sich die grundlegende Strukturrahmen zur Kodierung des Wörterbuchs gut erläutern: Als kleinste Einheit gilt der Wörterbucheintrag <entry>, mit den beiden Kernbestandteilen <form> für die Beschreibung morphologischer und grammatischer Eigenschaften und <sense> für die Beschreibung von Bedeutung und Gebrauchsbedingungen, feiner differenzierbar in Definitorisches und Textbeispiele, hier in Gestalt literarischer Belege, die die Bedeutungsbeschreibung verdeutlichen. Daraus ergibt sich folgendes Schema, auf der Basis der Elementbezeichnungen der TEI für Wörterbücher:

```

-----<entry>
|   -----<form>
|   |   <form> ... </form>
|   -----</form>
|   -----<sense>
|   |   <def> ... </def>
|   |   <cit> ... </cit>
|   |   <cit> ... </cit>
|   -----</sense>
|   -----</entry>

```

Eingefügt in den umfassenden Strukturrahmen, wie ihn die ‚Guidelines‘ der TEI in der Version P5 auf der Basis von XML und Schemakoncepten für validierbare elektronische Texte vorsehen, führt dies zu folgender Kodierung von Einzeldateien für jeden Wörterbucheintrag:

```

<TEI ...><teiHeader> ... </teiHeader><text><body><div>

                                <entry xml:id="cwds1_000045_aar"> ... </entry>

                                </div></body></text>

</TEI>

```

7. Definition des TEI-Schemas

Voraussetzung dafür, dass der erfasste elektronische Wörterbuchtext zu validen TEI-Dateien verarbeitet werden kann, ist die Erstellung eines TEI-Schemas, das genau die Dateistrukturen für das XML-Dateiformat vorgibt, nach denen jede einzelne Wörterbuchdateien während des Kodierungsverfahrens validiert werden kann. Als Werkzeug zur Generierung dieses Schemas steht das Programm ‚Roma‘ auf der Webseite der TEI zur Verfügung¹². Es erlaubt auf Grund des modularen Aufbaus der TEI-Rahmenrichtlinien, die Kodierungen an die Gegebenheiten und Erfordernisse des Kodierungsprojekts anzupassen.

Für die Kodierungen des Campe-Wörterbuchs werden Elementen aus den Modulen ‚core‘, ‚tei‘, ‚header‘, ‚textstructure‘, ‚analysis‘, ‚dictionaries‘, ‚gaiji‘ und ‚transcr‘ benötigt. Es müssen jedoch nicht alle in den jeweiligen Modulen definierten Elemente übernommen werden. Vielmehr können mit ‚Roma‘ nur diejenigen Elemente ausgewählt werden, die in dem Projekt auch zur Anwendung kommen. Als Resultat des Generierungsprozesses entsteht dabei nicht nur das für das Projekt angepasste Schema als XML- oder RelaxNG-Schema. Es kann auch eine XML-Datei erzeugt werden, die sowohl genau die getroffene Auswahl dokumentiert, als auch dazu verwendet werden kann, ein Schema nach gleichem Muster wieder zu generieren oder das generierte Schema an veränderte Bedingungen anzupassen.

Die Beschreibung des für die Zwecke dieses Berichts neu generierten Schemas ist als Anhang 2 beigelegt.

8. Textkodierung

Ausgangspunkt für die Textkodierung von Campes Wörterbüchern ist zum einen der elektronisch erfasste Text der gedruckten Bände, zum anderen die in dem TEI Schema verankerte Strukturdefinition. Der elektronische Basistext für die Textkodierung des Wörterbucheintrags ‚Der Aar‘ hat dabei folgende Form:

```
$0002.35 <negEZ>*<W+2> Der Aar</W+2>, <W+1>des -- es</W+1>, oder -- <W+1>en,
<abbr>Mz.</abbr> die -- e</W+1>, oder -- <W+1>en</W+1>, die alte
$0002.36 Benennung aller großer Raubvögel, besonders aber des Adlers, die
$0002.37 noch <geo>in <abbr>N. D.</abbr></geo> üblich ist und <style>bei
Dichtern</style> vorkommt.</negEZ>
$0002.38 <shi>Ein kühner <W+1>Aar</W+1> theilt mit gewalt'gen Schwingen
$0002.39 Die Lüfte, -- -- -- -- -- <author>^#s+Schreiber^#s-</author>.
$0002.40 Bald werdet ihr im Meer der Haien, am Gestade
$0002.41 Der <W+1>Aaren</W+1> Beute sein. -- <author>^#s+Ramler^#s-
</author>.</shi>
```

Jeweils am Zeilenanfang, markiert durch ‘\$’-Zeichen, steht die Seiten- und Zeilenreferenz innerhalb des jeweiligen Bandes, die Kodierung <negEZ> steht für ‚negativen Einzug‘ und markiert damit ein ausgerücktes Stichwort, danach folgt mit dem * eine Kennzeichnung der zeitlichen Zuordnung des Stichworts, die typographische Markierung <W+2> signalisiert, dass hier ein Stichwort beginnt, da nur

¹² <http://www.tei-c.org/Roma/>

Stichwörter in einer zwei Punkt größeren Schrift gesetzt sind. Die Formbeschreibung des Stichworts erfolgt in einer einen Punkt größeren Schrift und erlaubt es damit hier, über typographische Merkmale die Grenze zwischen Formbeschreibung und Bedeutungsbeschreibung zu bestimmen. Die Belegzitate sind einerseits durch Einrückungen `<shi>...</shi>` gekennzeichnet, andererseits durch Autorennamen in gesperrter Schrift (`#s+ ... #s-`), die bei der Vorbereitung bereits mit dem Element `<author> ... </author>` markiert wurden. Hinzukommt, dass das Stichwort in Zitaten in einer einen Punkt größeren Schrift `<W+1>` hervorgehoben wird. Mit diesen typographischen Indizien sind die Kontextbeispiele und ihre Autoren eindeutig zu identifizieren. Die räumliche Zuordnung „in N. D.“ und die Stilschicht „bei Dichtern“ ist über Listen eingesteuert.

Aufgabe der am Projekt beteiligten Informatiker war es, diese typographischen Informationen und die Rückschlüsse auf die Textumgebung, die man aus ihnen ableiten kann, in eine Transformation zu überführen, die valide XML-Strukturen nach den Vorgaben des TEI-Schemas erzeugt. Erreicht wurde dies mittels eines auf der Programmiersprache Prolog basierenden Programmierkonzepts, das Dietmar Seipel entwickelt hat und das von ihm und von Christian Schneider erfolgreich auf die Campe-Daten angewendet werden konnte. Einige Code-Fragmente sollen einen Einblick in die Programmabläufe vermitteln und verdeutlichen, wie die Transformationsregeln definiert sind, mit denen Eintragsstrukturen (`entry` und `headword`), grammatische Kennzeichnungen (`form_determiner`, `form_headword`, `form_inflected`), aber auch lineare Textstrukturen (`campe_linebreak`) und Sonderzeichen (`campe_punctuation`) bearbeitet werden¹³.

8.1 Implementierung in Prolog

Im folgenden werden einige der zentralen Programmierkonzepte für die Annotation des Campe Wörterbuchs vorgestellt. Prolog-Programme sind meist etwa 5 bis 10-mal kompakter als entsprechende Java-Programme, was sich in der Anzahl der Codezeilen des neu entwickelten Programmes widerspiegelt (1.000 Programmzeilen). Hinzu kommt die zugrunde liegende Prolog-Bibliothek DDK (DisLog Developers' Toolkit, 150.000 Programmzeilen) von Dietmar Seipel, aus der speziell die mächtige XML-Anfrage-, Transformations- und Update-Sprache FnQuery (etwa 6.000 Programmzeilen) verwendet wurde. Diese benutzt denselben Pfadformalismus wie die Standard-Sprachen XQuery (Anfragen) und XSLT (Transformationen) und erweitert deren Funktionalität um wesentliche Features.

Die Entwicklungszeit für Prolog-Programme liegt weit unter der für die entsprechenden Java-Programme. Allerdings sollte man beachten, dass aufgrund der Kompaktheit und der mathematischen/logischen Ausdrucksweise von Prolog (Programming in Logic) der Aufwand zum Schreiben einer einzelnen Programmzeile natürlich etwas höher liegt als bei Java, da ja eine Prolog-

¹³ Die Arbeiten wurden im Mai und im August auf zwei internationalen Tagungen vorgestellt: Christian Schneider, Dietmar Seipel, Werner Wegstein, Klaus Prator: Declarative Parsing and Annotation of Electronic Dictionaries, in: Bernadette Sharp, Michael Zock (Eds.), Natural Language Processing and Cognitive Science. Proceedings of the 6th International Workshop on Natural Language Processing and Cognitive Science – NLPCS 2009 (in conjunction with ICEIS 2009), Milan, Italy, May 2009, S. 122 – 132.

Christian Schneider, Dietmar Seipel, Werner Wegstein: Schema and Variation: Retro-Digitizing Printed Dictionaries, in: Proceedings of the Third Linguistic Annotation Workshop, sponsored by the Association for Computational Linguistics Special Interest Group for Annotation (ACL-SIGANN), (in conjunction with the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing), 2. – 7. August, Singapore 2009, S. 82 – 89. [url = <http://www.aclweb.org/anthology/W09-3014>].

Eine weitere Präsentation ist für die Jahrestagung der TEI 2009 in Ann Arbor als Vortrag angenommen worden:

Christian Schneider, Dietmar Seipel, Werner Wegstein (University of Würzburg): Declaratively Creating and Processing XML/TEI Data. Paper accepted 10 July 2009 for the TEI Members Meeting, November 2009, Ann Arbor (USA).

Programmzeile etwa 5 bis 10 Java-Programmzeilen entspricht.

Neben der kürzeren Entwicklungszeit hat die kompakte Darstellung auch noch den wesentlichen Vorteil, dass sie die Änderungsfreundlichkeit und die Anpassbarkeit der Programme an neue oder verfeinerte Anforderungen deutlich verbessert.

8.1.1 Benutzerschnittstelle

Die Benutzerschnittstelle der Implementierung ist das Kommandozeilen-Prädikat `campe_annotate_file`, welches eine Eingabedatei mit vielen Wörterbucheinträgen in je eine annotierte TEI-Datei mit genau einem Wörterbucheintrag transformiert. Die Pfade der Ausgabedateien bestehen alle aus einem Präfix, das vom Benutzer für zusätzliche Pfadangaben genutzt werden kann, gefolgt vom Identifikator des Wörterbucheintrages und der Endung ".xml".

```
campe_annotate_file(File, Prefix) :-  
    campe_read_file(File, Campe),  
    forall( campe_find_entry(Campe, Entry_1),  
        ( campe_annotate_entry(Entry_1, Entry_2),  
          Id := Entry_1/attribute::id,  
          concat([Prefix, Id, '.xml'], Path),  
          dwrite(xml, Path, Entry_2) ).
```

Nachdem die Eingabedatei eingelesen wurde (`campe_read_file`), werden in einer Schleife der Reihe nach die Wörterbucheinträge gefunden (`campe_find_entry`), annotiert (`campe_annotate_entry`) und in je eine Ausgabedatei geschrieben (`dwrite`).

```
campe_find_entry(Campe, Entry) :-  
    member(Tag, [tmpEntry, 'P']),  
    Entry := Campe/descendant::Tag,  
    'W_2' := Entry/nth_child::1/tag::' '*.
```

Für die Identifikation der vorannotierten Strukturen durch den Prolog-Parser, wurden im Vorfeld bestehende Tags in valide XML-Elemente umgeformt (z.B. 'W+2' ---> 'W_2')

8.1.2 Annotation einzelner Wörterbucheinträge

Das zentrale Prädikat `campe_annotate_entry` zur Annotation findet zuerst das Lemma eines Eintrages (Headword) und wendet daraufhin eine Reihe von Transformations-Schritten (`campe_transform`) für eine fein granulierte Auszeichnung an.

```
campe_annotate_entry(Entry_1, Entry_2) :-  
    campe_parse_headword_content(Entry_1, Headword),  
    apply_all(Entry_1, Entry_2, [  
        campe_transform(headword, Headword),  
        campe_transform(inflected),  
        campe_transform(form_headword),  
        campe_transform(form_determiner),  
        campe_transform(sense),  
        campe_transform(antiqua),  
        campe_transform(large),  
        campe_transform(linebreak),  
        campe_transform(punctuation) ]).
```

Das Headword wird aus dem ersten Kind-Element des zu untersuchenden Wörterbucheintrags gebildet und als form-Element in TEI repräsentiert.

```
campe_parse_headword_content(Entry, Headword) :-
    Content := Entry/nth_child::1/content::'*,
    entry_split_list_at_position([" "], Content, Ls),
    flatten(Ls, Ls_1),
    campe_parse_headword(Headword_1, Ls_1, []),
    flatten(Headword_1, Headword_2),
    !,
    Headword = form:[]:Headword_2.
```

Das Prädikat ‚campe_parse_headword‘ findet eine Folge von Elementen (form_) im zu untersuchenden Wörterbucheintrag; diese können Tags, Unicode-Zeichen oder form-Elemente sein. Dazu werden Definite Clause Grammars (DCG) aus Prolog benutzt, welche die Regeloperatoren "-->" und "==">" verwenden.

```
campe_parse_headword(Xs) -->
    sequence('*', form_, Xs).

form_(Xs) -->
    ( tag(Xs)
      ; unicode(Xs), !
      ; form(Xs) ).

tag([T:As:Es]) --> [T:As:Es].

unicode ==>
    [X], { atomic(X), campe_is_unicode_char(X) }.
```

Speziell die für TEI erzeugten form-Elemente können eine äußerst variable Struktur haben; deswegen müssen sehr viele alternative Formen angegeben werden, die in Prolog mit ";" (für oder) verknüpft werden.

```
form:[type:lemma] ==>
    ( form_determiner, conjunction, form_determiner,
      form_headword, !
      ; form_determiner, form_determiner, form_determiner,
      form_headword, !
      ; form_determiner, form_determiner, form_headword, !
      ; form_determiner, form_headword, !
      ; form_headword
      ; [X], { X = _:_:_ } ).
```

Die Erkennung der einzelnen Komponenten mithilfe der Prolog-Prädikate ‚form_determiner‘, ‚conjunction‘ und ‚form_headword‘ kann hier aus Platzgründen nicht näher beschrieben werden. Bereits die DCG-Regeln von Prolog stellen eine entscheidende Hilfe bei der kompakten Implementierung von Parsern dar, da der Ein- und Ausgabestrom generisch behandelt wird und unterdrückt werden kann, was die Les- und Wartbarkeit des erzeugten Programmcodes stark vereinfacht.

Eine weitere Vereinfachung, die Extended Definite Clause Grammars (EDCG) wurde von Dietmar Seipel entwickelt. Hier können die erzeugten XML-Elemente ebenfalls generisch behandelt und somit unterdrückt werden kann.

Dadurch reduziert sich der zu schreibende Prolog-Code auf etwa ein Drittel im Vergleich zu Standard-Prolog. Die kompaktere Schreibweise von Standard-Prolog im Vergleich zu prozeduralen oder objekt-orientierten Sprachen (wie etwa C, C++ oder Java) kann somit noch weiter vereinfacht werden. Die erzeugten Parser bieten durch diese Art der Programmierung ein hohes Maß an

Flexibilität, da Programmteile durch deren einfache Lesbarkeit schnell modifiziert werden können. Dies war für die Programmentwicklung im Umfeld einer sich ab und zu zwangsläufig ändernden TEI-Spezifikation äußerst hilfreich.

8.1.3 Transformationsregeln

Es wurden verschiedene Transformationsregeln (`campe_transform`) entwickelt, die unterschiedliche Fälle abdecken, was eine Feinstrukturierung der annotierten Einträge ermöglicht.

```
campe_transform(headword, Headword, Entry_1, Entry_2) :-  
    fn_item_transform(--->(headword, Headword), Entry_1, Entry_2).  
campe_transform(Type, Entry_1, Entry_2) :-  
    member(Type, [sense, antiqua, large,  
        form_headword, form_determiner, linebreak]),  
    fn_item_transform(--->(Type), Entry_1, Entry_2).  
campe_transform(punctuation, Entry_1, Entry_2) :-  
    fn_item_transform(--->(punctuation), Entry_1, Entry_2),  
    fn_item_purify(Entry, Entry_2).  
campe_transform(inflected, Entry_1, Entry_2) :-  
    fn_item_transform(--->(inflected), Entry_1, Entry_2),  
    fn_item_transform(--->(inflected_), Entry, Entry_2).
```

Zu jedem Aufruf der Form `fn_item_transform(--->(…), Entry_1, Entry_2)` muss es eine entsprechende Prolog-Regel der Form `--->(…, Entry_1, Entry_2)` geben.

Die rekursive Anwendung der letzteren Prolog-Regeln wird vom Prädikat `fn_item_transform` aus dem DDK übernommen. Beispielhaft sind im folgenden die Transformations-Regeln für `headword`, `form_headword`, `form_determiner` und `sense` angegeben, die jeweils für die Transformation der Einträge in getrennte form- und sense-Blöcke, sowie die Auszeichnung der Artikel für Substantive notwendig sind.

```
--->(headword, Headword, X, Y) :-  
    Headword = _:_:_,  
    X = 'W_2':_:_:,  
    Y = Headword,  
    !.  
--->(headword, _, X, X).  
  
--->(form_headword, X, Y) :-  
    X = form_headword:_:Es,  
    Y = form:[type:headword]:[orth:Es],  
    !.  
  
--->(form_determiner, X, Y) :-  
    X = form_determiner:_:Es,  
    Y = form:[type:determiner]:[orth:Es],  
    !.  
  
--->(sense, X, Y) :-  
    X = tmpEntry:As_1:Es_1,  
    Es_1 = [form:[]:Cs|Sense],  
    Es_2 = [form:[]:Cs, sense:[]:Sense],
```

```
increment_counter(Counter),
[Headword] := X/descendant::form::[@type=headword]/orth/content::'*,
downcase_atom(Headword, H),
concat(['cwds1_', Counter, '_', H], ID),
append(As_1, [id:ID], As_2),
Y = entry:As_2:Es_2,
!.
--->(sense, X, Y) :-
X = 'P':As:[form:[]:Cs|Sense],
Y = entry:As:[form:[]:Cs, sense:[]:Sense],
!.
```

8.2 Das Resultat als TEI-Datei

Als Resultat der Textkodierung wird für jeden Wörterbucheintrag eine separate Datei erzeugt. Zur Veranschaulichung soll hier nur der in den Transformationsbeispielen behandelte Ausschnitt wiedergegeben werden:

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="de">
<teiHeader> ... ..
</teiHeader>
<text>
<body>
<div>
<entry xml:id="cwds1_00045_aar">
<lb n="0002.035"/>
<usg type="time"><g>*</g></usg>
<form>
<form type="lemma">
<gramGrp>
<pos value="noun"/>
<gen value="m"/>
</gramGrp>
<form type="determiner"><orth>Der</orth></form>
<form type="headword" norm="aar"><orth>Aar</orth></form>
</form>
<c>,</c>
<form type="inflected">
<gramGrp>
<case value="genitive"/>
<number value="singular"/>
</gramGrp>
<form type="determiner"><orth>des</orth></form>
<form type="headword"><orth><oVar><oRef/> - es </oVar></orth></form>
<c>,</c> oder
<form type="headword"><orth><oVar><oRef/> - en </oVar></orth></form>
</form>
<c>,</c>
<form type="inflected">
<gramGrp>
<case value="nominative"/>
<number value="plural"><abbr>Mz.</abbr></number>
```

```
</gramGrp>
  <form type="determiner"><orth>die</orth></form>
  <form type="headword"><orth><oVar><oRef/> - e </oVar></orth></form>
  <c>,</c> oder
  <form type="headword"><orth><oVar><oRef/> - en </oVar></orth></form>
</form>
<c>,</c>
</form>
<sense>
  ... ..
</sense>
</entry>
</div>
</body>
</text>
</TEI>
```

Die vollständigen Kodierungsdaten für das Stichwort ‚Der Aar‘ in aktualisierter Form enthält der Anhang 3, die detaillierte Beschreibung der Kodierung von Header und Wörterbucheintrag die ‚TextGrid-Kernkodierung‘¹⁴.

Freilich sind die lexikographischen Strukturen von Campes Wörterbüchern nicht durchgängig so klar strukturiert wie der Musterfall ‚Der Aar‘ und es sind auch noch weitere Beschreibungsaspekte zu berücksichtigen, die hier, wo es um die Beschreibung der Arbeitsschritte geht, nicht alle zur Sprache kommen sollten. Sie sind in Mirjam Blümms ‚Leitlinien zur Kodierung des Campe-Wörterbuchs in TEI P5‘ ausführlich beschrieben¹⁵: außer den Bereichen ‚form‘ und ‚sense‘ sind dies ‚cit‘, ‚name‘, ‚foreign‘, ‚xr‘, ‚note‘, ‚re‘ und ‚usage‘ sowie die Grundlagen der linearen Textstrukturen.

9. Präsentation

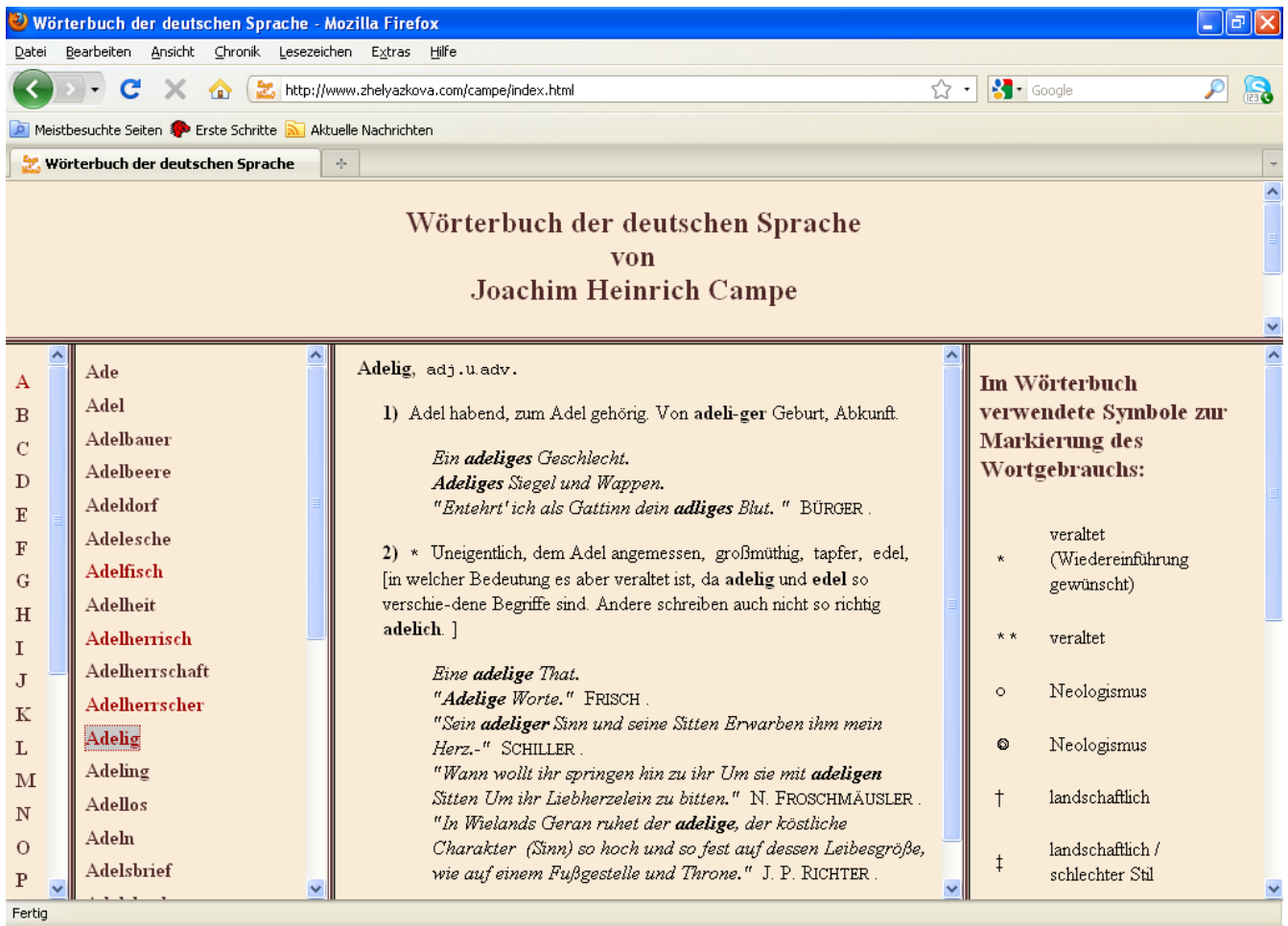
Ohne die Möglichkeit zu differenzierter Abfrage und zu freier Nachnutzung und Weiterverwendung der Daten sind retrodigitalisierte Texte wenig nützlich. Daher hat es von Anfang Überlegungen dazu gegeben, wie das Wörterbuchmaterial für eine möglichst umfassend nutzbare Präsentation aufbereitet werden könnte. Die Diskussion darüber ist innerhalb des TextGrid-Projekts in Verbindung mit dem Sammlungskonzept unter Integration der Metadaten noch nicht abgeschlossen. Parallel dazu haben aber Studierende in Seminaren zur Thematik ‚EDV-Philologie‘ über diese Fragen diskutiert und sich in Praktika im TextGrid-Projekt in Referaten, Präsentationen und Abschlussarbeiten damit auseinandergesetzt.

Ein Lösungsvorschlag, der es verdient, zum Abschluss als Modell für eine Präsentationsoberfläche vorgestellt zu werden, stammt von Nina Zhelyazkova¹⁶ und zeigt sehr überzeugend, wie die XML-Datenstrukturen für in einer Web-Umgebung weiter konvertiert werden können.

¹⁴ Link: http://www.textgrid.de/fileadmin/TextGrid/reports/Textsortenspezifische_Kernkodierung_2009.pdf

¹⁵ <http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/Leitlinien-zur-Kodierung-des-Campe-Woerterbuchs-in-TEI-P5.pdf>

¹⁶ Nina Zhelyazkova, Die Online-Darstellung von Joachim Heinrich Campes „Wörterbuch der deutschen Sprache“ (1807 – 1813). Abschlussarbeit im Aufbaustudiengang ‚Linguistische Informations- und Textverarbeitung‘, Würzburg 2008 (Betreuer Dr. Stephan Moser). Link: <http://www.zhelyazkova.com/campe/index.html> [Stand 19.10.2009]



Nina Zhelyazkova: : <http://www.zhelyazkova.com/campe/index.html>

10. Schlussbemerkung

Konzepte für die Digitalisierung und Richtlinien für die Textkodierung sind keine statischen Größen, sondern sie stehen für lebendige Entwicklung und permanente Verbesserung. Während der Arbeit an dem Projekt wurden beispielsweise die TEI-Richtlinien, auch und gerade für den Bereich ‚Wörterbücher‘ wiederholt grundlegend fortentwickelt, präzisiert und vereinfacht. Auch dieser Bericht über die Arbeitsschritte bedarf daher, wie die einzelnen Arbeitsvorgänge selbst ebenfalls permanenter Weiterentwicklung, Präzisierung und Vereinfachung.

11. Anhang

Anhang 1: Richtlinien für die manuelle Texterfassung
Anhang 2: XML-Datei zur Generierung des Schemas
Anhang 3: Textkodierung ‚Der Aar‘

BUCHSTABENMUSTER UND KODIERUNGSVORGABEN FÜR DIE TEXTERFASSUNG

Campe: Wörterbuch der deutschen Sprache &

Campe: Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke

Zeichensatz: UNICODE UTF-8 (soweit belegt)

Kodierung für komplexe Buchstaben mit diakritischen Zeichen wie in der Tabelle angegeben

Die Mustertranskription (schwarz) gilt jeweils nur für den rot markierten Buchstaben.

FRAKTUR		ANTIQUA	ANTIQUA
Großbuchstabe	Kleinbuchstabe	Großbuchstabe	Kleinbuchstabe

<u>Ans</u> age,	F <u>a</u> ckeln,	<u>A</u> dverbium	<u>a</u> dj. u. <u>a</u> dv.
A (A <u>n</u> sage)	a (F <u>a</u> ckeln)	A (A <u>d</u> verbium)	a (<u>a</u> dj. u. <u>a</u> dv.)

-----	ü <u>b</u> er <u>d</u> á <u>ch</u> en,	<u>Á</u> lb <u>u</u> m	Admir <u>a</u> ble
Á	á (Ü <u>b</u> er <u>d</u> á <u>ch</u> en)	À (Á <u>b</u> u <u>m</u>)	á (Admir <u>a</u> ble)

-----	-----	<u>Ā</u> cciden <u>z</u>	Admir <u>ā</u> l.
		Ā (Ā <u>a</u> cciden <u>z</u>)	ā (Admir <u>ā</u> l)

-----	-----		
			â

<u>Ä</u> rmel	B <u>a</u> um <u>s</u> ä <u>g</u> e	-----	<u>P</u> ä <u>d</u> ag <u>o</u> g
Ä (Ä <u>a</u> rmel)	ä (B <u>a</u> um <u>s</u> ä <u>g</u> e)	Ä	ä (Pä <u>d</u> ag <u>o</u> g)

-----	Durchächzen	-----	-----
%/Ä	%/ä (Durch%/ächzen)	%/Ä	%/ä

-----	-----	Æquivalíren.	æsthetischen læsæ Majestátis
		Æ (Æquivalíren)	æ(æstehthischen, læsæ Majestátis)
-----	-----	Ære	Et cætera
		Æ (Ære)	æ (Et cætera)

-----	-----	-----	Unitæt
		%-Æ	%-æ (Unit%-æt)

Baumsåge	Zabel,	Blumíren	Admiráble
B (Baumsäge)	b (Zabel)	B (Blumíren)	b (Admiráble)

Catel	Sociner	Carétte	Accouchíren
C (Catel)	c (Sociner)	C (Carétte)	c (Accouchíren)

-----	Kuckuck	-----	-----
	[Ligatur] ck (Kuckuck)		

-----	-----	-----	Façón
-------	-------	-------	--------------

		Ç	ç (Façon)
-----	Pfandlich,	-----	-----
	[Ligatur] ch (Pfandlich)		

Die Diele	Stau<u>d</u>e,	(Don<u>a</u>x	adj. u. adv.
D (Die Diele)	d (Stau <u>d</u> e)	D (Don <u>a</u> x)	d (adj. u. adv.)

Die E<u>b</u>ene,	Ver<u>d</u>ien<u>e</u>n,	El<u>a</u>stisch	Ad<u>v</u>erbium
E (Die E <u>b</u> ene,)	e (Ver <u>d</u> ien <u>e</u> n,)	E (El <u>a</u> stisch)	e (Ad <u>v</u> erbium)

-----	Ü<u>b</u>erb<u>e</u>ten	É<u>n</u>tre - p<u>o</u>t	Augm<u>e</u>nt
	é (Ü <u>b</u> erb <u>e</u> ten)	É (É <u>n</u> tre – pot)	é (Augmm <u>e</u> nt)

-----	-----		Gazett<u>i</u>er
		Ē	ē (Gazett <u>i</u> er)

-----	Die Alo<u>e</u>,	-----	Ob<u>o</u>ë.
	ë (Alo <u>e</u>)		ë (Ob <u>o</u> ë)

-----	-----	-----	B<u>e</u>te
		Ê	ê (B <u>e</u> te)

F<u>a</u>ckeln,	stift<u>e</u>n.	F<u>a</u>ntást	Ben<u>e</u>ficiát
------------------------	------------------------	-----------------------	--------------------------

F (Fackeln)	f (stiften)	F (Fantást)	f (Beneficiāt)
-----	Waffenlo <u>s</u>	-----	Effron <u>té</u>
	[Ligatur] ff (Waffenlos)		[Ligatur] ff (Efffronté)

-----	flacher	-----	Nelkenfloren
	[Ligatur] fl (flacher)		[Ligatur] fl (Nelkenflore <u>n</u>)

-----	Sa <u>f</u> tring	-----	-----
	[Ligatur] ft (Saft <u>r</u> ing)		

<u>G</u> algen <u>s</u> äule,	Baum <u>s</u> ä <u>g</u> e	<u>G</u> ráphic	Aú <u>g</u> ur
G (Galgensäule)	g (Baumsä <u>g</u> e)	G (G <u>r</u> aphic)	g (Aú <u>g</u> ur)

<u>H</u> elden	Xan <u>t</u> hus	<u>H</u> ésperus	<u>G</u> ráphic
H (Helden)	h (Xan <u>t</u> hus)	H (H <u>e</u> sperus)	h (Grap <u>h</u> ic)

<u>I</u> nhalt	w <u>i</u> e i <u>s</u> t d <u>i</u> e <u>s</u>	<u>I</u> ncontin <u>e</u> nz	<u>G</u> ráphic
I (Inhalt)	i (w <u>i</u> e i <u>s</u> t d <u>i</u> e <u>s</u>)	I (Incompetenz)	i (Grap <u>h</u> ic)

	Ü <u>berf</u> ir <u>n</u> issen	<u>I</u> nd <u>e</u> x.	Blum <u>i</u> ren
Í	í (Ü <u>berf</u> ir <u>n</u> issen)	Í (Í <u>n</u> dex)	í (Blum <u>i</u> ren)

-----	Durch <u>b</u> it <u>t</u> ern	-----	-----
	%i (Durchb <u>%i</u> tt <u>e</u> rn)		

-----	-----		St <u>ī</u> l
		ī	ī (Stīl)

-----	-----		t <u>ī</u> tre commun
		î	î (t <u>ī</u> tre commun)

-----	Der <u>Ũ</u> i,	-----	Concho <u>ī</u> de
	ï (Aï)		ï (Conchoïde)

<u>J</u> un <u>k</u> er. <u>J</u> ahr	<u>j</u> e	-----	adj. u. adv.
J (nur +Vokal) I=≠J (Junker, Jahr)	j (je)	J	j (adj. u. adv.)

<u>K</u> loster.	<u>J</u> un <u>k</u> er.	<u>K</u> alénder.	<u>K</u> akerl <u>ä</u> k.
K (Kloster)	k (Junker)	K (Kalénder)	k (Kakerlak)

<u>L</u> acherhaft,	<u>G</u> algen <u>s</u> äule,	<u>L</u> aviren	<u>S</u> qual <u>u</u> s
L (Lacherhaft)	l (Galgensäule)	L (Laviren)	l (Squalus)

<u>M</u> undgut,	<u>Ũ</u> m <u>s</u> tapeln,	<u>M</u> ythen	Adverbium <u>m</u>
M (Mundgut,)	m (Ũmstapeln,)	M (Mythen)	m (Adverbium)

<u>N</u> achblasen,	T <u>a</u> nde <u>n</u>	<u>N</u> ónsens	In <u>co</u> ntin <u>e</u> n <u>z</u>
N (N <u>a</u> chblasen,)	n (T <u>a</u> nde <u>n</u>)	N (N <u>a</u> ónsens)	n (In <u>co</u> ntin <u>e</u> n <u>z</u>)

<u>O</u> sterapfel,	<u>v</u> on	<u>O</u> rgánisch.	<u>o</u> rganisiren
O (O <u>s</u> terapfel)	o (v <u>o</u> n)	O (O <u>s</u> rganisch)	o (o <u>r</u> ganisieren)

-----	ü <u>b</u> erd <u>o</u> rren	<u>O</u> rient	Antip <u>o</u> de
	ó (Ü <u>b</u> erd <u>o</u> rren)	Ó (Ó <u>r</u> ient)	ó (Antip <u>o</u> de)

-----	-----		B <u>o</u> ogspriet
		Ō	ō (B <u>o</u> ogspriet)

-----	-----		H <u>o</u> tél
		Ô	ô (H <u>o</u> tél)

<u>Ö</u> sterreich	<u>G</u> öttern	-----	Vapor <u>ö</u> ,
Ö (Ö <u>s</u> treich)	ö (Göt <u>t</u> tern)	Ö	ö (Vapor <u>ö</u> s)

-----	ü <u>b</u> erkr <u>ö</u> p <u>f</u> en	-----	-----
%/Ö	%/ö Ü <u>b</u> erkr <u>ö</u> p <u>f</u> en)	%/Ö	%/ö

<u>P</u> fandlich,	ü <u>m</u> stap <u>e</u> ln,	<u>P</u> ädag <u>o</u> g	H <u>e</u> sperus
P (P <u>f</u> andlich)	p (Ü <u>m</u> stap <u>e</u> ln)	P (P <u>f</u> ädag <u>o</u> g)	p (H <u>e</u> sperus)

<u>Qual</u> ,	Durch <u>qual</u> men	<u>Qual</u> ität.	<u>Squal</u> us
Q (Qual)	q	Q (Qualität)	q (Squalus)

<u>Ras</u> edichter,	Lach <u>er</u> haft,	<u>Ran</u> ziön	Marg <u>uer</u> ite).
R (Rasedichter)	r (Lacherhaft)	R (Ranzion)	r (Marguerite)

<u>Spa</u> ße,	<u>Waff</u> enlos, (Waffenlos)	<u>Squal</u> us	<u>intr</u> s. u. <u>tr</u> s. (intrs. u. trs.)
-----	<u>Ras</u> edichter, [Schaft- "s"]	-----	<u>Nón</u> sens (Nónsens)
S	s (Rasedichter)	S	<u>Bá</u> ßs (Báßs)

-----	<u>Grö</u> ße	-----	-----
	ß (Größe)		

<u>Sch</u> ellen	<u>Was</u> chhaft,	<u>Sch</u> éma	<u>Elást</u> isch
Sch (Schellen)	sch (Waschhaft)	Sch (Schéma)	sch (Elástisch)

-----	<u>sicht</u> bar	-----	-----
	[Ligatur] si (sichtbar)		

-----	Verh<u>ä</u>ltnisse	-----	-----
	[Ligatur] ss (Verh <u>ä</u> ltnisse)		

-----	Úmstap<u>e</u>ln,	-----	-----
	[Ligatur] st (Úm <u>st</u> apeln)		

<u>T</u>anden	K<u>a</u>rte	<u>T</u>échnic	Ábcap<u>i</u>tteln
T (T <u>a</u> nden)	t (K <u>a</u> rte)	T (T <u>a</u> chnic)	t (Ábcap <u>i</u> tteln)

-----	Sp<u>a</u>tze,	-----	-----
	[Ligatur] tz (Sp <u>a</u> tze)		

<u>U</u>ndienst,	Qual,	<u>U</u>nique	Adverb<u>i</u>um
U (U <u>n</u> dienst)	u (Q <u>a</u> l)	U (U <u>n</u> ique)	u (Adverb <u>i</u> um)

<u>U</u>mstapeln,	Überf<u>u</u>nkeln	<u>U</u>ranus	A<u>u</u>gur
Ú (Ú <u>m</u> stapeln)	ú (Überf <u>u</u> nkeln)	Ú (Ú <u>a</u> ranus)	ú (A <u>u</u> gur)

-----	-----	-----	Bless<u>u</u>r
		Ū	ū (Bless <u>u</u> r)

<u>U</u>berd<u>a</u>chen,	Betr<u>u</u>ger	-----	<u>I</u>üstiren
Ü (Ü <u>b</u> erd <u>a</u> chen)	ü (Betr <u>u</u> ger)	Ü	ü (I <u>u</u> stiren)

	<u>D</u> urchbegleiten	-----	-----
%)%/U (%)%/U)	%)%/ (D%)%/urchbegleiten)		

<u>Ü</u> berkriegen	<u>Ü</u> berf <u>ü</u> hlen	-----	-----
%/Ü (%/Überkriegen)	%/ü (Überk%/ühlen)		

<u>V</u> erdienen ,	<u>v</u> ereiniget	<u>V</u> alétt	adj. u. adv.
V (Verdienen,)	v (vereiniget)	V (Valétt)	v (adj. u adv.)

<u>W</u> affenloß ,	<u>w</u> eiterer	<u>W</u> oole	<u>w</u> hole
W (Waffenlos)	w (weiterer)	W (Woole)	w (whóle)

<u>Y</u> ,	<u>M</u> urray	<u>Y</u> ard	<u>M</u> ythen
Y	y	Y (Yard)	y (Mythen)

<u>X</u> anthus	<u>X</u> erifon	<u>X</u> erophagie	(Don <u>a</u> x)
X (Xanthus)	x	X (Xerophagie)	x (Donax)

<u>Z</u> abel,	<u>z</u> . B.	<u>Z</u> ébra	<u>Ä</u> cciden <u>z</u>
Z (Zabel)	z (z. B.)	Z (Zébra)	z (Äccidenz)

Interpunktion:

Chaussüre (spr. Schoßführ')	
' (Schlossuhr')	Hochkomma

Vúlgo, ingemein, gewöhnlich,	
, (Vúlgo,)	Komma

Volcān. S. Vulcan.	
. (Volcān.)	Punkt

(s. Senden)	
() (s. Senden)	runde Klammer

Daß Einschließzeichen [] oder (), welches	
[] ([] oder)	eckige Klammer

Asyndeton, in der Redekunst, eine Belebung (Figur), nach welcher mehrere Wörter und Sätze ohne Verbindungswörter auf einander folgen; z. B. Ruhe, Zufriedenheit, Vermögen, Gesundheit, Leben — Alles ist für mich dahin!	
! (dahin!)	Ausrufezeichen
Astronómisches. Ich beziehe mich auf Das, was ich oben unter (Ästhetisch) über die veralteten Beiwörter lehrig und künftig angemerkt habe. Würden diese wieder erneuert, und was könnte uns daran hindern? so könnten wir sternlehrig für astronómisches, gegenständ-	
? (?)	Fragezeichen

Der Fliegenstein, des <u>—s</u> , Mz. die <u>—e</u> .	1. bei grammatischen Angaben
Fußbreit, adj. u. adv. eines Fußes breit, die Breite eines Fußes habend, sowol eines natürlichen Fußes, als desjenigen Längenmaßes, welches Fuß heißt. Es ist noch nicht einmahl fußbreit. Auch als Grundwort für etwas, das eines Fußes breit ist. <u>—</u> und nicht ein Fußbreit Erde Gewann es ihm, das grause Menschenopfer. Schiller.	2. bei Verszitaten
Reifen, v. intrs. u. trs. bei den Schiffern, das Seil an das Segel nähen. D. Reifen. D. <u>—ung</u> .	3. bei Ableitungen
langer Strich — (Kodierung als Geviertstrich: Alt+Strg+Num-)	

Astronomie, die Sternlehre, Sternwissenschaft. Sternkunde (wo= für man auch Himmelskunde sagen kann), sollte nur von der in= wohnenden Kenntniß der Sterne gebraucht werden. Burja aber weist diese Benennung demjenigen Theile der Sternlehre an, welcher Lat. astronomia sphärica, und den Namen Sternwissenschaft dem=	wo = für in= wohnenden dem= ...
Athenienser, besser die Athener, so wie Athenisch für Athenien= sisch. G. Italiener. Einer meiner freundschaftlichen Beurtheiler eifert sehr über die Verkürzungen, wie er es nennt, der alten Böl= ker= und Städtenamen, und verlangt, daß wir nicht Athen, Theben Rom, sondern Athenā, Thebā, Roma, folglich auch Athenāer,	- (bei Antiquaschrift) = (bei Fraktur) (+Bindestrich im Wort)
Silbentrennstrich (Unterschied Antiqua - Fraktur) einheitliche Kodierung als Trennstrich -	

○ Der Farbensduft, des —es, <u>Mz.</u> die —düfte, gleichsam ein Dufst von Farben, sanfte, feine Farben, von welchen das Bild wie von einem Dufte überzogen ist. »— an allen Gegenständen vertrocknete der feuchte, weiche Farbensduft, der bisher herbftlich über ihnen ra- gend gehangen.« <u>J. P. Richter.</u>	Anführungszeichen Anfang und Ende eines Zitats
» « (ist. »— an) (gehangen.« J.)	
Fortschwimmen, v. unregelm. (f. Schwimmen). <u>I)</u> ntr. mit fein, vom Wasser fortgetragen werden. Der Kahn, das Holz ist fortgeschwommen. <u>II)</u> intrs. mit fein. 1) Sich durch Schwimmen entfernen. <u>2)</u> Fortfahren zu schwimmen. — Das Fortschwimmen.	Nummerierung erste Ebene mit römischen, weitere mit arabischen Ziffern
I) II) 1) 2) (. I) nrt) (. II) intrs.) (sein. 1) Sich) (. 2) Fortfahren)	
Atheist, Gottläugner würde deutlicher als Gottesläugner sein, aber das letzte ist gebräuchlich, das erste nicht. Der verkappte Realis de Vienna (Gabriel Wagner, S. Herder's Briefe über die Humanität Thl. 4. S. 17 folg.) hat Ungötter (wie Abgötter) dafür: »Ein Unkrist ist kein Ungötter.« Allein in beiden Wörtern müßte es Götterer, statt Götter heißen; um sie nicht mit der Mehrzahl von Abgott und Ungott zu verwechseln. Da es ehemahls ein Aus:	Doppelpunkt und Semikolon
: ; (dafür: »Ein) (heißen; um)	

Die Buchstaben des griechischen Alphabets:

ά έ ή ί ύ α β γ δ ε ζ η ι κ λ μ ν ξ ο π ρ ς σ τ υ φ χ ψ ω ι̇ υ̇ ό ύ ώ

Α Ε Η Ι Ο Υ Ω ι̇ Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω Ι̇ Υ̇

Ziffern:

0	0	0	0	1	1	1	1	3	3	3	3	8	8	8	8
0	0	0	0	1	1	1	1	3	3	3	3	8	8	8	8
0	0	0	0	1	1	1	1	3	3	3	3	8	8	8	8
0	0	0	0	1	1	1	1	3	3	3	3	8	8	8	8
0	0	0	0	1	1	1	1	3	3	3	3	8	8	8	8
1	1	1	1	1	1	1	1	4	4	4	4	8	8	8	8
1	1	1	1	1	1	1	1	4	4	4	4	9	9	9	9
1	1	1	1	1	1	1	1	4	4	4	4	9	9	9	9
1	1	1	1	1	1	1	1	5		5	5	9	9	9	9
1	1	1	1	1	1	1	1	6	6	6	6	9	9	9	9
1	1	1	1	1	1	1	1	6	6	6	6	502			502
1	1	1	1	1	1	1	1	7	7	7	7				

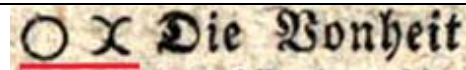
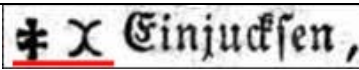
NB! Die Ziffern sind Minuskel- bzw. Medieval-Ziffern und halten nicht die Linie!- vgl. 502)

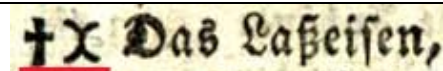
Vorsicht bei den Unterschieden zwischen Buchstabe „O“ und Ziffer „0“, zwischen Buchstabe „l“ und Ziffer „1“.

Sonderzeichen für Verwendungsangaben (nur in Campe: Deutsches Wörterbuch) mit ihren Unicodewerten (hex.):

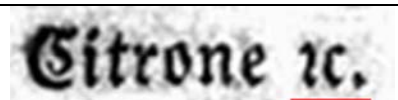
 Der Felsenfall,	Kreis	 Der Dichterquell,	Kreis im Kreis
○	○	☉	◉ 9737 SUN, miscellaneous symbols
 Daß Felsenherz	leeres Dreieck	 Die Felsenbrust,	Dreieck mit Innenkreis
△	△ 9651 WHITE UP-POINTING TRIANGLE, geometric shapes	◬	◤ 9708 WHITE UP-POINTING TRIANGLE WITH DOT, geometric shapes
 Die Überart	Kreuz	 Der Walz	Doppelkreuz
Ϯ	Ⲉ 1006 COPTIC CAPITAL LETTER DEI, greek and coptic	{DK}	
 Evangelisch	Doppelhalbkreis, seitlich offen	 Der Zeterjunge	Doppelhalbkreis, durchgestrichen
{DHK}		{DHKG}	
 Wältigen	Stern	 Wältig	zwei Sterne
*	*	**	**
 Überallmächtig	umgekehrtes Venuszeichen		
♁	♁ 9793 EARTH		

Auch Kombinationen aus zwei Zeichen sind möglich (Auswahl):

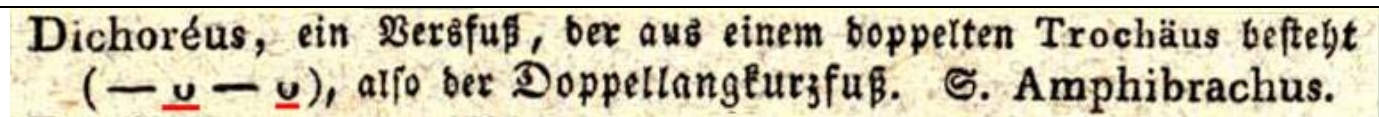
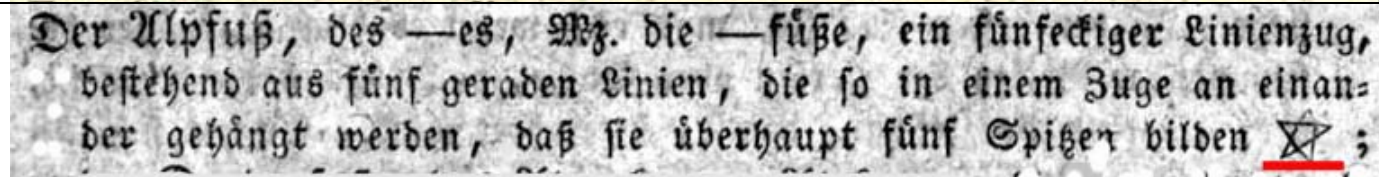
		
○{DHK}		{DK}{DHK}

		
Ϯ{DHK}		*△

Sonstiges:

	etc.
etc.	

Bei Zeichenformen, die nicht aufgeführt sind, ein Schmierzeichen  einfügen

	(—  — ),
	bilden  ;

Zeilenfall wie im Original (Zeilenumbruch durch Enter-Taste)

Silbentrennung am Zeilenende beibehalten und wiedergeben wie beschrieben als Trennstrich

Leerräume am Anfang, innerhalb oder am Ende der Zeile = 3 Blanks

Schriftauszeichnungen (schachtelbar)

Fettdruck	Anfang:	#f+	Ende:	#f-
Kursiv		#/+		#/-
Sperrung		#s+		#s-
Antiqua		#a+		#a-

Schriftgrößenunterschiede

Grundschrift wird nicht markiert

Größerer Schriftfont

Anfang: #l+ Ende: #l- (el)

Hebräisch

Anfang: #hr+ Ende: #hr- (dazwischen Schmierzeichen)

Seitenzahl zuerst erfassen, anschließend den linken Kolumnentitel und den Text der linken Spalte, gefolgt vom rechten Kolumnentitel und dem Text der rechten Spalte.

Kodierung für die Seitenzahl:

Anfang: #n+ Ende: #n-

Kolumnentitel (rechts und links):

#@+ #@-

Reklamanten und Kustoden am Seitenende entfallen.

Musterseiten:

S. 17 Band 1: Seite 74

S. 18 Band 1: Seite 675

S. 19 Band 6: Seite 216

S. 20 Band 6: Seite 609

Reich abziehen, ihn abseiten. Wein und Bier abziehen, d. h. in ein andres Gefäß oder in Gläsern füllen. Brantwein, Weingeist abziehen, ihn mit den flüchtigen und kräftigen Theilen anderer Dinge, als Kräuter u. durch die Wärme verbinden, und ihm ihre Kraft mittheilen. Blumen, Kräuter u. abziehen, ihre flüssigen und flüchtigen Theile, durch eine darauf gegossene Flüssigkeit in der Wärme ausziehen lassen. Abgezogene Wasser. Daher auch unzeitiglich, das Wasser aus einer Gasse absondern. — weil die Sprache aus allen Mundarten abgezogen (desillirt) und emporgetrieben (sublimirt) wird. — W ü d i g e r. — Im Weinbaue u. ablegen, absinken, auf die Art, daß die Enden der Reinfäden in die Erde abgebogen werden bis sie anwurzen. — In Salzwerken, die Pfanne abziehen, sie vom Gerste losmachen und ausheben. Bei den Karschnern, ein Fell abziehen, die Fleischseite desselben völlig rein machen. Bei den Fiskern, die Arbeit abziehen, sie mit der Fischseite glatt und glänzend machen. — Bei den Färbern, die Farbe abziehen, sie aus einem gefärbten Zeug ausziehen; auch abblöden. — Bei Buch- und Kupferdruckern, einen Bogen abziehen, ihn abdrucken, doch nur von einzelnen, z. B. von Durchschlagen. Eine Kupferplatte abziehen, ein Probeklatt von derselben abdrucken; auch überhaupt Abdrücke davon nehmen. Er hat von dieser Platte nur 200 Abdrücke oder Blätter abziehen lassen. — Bei den Metallarbeitern heißt diejenigen Stücke, welche zusammengelegt werden sollen, abziehen, so viel als sie ganz gerade seien. Auf ähnliche Art bedeutet Gewicht abziehen, sie eichen, wobei oft mit der Feile noch etwas abgenommen werden muß. — Bei den Waffenschneidern, abmessen, einen Schwant an den Tag bringen, daß man am Tage weiß, wo man in der Grube bauet. — Bei Schreinerwerkern, Messerschneidern u. ein Messer und dergl. abziehen, dasselbe, nachdem es geschliffen worden, zur Wegschaffung des Grathes auf dem Abziehsteine oder auf einem lederen Riemen streichen, und dadurch glatt und scharf machen. — Bei den Schuftern, das Leder abziehen, es mit Winstein abreiben. — In den Küchen, die Suppe mit einem Ei abziehen, ein Ei hineinsetzen, und ihre Verteilung dadurch vollenden. 3) In einigen unzeitlichen Bedeutungen. (1) Dem Gemüthe nach von etwas entfernen. Einen von seinem Vorleser, von einem Plane, einer Partei, vom Wege der Jugend u. abziehen. Sein Herz von etwas abziehen. So auch, sich von der Welt abziehen. (2) Der Zahl, dem Maße, dem Gewicht nach vermindern. In der Rechnungskunst besonders heißt abziehen (subtrahiren), von einer größeren Zahl eine kleinere wegnehmen, um zu sehen wie viel von jener übrigbleibt. Dann, Etwas etwas abziehen, von dem, was er bekommen hat, etwas zurückbehalten. Einem etwas von seinem Lohne, vom geforderten Preise, von dem ihm bestimmten Theile des Geldes u. abziehen. So auch, sich etwas abziehen. (3) In Gedanken ganz absondern, ganz allein denken, die allgemeinen Merkmale und Bestimmungen der Dinge für sich betrachten (abstrahiren). Daraus entsteht ein abgezogener Begriff. Abgezogen betrachtet erscheint die Sache so. Abziehen ist in dieser Bedeutung keinesweges verwerflich oder dem Absondern nachzusehen. S. G. W e r d e n u n g s w ö r t e r b. A b s t r a c t. II) rec. Sich abziehen. 1) Sich durch vieles Ziehen entkräften, vom Zugvieh. 2) Von den frisch gedruckten Bogen, die sich abziehen, wenn sie beim Schlagen und Pressen die Schwärze fahren lassen. 3) Sich zurückziehen. Gegen Abend ward man gewahr, daß das bisher bei Dohersicht gefundene Korps (der Heerhaufen) sich abgezogen hatte. II) g e n. III) intrs. mit sich, sich mit seinem Gewichte u. von einem Orte entfernen. Der Feind zog wieder ab. Die Wache zieht ab. Unzeitlich. Mit Schimpf und Schande, mit einer langen Nase abziehen. Du kannst nur immer abziehen, verächtlich für gehen. Vom Gefinde gebraucht, bedeutet es aus dem Dienste ziehen. In Weinachten zieht das Gefinde gewöhnlich ab. Schematis bedeutet abziehen überhaupt, sich mit seiner Habe aus einer Stadt oder Gegend wegzugehen, wofür man aber gewöhnlicher

wegziehen sagt. Unzeitlich vom Stande u. für, sich entfernen. Der Rauch zieht durch die Röhre ab. — Das Abziehen. Die Abziehung. Der Abzieher, des —s, d. Wz. w. d. Gz. der abziehet oder der etwas abziehet; ungewöhnlich. In der Bergbauerkunst, ein Muskel, der ein Glied vom andern abziehet, d. h. entfernt. Der lange Abzieher des Daumens (Musculus abductor pollicis longus) zieht den Daumen vom Zeigefinger ab. Der kurze Abzieher des Daumens (Musc. abd. pollicis brevis). Der Abzieher des kleinen Fingers (Musc. abd. digiti minimi). Wie man n. Der Abzieher der großen Zehe (Musculus abductor hallucis) zieht die große Zehe von den übrigen ab. Der Abzieher der kleinen Zehe (Musculus abductor digiti minimi pedis). Der s.

Die Abziehseile, Wz. die —n, eine breite und dicke flächerte Seile, welche sehr fein gehauen ist, mit welcher die Metallarbeiter die zusammengelegten Stücke gerade und glatt stellen.

Der Abziehmuskel, des —s, Wz. die —n, in der Bergbauerkunst, eine Senkung derjenigen Muskeln, durch deren Hilfe die Glieder, an welchen sie sich befinden, auswärts oder seitwärts bewegt, gleichsam abgezogen werden.

Der Abziehstein, des —es, Wz. —e, ein feiner Schieferstein, auf welchem Schneidwerkzeuge mit Baumöl abgezogen werden.

© Das Abziehungsvermögen, des —s, o. Wz. das Vermögen der menschlichen Seele, sich allgemeine Begriffe von den Dingen abzuziehen. Auch das Abziehungsvermögen. S. G. W e r d e n u n g s w ö r t e r b. A b s t r a c t i o n s v e r m ö g e n.

○ Die Abziehhülse, Wz. die —en, in der Rechnungskunst, diejenige Zahl, die von einer größeren abgezogen werden soll (Subtrahens).

Das Abziehzeug, des —es, Wz. die —e, alles dasjenige Gerath, welches zum Abziehen des Brantweins gebraucht wird.

Abziehen, v. I) intr. mit haben, auf einen Zweck hingearbeitet sein. Das zieht dahin ab. Es zieht auf nichts Gutes ab. II) intrs. auf etwas zielen, sein Augenmerk richten (visiren). Auf einen festen Punkt abziehen. III) trs. etwas zum Ziel nehmen, zu erreichen suchen. Die abgezielte Wirkung thun. A t t i s i s c h e s W e s e n. „Zum mindesten wird bei dieser Art von Rede nichts Körperliches abgezielt.“ Wieland. „Ob Luna selbst dabei etwas abzielte, entschwebet die Geschichte nicht.“ Der s. — Das Abziehen. Die Abziehung. Abzimmern, v. trs. 1) Gehrig, fertig zimmern. Ich habe schon die Schwellen abzimmern lassen. 2) Durch zimmern, mit der Stimmgabel absondern. Epäne abzimmern. Ein Stück vom Balken abzimmern. Das Abzimmern. Die Abzimmern.

Abzirkeln, v. trs. mit dem Zirkel abmessen. Eine Zeichnung abzirkeln. Dann unzeitiglich, genau, ängstlich abmessen, bestimmen. Wer kann alles abzirkeln. Er will alles abgezirkelt haben. Das Abzirkeln. Die Abzirkelung.

Abzittern, v. I) intr. 1) Mit sein, zitternd abfallen. Das Laub zittert von dem Baume ab. Die Hitznadel wird abzittern. 2) Mit haben, sehr zittern. Man ließ ihn drängen in der Kälte abzittern. II) intrs. mit sein, sich zitternd entfernen. — Das Abzittern.

X Abziehen, v. trs. 1) Durch Ziehen, Saugen, das Saftige oder Fleischichte an einer Sache absondern. 2) Durch vieles Ziehen entkräften. Das Kind zieht seine Mutter ganz ab. Das Abziehen. Die Abziehung. Ist abgezottelt. II) X intrs. zottelnd sich entfernen. S. Zotteln. Er abzozottelt. II) X trs. so viel als abkaufen. Der Waff ist sehr abzozottelt. — Das Abzotten. Die Abzottelung.

Die Abzucht, Wz. die —ächte (—en). 1) Eine Zucht von Thieren, die von einer gewissen Gattung abstammen. Die Pferde in Schiti sind eine Abzucht der Spanische. 2) + Ein Ort, durch welchen oder in welchem eine Flüssigkeit abgezogen oder abgeleitet wird. In den Gärten werden unter den Linden besonders die Gräben, welche kreuzweise unter den Linden fortgeführt sind. Abzucht ist nichts anderes als Abzug, nur in anderer Form und daß es vorzüglich von neuen Gräben gebraucht wird.

Conglomeriren, zusammenhäufen, ballen. Auch zusammenfammengesetzt hatte.“ Sagen der Vorzeit. „Ich will euch schwören lassen, über der grauen Tiefe meines Stendes an einem Strohhalme, bis euer Blut, wie Hagel, sich zusammenkugelt.“ *Eubolus*.

Conglutinat, etwas Zusammengeleimtes.

Conglutination, das Zusammenleimen, Zusammenkleben.

Conglutiniren, zusammenleimen.

Congregation. 1) Die Versammlung; 2) eine Ordensverbindung oder Würdigung; 3) eine Rathversammlung der Cardinäle, also die Hauptpriesterversammlung. *B.* Die *Congregatio de propaganda fide* zu Rom ist eine von den Päpsten angeordnete Stelle (*Collegium*), welche die Ausbreitung des Röm. Kathol. Glaubens zum Zweck hat, also eine Glaubensveranstaltung, oder ein geistliches Werckhaus; erstlich gesprochen ein Ausbreitungserath oder eine Ausbreitungsanstalt. (3 u. f.) *B.* hat Vesehrungsath und Vesehrungsanstalt dafür vorgeschlagen.

Congress. 1) Der Zusammentritt oder die Zusammenkunft abgeordneter Staatsbevollmächtigter; 2) Die Gesellschaft solcher Bevollmächtigten. Ein Friedenscongress ist eine Friedensversammlung. Den Congress der Amerikanischen Staaten kann man durch Staatenversammlung verstehen. (3 u. f.) „Staatsversammlung, welches ehemals dafür angelegt war, schien mir minder paßlich, weil bei dem Amerikanischen Congress, wie bei den Versammlungen in verschiedenen Staaten, aus welchen jener Kreislauf besteht, also in den ehenmaligen vereinigten Niederlanden, sich die Abgeordneten der verschiedenen Staaten selbst, versammeln; bei einer Staatenversammlung aber bloß die ersten Diener eines und desselben Staats unter Vorbehalt des Staatsoberhauptes zusammenkommen.“ *B.* Ich finde diese Bemerkung vollkommen gegründet, und habe diesem zufolge den Ausdruck Staatsversammlung, als nicht hiehergehörig, weggelassen.

Congruent, flüßig, übereinstimmend oder übereinstimmig, zusammenfließend oder zusammenflüßig. In der Mathematik werden zwei Figuren, die nicht bloß einander ähnlich (*similes*), sondern auch gleich (*aequales*) sind, d. i. einerlei Flächeninhalt und einerlei Umfang haben, folglich sich einander vollkommen decken, congruent, auf Deutsch ähnlich gleich genannt. Dieses Deutsche Wort hat *Wurja* dafür angelegt. (3 u. f.) Besser gleichähnlich, welches v. *Wintersfeld* dafür gebraucht hat.

Congruenz, das Übereinkommen, die Fügigkeit, das Zusammenpassen. (3 u. f.) Auch die Gleichähnlichkeit, in maßähnlicher Bedeutung. Dieses Wort ist schon von *Wintersfeld* gebraucht worden. Ich habe auch Einerleiheit dafür versucht: „Die vollkommene Einerleiheit beider Figuren, vermöge welcher sie sich einander decken.“

Congruiren, übereinstimmen, gleich sein, zusammenpassen; in der Erdkunde, gleichähnlich sein.

Cónisch, kegelförmig, kegelticht. (3 u. f.) *Wos* hat in Regelgestalt dafür gesagt:

„Doch in Regelgestalt erhebt sich der schlanke Cupressus.“

Conjectur, eine Vermuthung, Muthmaßung; eine muthmaßliche That.

Conjecturiren, vermuthen, muthmaßen.

Conjugation, in der Sprachlehre, die Abwandlung, *Πορωτική*. *Eschottel*, welcher unter Abwandlung die Declination versteht, nennt die Conjugation Zeitwandlung. *Клопшод* hat in den Grammaticischen Gesprächen das Wort Umbildung dafür, wie es scheint, mit Recht beliebt, so wie Umwandlung für Declination. Den guten Grund, den er dabei vor Augen gehabt zu haben scheint, siehe unter Declination.

Conjugiren, abwandeln. *Πορωτική*. *Клопшод* besser, *амбилден*. *S.* Declination.

Conjugal, ehelich.

Cóinspector, der Mitaufseher.

Cóinspectorat, das Mitaufseheramt.

Conjunction. 1) überhaupt die Vereinigung, Verbindung; 2) in der Sprachlehre, das Bindewort. Andere, wie *Eschottel*, in seiner Lat. Sprachlehre, haben Bindewort, und *Eschottel*, Bindewort dafür gewählt. Auch *W.* hat sich für diesen letzten Ausdruck erklärt. „Dieser Redetheil sagt er (in dem *Lehrgebäude*); hat im Lateinischen den Namen von seinem Amte, von dem Zusammenfügen, weil er Sätze und ihre Glieder zusammenfüget. Nimmt man diesen Ausdruck in seiner weitesten Bedeutung, in welcher auch die Trennung, Auslöschung u. s. f. eine Art der Zusammenfügung ist, indem dadurch die Dinge wenigstens in den Sätzen neben einander gestellt werden: so ist er untadelt. Nicht so der Deutsche Ausdruck Bindewort, welcher die bloß verbindende Bedeutung zu bestimmt bezeichnet, und daher nur auf einen sehr kleinen Theil derselben paßt. Bindewort (warum nicht Bindewort?) ist etwähliger.“ Aber was hindert uns, dem Ausdrücke Bindewort eben die Enfscheidung zu Etatten kommen zu lassen, welche *W.* dem gleichbedeutenden Bindewort angedeihen läßt, und zu sagen, es werde damit nur auf die Verbindung der Sätze und ihrer Glieder gesehen, ohne alle Rücksicht, ob die darin enthaltenen Begriffe verbunden oder getrennt werden sollen? Daß jenes die verbindende Bedeutung bestimmter bezeichnet, als dieses, kann ich nicht finden.

Conjunctio causalis, das begründende Füge = oder Bindewort.

Conjunctio conditionalis, das bedingende Füge = oder Bindewort. **Conjunctis viribus**, mit vereinigten oder vereinten Kräften; mit Kraftverein. *Leges* hat *W.*

Conjunctiv, in der Sprachlehre. Einige haben die verbindende Weise, Andere die ungewisse Weise oder Form, und noch Andere, wie *W.* *na*, etwas dunkel, die Bind-art dafür einführen wollen. (3 u. f.) Nach oft wiederholtem Nachdenken über die Verdeutschung der Lateinischen sprachlichen Kunstwörter bin ich für Indicativus und Conjunctivus bei den Ausdrücken ungebundene und gebundene Weise stehen geblieben, weil ein im Coniunctivo stehendes Ausageswort, um die Aussage verständlich und vollständig anzugeben, außer dem Grundworte (Substantiv, auch noch einer Verbindung mit andern Wörtern bedarf, oder an diese gebunden ist. *S.* meinen Versuch einer genauern Bestimmung und Verdeutschung der sprachlichen Kunstwörter, Straußsch. 1804.

Conjunctur. Dieses Wort können wir durch das alte Zeitsäule, oder durch Zeitsumstände, und durch Umstände (schlechtweg übersetzen. Unter den jetzigen Umständen; bei den jetzigen Zeitsäulen oder Zeitsumständen. Man muß die glücklichen Zeitsumstände oder das Zusammenreffen der glücklichen Zeitsumstände begreifen.

Conjungere, vereinigen, auch einigen und verbinden. **Conjurant**, *Frang.* **Conjuré** (*Fr.* Konguré), der Mitverschworne **Conjuratio**, die Verschwörung.

Conjuriren, sich verschwören, eine Verschwörung eingehen.

Connectiren, verbinden, verknüpfen. Man hört es auch, wiewol unrichtig, für zusammenhängen. Diese beiden Sätze connectiren nicht, d. i. sie hängen nicht zusammen.

Connetabel, *Frang.* **Connetable** (*Fr.* Konnetab^l), der Kronseibherr.

Connéx, verbunden, verknüpft. **Connéxa**, verbundene Dinge.

Connexión oder **Connection**, die Verbindung oder Verknüpfung; oft auch bloß die Bekanntheit.

Connivenz. *Grisch* hat das veraltete Unachtung dafür, welches wieder in Umlauf gebracht zu werden verdient. *Conj* haben wir auch Nachsicht. (3 u. f.) Es scheint, daß Nichtachtung ihnen keinen Vortheil vorzuziehen sei.

Conniviren, nachsehen; im gemeinen Leben, Ein Auge zudrücken, auch durch die Finger sehen; ein maßetischer Ausdruck! Der Con-

Vogue, en (spr. ang Woge) sein', im Schwange, am Breite sein, viel gelten, gesucht sein.

Volatil (spr. Wolong), der Federball

Volatiliseren, in der Schwebekunst, flüchtig machen, versflüchtigen.

Man könnte auch flüchtigen schließlich dafür sagen. Im D. D. sagt man schon sich flüchtigen für sich auf die Flucht begeben oder flüchten. Versflüchtigen kann man nur von der Handlung sagen, wodurch man etwas flüchtig Gemachtes versliegen läßt.

Volatilität, die Flüchtigkeit, die Eigenheit, da etwas versliegt.

Vôle. S. Woole.

Volcän. S. Vulcan.

Volontair (spr. Wolongtär), ein Freiwilliger.

Volontairement (spr. wolongtärémang), freiwillig, aus freiem Willen, absichtlich, vorsätzlich. Ich habe auch willentlich dafür gefunden: das, was auch in meinen Grundrissen Vertiges sein mag, hat auf die Absichten des Wägers willentlich nie Einfluß gehabt.

Vot.

Völte. 1) Auf der Reitschule, der Kreislaufer oder Kreisläufer. Man hat auch der Ringeltritt dafür gesagt. 2) In der Sprache der Spieler, der Kartenschlag oder das Kartenschlagen, d. i. die Kartenverteilung.

Völt, in der Tonkunst, umgewandelt! **Eschenburg**. (Zus.) Die Wöltung V. S. heißt, volt subito, schnell umgewandelt!

Völtigren, das Kunstspringen. Eft ist auch springen schließlich hinreichend, wenn nämlich der Zusammenhang die nähere Bestimmung gibt, d. B. er übt sich im Springen, Reiten, Fischen und Springen. (Zus.) Also ein Voltigeur (spr. Völtigöhr), ein Kunstspringer.

Völtigrekunst, die Springekunst. Auch die Voltigeurs beim Fecht.

Fechte kann man die Springer nennen.

Völbilität, die Gelassigkeit, d. B. der Zunge.

Völvolumen. 1) Die Dicke, die Masse, der Umfang. 2) Ein Band,

Pack, Rolle oder ein Bündel Schriften.

Völvoluminös, vieltheilig und dick, innes, wenn von einem Werke die Rede ist, welches aus vielen Theilen oder Bänden besteht, in welchem Falle man auch vielhändig dafür sagen könnte; dieses, wenn bloß ein Werk gemeint wird, welches viel Volumen hat, d'un grand volume. In Scherz bezeichnen wir dieses letzte auch durch dicksteig.

Völvolutios, volküstig; die Dichter sagen auch wölvolutiosend dafür.

Völvölute. 1) In der Naturbeschreibung, die Wölvölute. 2) In der Baukunst, die Schnecke, ein Schneckel an dem Jönischen Säulenknopf.

Völvören, brechen, sich erbrecnen oder übergeben.

Völvöriv, ein Brechmittel.

Völvördoehren. Ich habe, in den Beiträgen zur weitem Ansbild. der Deutsch. Sprache, vorschulmeistern dafür zu sagen versucht. (Zus.) Wir haben auch das gute Wort schulen für in die Schule nehmen, die Schule machen lassen, wovon man vorschulen für Einem die Schule vormachen, d. i. ihm zeigen, wie er etwas schulmäßig machen, thun oder sagen müsse (also ihm vordociren) bilden könnte. Das einfache schulen hat unter Andern J. P. Richter gebraucht: „Ich schule sowohl mein Sonnenroß, als mein Bagageros (meinen Packgaut) viel anders.“ Vorschulmeistern ist von Seume nachgesagt worden: „Da ich eben nicht viel zu thun habe, so will ich die die Stelle ein wenig vorschulmeistern.“

Völvöta, die Stimmen. Vota majora, die Stimmenmehrheit.

Völvören, stimmen.

Völvörung, die Stimmengabe oder Stimmensammlung.

Völvörungsnähe oder ein Ex voto, in der Römischen Kirche, ein Gemälde, zu dessen Schenkung sich Jemand durch ein Gelübde verbindlich gemacht hat; ein Gelübdebild. Wir haben auch in unserer alten Sprache das Aussagewort gelüben, welches Wächter, in den Sagen der Vorzeit, erneuert hat: „Der Mann war's, welcher als

Knabe mit ihm den Bund der Jugend gelübet hatte.“ Wie können also auch ein gelübetes Bild oder Gemälde sagen. An wölvörungsnähe und doch zugleich auch hinlänglich bezeichnend dürfte Wölvörungsbild sein. W. hat auch Wölvör- oder Gelübdebild dafür vorgezogen. Ein Ex votomähler, ist ein Gelübde oder Wölvörungsmähler.

Völvörungsschenk, ein Wölvörungsschenk. Nehmt mit diesen Wölvörungsschenken, diesen Wölvörungsgaben, auch mein Leben hin.“ Herder.

Völvörtafel, in der Römischen Kirche, eine unter oder neben einem Heiligenbilde in der Kirche aufgestellte Tafel, worauf irgend Etwas unter der Bedingung gelübet wird, daß ein gewisser Wunsch in Erfüllung gehe. Eine Wölvörtafel. Herder hat sie Wölvörtafel genannt.

Völvörum, die Stimme. Votum consultativum, die rathende oder Berathschlagungsstimme. Votum decisivum, die entscheidende oder Entscheidungstimme. Plurima vota, die meisten Stimmen, die Stimmenmehrheit.

Völvörumänna, die Menschenstimme, in der Drey.

Völvörumän. 1) Ein feuerpeitender Berg, ein Feuerschlund. (Zus.)

Auch ein Feuerberg.

Ein Chaos (Kumpen) ausgebanter Schläcken, In die ein Feuerberg zuletzt zusammenfällt. Wieland.

2) In der Pöbellehre, der Feuergeist.

Völvörnäst, entgegengesetzt dem Neptunist; zwei Parteien in der Erdbildungslehre, wovon die eine die meisten Veränderungen des Erdballs und seiner Bestandtheile durch Feuer, die andere durch Wasser entstehen läßt. Man könnte jene Feuerfreunde oder Feuerwerker, diese Wasserfreunde oder Wasserwerker nennen; in Scherz Feuermänner und Wassermänner. S. auch Neptunist. **Völvörnä**, die in der Römischen Kirche angenommene lat. Übersetzung des Bibel, die Gemeinbibel.

Völvör (spr. wölvör), gemein.

Völvö, ingemein, gewöhnlich, gemeinhin, dem gemeinen Sprachgebrauche nach.

W.

Wahlcapitulation, der Wahlhebing oder die Wahlbedingungen.

Wählconvent, die Wahlversammlung.

Walpurgis, der Name eines sogenannten Heiligen in der Römischen Kirche, dem der erste Mai gewidmet ist. Wenn wir Walpurgstag, Walpurgsnacht und Walpurgsfest sagen, so sind diese Ausdrücke Deutsch.

Wardlein, wäre Deutsch, wenn die undeutsche Betonung ihm nicht etwas Fremdes gäbe. Man sage der Wähler oder Wählwähler, und es ist ganz Deutsch. Etlicher hat Wählprüfer.

Wardiren, schägen, den Werth bestimmen. (Zus.) Ein ursprünglich Deutsches Wort, nur durch die undeutsche Endung iren verunstaltet, es heißt, wöhren. Das Grundwort Wöhren ist im Mänschen noch ganz gebräuchlich. Im R. D. haben wir auch ein Auslagewort wärdiren dafür.

Wardiren, Watten unterlegen; mit Einem Worte, watten.

Wägewood (spr. Wölvöwöde), eine schwärzliche, auch blaue und weiße porzellanartige Masse, aus welcher der Gefäße gleiches Namens Köpfe, Prachtskränze und Medaillen aller Art verfertigt läßt. In weiterer Bedeutung bezieht die Benennung Wägewood auch das sogenannte Englische Steingut unter sich. S. Fayance.

Man könnte in allgemeinen Wölvöwödegut dafür sagen. **Wölvö-atome**, welches Herder hat, hätte wol sätlich mit Wölvöstoffe vertauscht werden können.

West. S. Nord.

```
<?xml version="1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>My TEI Extension</title>
        <author>generated by Roma 3.4</author>
      </titleStmt>
      <publicationStmt>
        <p>for use by whoever wants it</p>
      </publicationStmt>
      <sourceDesc>
        <p>created on Thursday 09th October 2008 02:43:21 PM</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <front>
      <divGen type="toc"/>
    </front>
    <body>
      <p>My TEI Customization starts with modules tei, core, textstructure and header</p>
      <schemaSpec ident="myTEI" docLang="en" start="TEI" xml:lang="en">
        <moduleRef key="core"/>
        <moduleRef key="tei"/>
        <moduleRef key="header"/>
        <moduleRef key="textstructure"/>
        <elementSpec module="core" ident="add" mode="delete"/>
        <elementSpec module="core" ident="addrLine" mode="delete"/>
        <elementSpec module="core" ident="address" mode="delete"/>
        <elementSpec module="core" ident="analytic" mode="delete"/>
        <elementSpec module="core" ident="binaryObject" mode="delete"/>
        <elementSpec module="core" ident="choice" mode="delete"/>
        <elementSpec module="core" ident="corr" mode="delete"/>
        <elementSpec module="core" ident="del" mode="delete"/>
        <elementSpec module="core" ident="desc" mode="delete"/>
        <elementSpec module="core" ident="distinct" mode="delete"/>
        <elementSpec module="core" ident="divGen" mode="delete"/>
        <elementSpec module="core" ident="editor" mode="delete"/>
        <elementSpec module="core" ident="email" mode="delete"/>
        <elementSpec module="core" ident="emph" mode="delete"/>
        <elementSpec module="core" ident="expan" mode="delete"/>
        <elementSpec module="core" ident="gap" mode="delete"/>
        <elementSpec module="core" ident="gloss" mode="delete"/>
      </schemaSpec>
    </body>
  </text>
</TEI>
```

```
<elementSpec module="core" ident="headItem" mode="delete"/>
<elementSpec module="core" ident="headLabel" mode="delete"/>
<elementSpec module="core" ident="index" mode="delete"/>
<elementSpec module="core" ident="item" mode="delete"/>
<elementSpec module="core" ident="l" mode="delete"/>
<elementSpec module="core" ident="label" mode="delete"/>
<elementSpec module="core" ident="lg" mode="delete"/>
<elementSpec module="core" ident="list" mode="delete"/>
<elementSpec module="core" ident="listBibl" mode="delete"/>
<elementSpec module="core" ident="measure" mode="delete"/>
<elementSpec module="core" ident="measureGrp" mode="delete"/>
<elementSpec module="core" ident="meeting" mode="delete"/>
<elementSpec module="core" ident="mentioned" mode="delete"/>
<elementSpec module="core" ident="milestone" mode="delete"/>
<elementSpec module="core" ident="num" mode="delete"/>
<elementSpec module="core" ident="postBox" mode="delete"/>
<elementSpec module="core" ident="postCode" mode="delete"/>
<elementSpec module="core" ident="ptr" mode="delete"/>
<elementSpec module="core" ident="reg" mode="delete"/>
<elementSpec module="core" ident="relatedItem" mode="delete"/>
<elementSpec module="core" ident="resp" mode="delete"/>
<elementSpec module="core" ident="respStmt" mode="delete"/>
<elementSpec module="core" ident="rs" mode="delete"/>
<elementSpec module="core" ident="said" mode="delete"/>
<elementSpec module="core" ident="series" mode="delete"/>
<elementSpec module="core" ident="sic" mode="delete"/>
<elementSpec module="core" ident="soCalled" mode="delete"/>
<elementSpec module="core" ident="sp" mode="delete"/>
<elementSpec module="core" ident="speaker" mode="delete"/>
<elementSpec module="core" ident="stage" mode="delete"/>
<elementSpec module="core" ident="street" mode="delete"/>
<elementSpec module="core" ident="teiCorpus" mode="delete"/>
<elementSpec module="core" ident="time" mode="delete"/>
<elementSpec module="core" ident="unclear" mode="delete"/>
<elementSpec module="header" ident="appInfo" mode="delete"/>
<elementSpec module="header" ident="application" mode="delete"/>
<elementSpec module="header" ident="biblFull" mode="delete"/>
<elementSpec module="header" ident="cRefPattern" mode="delete"/>
<elementSpec module="header" ident="catDesc" mode="delete"/>
<elementSpec module="header" ident="catRef" mode="delete"/>
<elementSpec module="header" ident="category" mode="delete"/>
<elementSpec module="header" ident="classCode" mode="delete"/>
<elementSpec module="header" ident="classDecl" mode="delete"/>
<elementSpec module="header" ident="creation" mode="delete"/>
<elementSpec module="header" ident="distributor" mode="delete"/>
```

```
<elementSpec module="header" ident="edition" mode="delete"/>
<elementSpec module="header" ident="editionStmt" mode="delete"/>
<elementSpec module="header" ident="extent" mode="delete"/>
<elementSpec module="header" ident="funder" mode="delete"/>
<elementSpec module="header" ident="geoDecl" mode="delete"/>
<elementSpec module="header" ident="handNote" mode="delete"/>
<elementSpec module="header" ident="idno" mode="delete"/>
<elementSpec module="header" ident="interpretation" mode="delete"/>
<elementSpec module="header" ident="keywords" mode="delete"/>
<elementSpec module="header" ident="langUsage" mode="delete"/>
<elementSpec module="header" ident="language" mode="delete"/>
<elementSpec module="header" ident="namespace" mode="delete"/>
<elementSpec module="header" ident="notesStmt" mode="delete"/>
<elementSpec module="header" ident="principal" mode="delete"/>
<elementSpec module="header" ident="profileDesc" mode="delete"/>
<elementSpec module="header" ident="projectDesc" mode="delete"/>
<elementSpec module="header" ident="quotation" mode="delete"/>
<elementSpec module="header" ident="refState" mode="delete"/>
<elementSpec module="header" ident="refsDecl" mode="delete"/>
<elementSpec module="header" ident="rendition" mode="delete"/>
<elementSpec module="header" ident="samplingDecl" mode="delete"/>
<elementSpec module="header" ident="segmentation" mode="delete"/>
<elementSpec module="header" ident="seriesStmt" mode="delete"/>
<elementSpec module="header" ident="sponsor" mode="delete"/>
<elementSpec module="header" ident="stdVals" mode="delete"/>
<elementSpec module="header" ident="tagUsage" mode="delete"/>
<elementSpec module="header" ident="tagsDecl" mode="delete"/>
<elementSpec module="header" ident="taxonomy" mode="delete"/>
<elementSpec module="header" ident="textClass" mode="delete"/>
<elementSpec module="textstructure" ident="argument" mode="delete"/>
<elementSpec module="textstructure" ident="div1" mode="delete"/>
<elementSpec module="textstructure" ident="div2" mode="delete"/>
<elementSpec module="textstructure" ident="div3" mode="delete"/>
<elementSpec module="textstructure" ident="div4" mode="delete"/>
<elementSpec module="textstructure" ident="div5" mode="delete"/>
<elementSpec module="textstructure" ident="div6" mode="delete"/>
<elementSpec module="textstructure" ident="div7" mode="delete"/>
<elementSpec module="textstructure" ident="docAuthor" mode="delete"/>
<elementSpec module="textstructure" ident="docDate" mode="delete"/>
<elementSpec module="textstructure" ident="docEdition" mode="delete"/>
<elementSpec module="textstructure" ident="docImprint" mode="delete"/>
<elementSpec module="textstructure" ident="docTitle" mode="delete"/>
<elementSpec module="textstructure" ident="epigraph" mode="delete"/>
<elementSpec module="textstructure" ident="floatingText" mode="delete"/>
<elementSpec module="textstructure" ident="group" mode="delete"/>
```

```
<elementSpec module="textstructure" ident="imprimatur" mode="delete"/>
<elementSpec module="textstructure" ident="titlePart" mode="delete"/>
<elementSpec module="textstructure" ident="trailer" mode="delete"/>
<moduleRef key="analysis"/>
<elementSpec module="analysis" ident="cl" mode="delete"/>
<elementSpec module="analysis" ident="interp" mode="delete"/>
<elementSpec module="analysis" ident="interpGrp" mode="delete"/>
<elementSpec module="analysis" ident="m" mode="delete"/>
<elementSpec module="analysis" ident="phr" mode="delete"/>
<elementSpec module="analysis" ident="s" mode="delete"/>
<elementSpec module="analysis" ident="span" mode="delete"/>
<elementSpec module="analysis" ident="spanGrp" mode="delete"/>
<moduleRef key="dictionaries"/>
<elementSpec module="dictionaries" ident="colloc" mode="delete"/>
<elementSpec module="dictionaries" ident="dictScrap" mode="delete"/>
<elementSpec module="dictionaries" ident="entryFree" mode="delete"/>
<elementSpec module="dictionaries" ident="etym" mode="delete"/>
<elementSpec module="dictionaries" ident="hom" mode="delete"/>
<elementSpec module="dictionaries" ident="hyph" mode="delete"/>
<elementSpec module="dictionaries" ident="iType" mode="delete"/>
<elementSpec module="dictionaries" ident="lang" mode="delete"/>
<elementSpec module="dictionaries" ident="pRef" mode="delete"/>
<elementSpec module="dictionaries" ident="pVar" mode="delete"/>
<elementSpec module="dictionaries" ident="pron" mode="delete"/>
<elementSpec module="dictionaries" ident="stress" mode="delete"/>
<elementSpec module="dictionaries" ident="syll" mode="delete"/>
<moduleRef key="gaiji"/>
<elementSpec module="gaiji" ident="char" mode="delete"/>
<elementSpec module="gaiji" ident="charDecl" mode="delete"/>
<elementSpec module="gaiji" ident="charName" mode="delete"/>
<elementSpec module="gaiji" ident="charProp" mode="delete"/>
<elementSpec module="gaiji" ident="glyph" mode="delete"/>
<elementSpec module="gaiji" ident="glyphName" mode="delete"/>
<elementSpec module="gaiji" ident="localName" mode="delete"/>
<elementSpec module="gaiji" ident="mapping" mode="delete"/>
<elementSpec module="gaiji" ident="unicodeName" mode="delete"/>
<elementSpec module="gaiji" ident="value" mode="delete"/>
<moduleRef key="transcr"/>
<elementSpec module="transcr" ident="addSpan" mode="delete"/>
<elementSpec module="transcr" ident="am" mode="delete"/>
<elementSpec module="transcr" ident="damage" mode="delete"/>
<elementSpec module="transcr" ident="damageSpan" mode="delete"/>
<elementSpec module="transcr" ident="delSpan" mode="delete"/>
<elementSpec module="transcr" ident="ex" mode="delete"/>
<elementSpec module="transcr" ident="fw" mode="delete"/>
```

```

<elementSpec module="transcr" ident="handNotes" mode="delete"/>
<elementSpec module="transcr" ident="handShift" mode="delete"/>
<elementSpec module="transcr" ident="restore" mode="delete"/>
<elementSpec module="transcr" ident="space" mode="delete"/>
<elementSpec module="transcr" ident="subst" mode="delete"/>
<elementSpec module="transcr" ident="supplied" mode="delete"/>
<elementSpec module="transcr" ident="surface" mode="delete"/>
<elementSpec module="transcr" ident="zone" mode="delete"/>
<elementSpec module="core" ident="orig" mode="delete"/>
<elementSpec ident="w" module="analysis" mode="change">
  <attList>
    <attDef ident="orig" mode="change">
      <desc>gives the original string or is the empty string when the element
        does not appear in the source text.</desc>
      <datatype minOccurs="1" maxOccurs="1">
        <rng:text xmlns:rng="http://relaxng.org/ns/structure/1.0"/>
      </datatype>
    </attDef>
  </attList>
</elementSpec>
</schemaSpec>
</body>
</text>
</TEI>

```

Textkodierung ‚Der Aar‘

```
<?xml version="1.0" encoding="UTF-8"?>

<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="de">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Campe - Exemplarisches Wörterbuch, mit Header für Testzwecke</title>
        <author>Mirjam</author>
      </titleStmt>
      <publicationStmt>
        <availability>
          <p>für Testzwecke</p>
        </availability>
        <authority>Kompetenzzentrum für EDV-Philologie, Universität Würzburg</authority>
      </publicationStmt>
      <sourceDesc>
        <biblStruct xml:lang="de" type="dictionary">
          <monogr>
            <author>Joachim Heinrich Campe</author>
            <title>Wörterbuch der Deutschen Sprache</title>
            <title type="sub">Veranstaltet und herausgegeben von Joachim Heinrich Campe.
              Erster Theil. A - bis - E. </title>
            <imprint>
              <pubPlace>Braunschweig</pubPlace>
              <publisher>In der Schulbuchhandlung.</publisher>
              <date when="1807">1807</date>
            </imprint>
            <biblScope type="vol">1</biblScope>
            <biblScope type="pages" from="1" to="1">1</biblScope>
          </monogr>
        </biblStruct>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <editorialDecl>
        <normalization>
          <p>Formaterfassung der Lemmata:</p>
          <p>Alle Lemmata (Stichwort ggf. mit Artikel und Flexionsendungen) sind
            ausnahmslos in einem ca. 0,5 Punkt größeren Font gesetzt und ausgerückt;
            diese Formatierung wird nur an dieser Stelle notiert und nicht für jeden
            einzelnen Fall im Text durch <hi/> ausgezeichnet. </p>
          <p>Kodierung des Schriftfonts/-schnitts:</p>
          <p>Unterschieden wird zwischen normalem Schriftfont (unmarkiert) und größerer
            Schrift (markiert). Ebenso werden Antiqua- (markiert) und Frakturschrift
            /(unmarkiert) behandelt. Schriftsperrung wird ebenfalls markiert.</p>
        </normalization>
      </editorialDecl>
    </encodingDesc>
    <revisionDesc>
```

```
<change who="#MB">20060906 Erstkodierung</change>
<change who="#MB">20061108 Anpassung des Schemas</change>
<change who="#MB">20070420 Überarbeitung</change>
<change who="#MB">20070710 Überarbeitung</change>
<change who="#CS">20090630 Kodierung</change>
<change who="#WW">20091017 Textkontrolle</change>
<change who="#----">Übernahme in Bandkodierung</change>
</revisionDesc>
</teiHeader>
<facsimile>
  <graphic url="cwds1_c0002.tif"/>
</facsimile>
<text>
  <body>
    <div>
      <entry xml:id="cwds1_00045_aar">
        <lb n="0002.035"/>
        <usg type="time">
          <g>*</g>
        </usg>
        <form>
          <form type="lemma">
            <gramGrp>
              <pos value="noun"/>
              <gen value="m"/>
            </gramGrp>
            <form type="determiner">
              <orth>Der</orth>
            </form>
            <form type="headword" norm="aar">
              <orth>Aar</orth>
            </form>
          </form>
        </form>
        <pc>,</pc>
        <form type="inflected">
          <gramGrp>
            <case value="genitive"/>
            <number value="singular"/>
          </gramGrp>
          <form type="determiner">
            <orth>des</orth>
          </form>
          <form type="headword">
            <orth>
              <oVar>
                <oRef/> - es </oVar>
              </orth>
            </form>
            <pc>,</pc> oder
            <form type="headword">
              <orth>
                <oVar>
                  <oRef/> - en </oVar>
                </orth>
              </form>
            </form>
          </form>
        </form>
      </entry>
    </div>
  </body>
</text>
```

```
</orth>
</form>
</form>
<pc>,</pc>
<form type="inflected">
  <gramGrp>
    <case value="nominative"/>
    <number value="plural">
      <abbr>Mz.</abbr>
    </number>
  </gramGrp>
  <form type="determiner">
    <orth>die</orth>
  </form>
  <form type="headword">
    <orth>
      <oVar>
        <oRef/> - e </oVar>
      </orth>
    </form>
    <pc>,</pc> oder
    <form type="headword">
      <orth>
        <oVar>
          <oRef/> - en </oVar>
        </orth>
      </form>
    </form>
    <pc>,</pc>
  </form>
</sense>
<def> die alte <lb n="0002.036"/> Benennung aller großer Raubvögel
  <pc>,</pc> besonders aber des Adlers <pc>,</pc>
</def> die <lb
  n="0002.037"/> noch <usg type="geo"> in <abbr>N. D.</abbr>
  </usg> üblich ist und <usg type="style">bei Dichtern</usg> vorkommt
  <pc>.</pc>
  <lb n="0002.038"/>
  <cit rend="indented">
    <cit type="quote">
      <quote> Ein kühner <hi rend="large">Aar</hi> theilt mit gewalt'gen
        Schwingen <lb n="0002.039"/> Die Lüfte <pc>,</pc>
        <pc>-</pc> <pc>-</pc> <pc>-</pc> <pc>-</pc> <pc>-</pc>
      </quote>
    </cit>
    <bibl>
      <author n="#Schreiber">
        <hi rend="spaced">Schreiber</hi>
      </author>
      <pc>.</pc>
    </bibl>
  </cit>
  <lb n="0002.040"/>
  <cit type="quote">
```

```
<quote>Bald werdet ihr im Meer der Haien <pc>,</pc> am Gestade <lb  
n="0002.041"/> Der <hi rend="large">Aaren</hi> Beute sein  
<pc>.</pc> <pc>-</pc>  
</quote>  
<bibl>  
  <author n="#Ramler">  
    <hi rend="spaced">Ramler</hi>  
  </author>  
  <pc>.</pc>  
</bibl>  
</cit>  
</cit>  
</sense>  
</entry>  
</div>  
</body>  
</text>  
</TEI>
```