

# Historisches Zeitungskorpus

Protokoll der Doppelerfassung (Stand 27.11.2010)

## 1. Erfassungsumfang

1.0 Erfassungsvorlage: Scan-Images

1.1 Erfassungsumfang (Die Scan-Vorlage befindet sich auf 4 DVDs)

DVD	Erfassungsnr.	Ordnername	Image-Nr. von ... bis	Gesamtzahl der Image-Seiten	Bemerkung
DVD1	01	Allerneueste Europäische Welt- und Staatsgeschichte	Weltgeschichte0001—Weltgeschichte0161	161	
	02	Allgemeine Auswandererzeitung Rudolstadt1846-1848	Auswanderer0001—Auswanderer0317	317	
	03	Allgemeine Zeitung	Augsburg0001—Augsburg0139	139	
	04a	Arbeitgeber	Arbeitgeber(1)0001—Arbeitgeber(1)0114	114	2 Teile
	04b	Arbeitgeber	Arbeitgeber(2)0001—Arbeitgeber(2)0108	108	
	05	Bayerische Presse	Bayerische Presse0001—Bayerische Presse0187	187	
DVD2	06	Bayreuther Zeitungen	Bayreuther0001—Bayreuther0052	52	
	07	Berlinische Nachrichten	BerlinN00001—BerlinN00143	143	
	08	Berlinische Privilegierte Zeitung	BerlinPZ0001—BerlinPZ0080	80	
	09a	Conversationsblatt	Conversationsblatt(1)0001—Conversationsblatt(1)0042;	42	3 Teile
	09b	Conversationsblatt	Conversationsblatt(2)0001—Conversationsblatt(2)0026;	26	
	09c	Conversationsblatt	Conversationsblatt(3)0001—Conversationsblatt(3)0017	17	
	10	Das Heller-Blatt	Heller0001—Heller0083	83	
	11a	Das Pfennig-Magazin	Pfennig0001—Pfennig0094;	94	3 Teile
	11b	Das Pfennig-Magazin	Pfennig(2)0001—Pfennig(2)0093;	93	
	11c	Das Pfennig-Magazin	Pfennig(3)0001—Pfennig(3)0099	99	
	12	Das wohlfeilste Panorama des Universums Prag	Das wohlfeilste Panorama des Universums00001—Das wohlfeilste Panorama des Universums00235	235	
DVD3	13	Deutsche Auswandererzeitung	Auswandererzeitung0001—Auswandererzeitung0116	116	
	14	Europa	Europa0001—Europa0242	242	
	15a	Mainzer Journal	MainzerJournal(1)0001—MainzerJournal(1)0326;	326	2 Teile
	15b	Mainzer Journal	MainzerJournal(2)0001—MainzerJournal(2)0098	98	
	16	Morgenblatt für gebildete Leser	Morgenblatt0001—Morgenblatt0114	114	
	17a	Märkische Blätter	Märkische(1)0001—Märkische(1)0052;	52	2 Teile
	17b	Märkische Blätter	Märkische(2)0001—Märkische(2)0087	87	
	DVD4	18a	Social-Politische Blätter Berlin	Social-Politische Blätter(1873)0001—Social-Politische Blätter(1873)0037;	37

18b	Social-Politische Blätter Berlin	Social-Politische Blätter(1874)0001— Social-Politische Blätter(1874)0073	73	
19	Sonntagsblatt	Sonntagsblatt0001—Sonntagsblatt0208	208	
20	Tübinger Chronik	Chronik0001—Chronik0108	108	
21	Wiener Zeitung	Wiener0001—Wiener0082	82	

## 2. Dateinamenkonventionen

HZNNMXXY.ZZZ

HZ: Kürzel für das Projekt Historisches Zeitungskorpus

NNM: NN = Nr. der Erfassungseinheit; M = 0 oder Unterscheidungsbuchstabe a, b, c bei mehreren Teilen (s. Tabelle oben, Sp. 2).

XX: 2-stellige Nummer zur Bezeichnung der in der Datei erfassten Imageseiten, ggf. mit führender Null, Muster: 01 (= Image-S. 1-10), 02 (= Image-S. 11-20), 03 (= Image-S. 21-30) ... 10 (= Image-S. 91-100), 11 (= Image-S. 101-110) ... D.h. Die Zahl XX mal 10 entspricht i.d.R. der letzten in der Datei erfassten Imageseite.

Y: Erfassungsversion A, B, C

ZZZ: Kürzel für interne Arbeitsorganisation der Datenerfassung, für die weitere Datenverarbeitung irrelevant und kann zur Bezeichnung des Dateiformats nutzbar gemacht werden.

Dateiportion: 10 Image-Seiten (Die letzte Datei einer Erfassungseinheit kann davon abweichen)

## 3. Sonderzeichen

### 3.1 Allgemeines zur Sonderzeichenkodierung

- a) Die Kodierung der Sonderzeichen richtet sich soweit wie möglich nach TUSTEP-Konventionen. Der phonetische Zeichensatz (**#P+** ... **#P-**) kommt nur dann zum Einsatz, wenn ein Sonderzeichen weder mit ^ noch mit # oder % kodiert werden kann. Eine Abweichung davon stellen die Gedankenstriche dar, die jeweils durch **zwei Bindestriche** statt durch ^- nach TUSTEP (zur besseren optischen Kontrolle während der Eingabe) wiedergegeben sind. Projektspezifisch können Striche vorkommen, die noch länger als Gedankenstriche sind. Solche Striche werden dann **durch drei Bindestriche** wiedergegeben.

Hinweis: Es gibt innerhalb des TUSTEP-Zeichenvorrats manchmal mehrere Möglichkeiten, ein und dasselbe Zeichen auf zwei verschiedene Weisen zu kodieren, v. a., wenn man vom äußeren Bild des Zeichens ausgeht und die sachliche Bedeutung des Zeichens nicht erkennt. Z.B. ein hochgestelltes Zeichen X kann sowohl mit #H:X#G: als auch mit #'X, ein hochgestellter kleiner Kreis sowohl mit hochgestelltem Buchstaben o als auch mit #.\* dargestellt werden. Ein weiteres Beispiel ist das über einem Buchstaben übergesetzte v (#,vX oder %>X). Diese Fälle sollten bei der Weiterverarbeitung systematisch ermittelt, gesammelt und einheitlich behandelt werden.

- b) Mehrere Zeilen umfassende große geschweifte Klammern werden mit dem TUSTEP-Kode  **#(GAG)** (Klammer auf) und dem TUSTEP-Kode  **#(GZG)** (Klammer zu) abgebildet und für jede betroffene Zeile einschließlich des Inhalts wiederholt. In ähnlicher Weise wird die Kodierung  **#(RZG)** für die mehrere Zeilen große runde Klammer zu auch für jede betroffene Zeile wiederholt (bitte nachbearbeiten).

- c) Die über TUSTEP-Konventionen hinausgehenden, selbst definierten Kodierungen werden i.d.R. mit einer öffnenden geschweiften Klammer eingeleitet und mit einer schließenden geschweiften Klammer beendet (**{...}**), wobei zwischen Klein- u. Großschreibung i. d. R. nicht unterschieden wird. In besonderen Fällen, in denen das oder die übergesetzten Zeichen nicht dem oder den darunter stehenden Zeichen einzeln zuzuordnen sind, kommt die Kodierungsform  **<sup char="X">Y</sup>** zum Einsatz, wobei X und Y jeweils für ein oder mehrere Zeichen steht und X über Y gesetzt ist. Die Kodierungszeichen werden sinnvollerweise auf den ASCII-Standard-Zeichensatz beschränkt. Dabei kann die Kodierung eines über den TUSTEP-Zeichensatz hinausgehenden und mit geschweiften Klammern kodierten Sonderzeichens wegen der optischen Ähnlichkeit an ein anderes, aber ähnliches und nach TUSTEP kodierbares Sonderzeichen angelehnt definiert werden („tustep-analoge Sonderzeichenkodierung“).

Hinweis für die Datenkonvertierung: Die Kodierungszeichen sind für die Konvertierung in das TUSTEP-Format (das auf dem ASCII-Zeichensatz [7-bit, 128 Zeichen inkl. 32 Steuerzeichen] basiert) unproblematisch. Wenn die Daten aus dem TUSTEP-Format wiederum in ein anderes, nicht auf dem ASCII-Zeichensatz basierendes Dateiformat, etwa WINWORD, umgewandelt werden, sollten die „tustep-analoge“ Sonderzeichenkodierungen (s.o.) in den geschweiften Klammern als einzelne ASCII-Zeichen wie in ursprünglichen Erfassungsdaten ausgegeben werden (sonst wäre evtl. eine Anpassung der Tabelle der Sonderzeichenkodierungen im Erfassungsprotokoll nötig) und nicht wie die anderen normalen, nach TUSTEP-Konventionen mit mehreren ASCII-Zeichen kodierten Sonderzeichen behandelt werden. Denn diese können ja je nach der Gestaltung der Umwandlungsmodus (Codepage) in ein einziges Zeichen, d.h. in das im Zeichenvorrat des Zielformats verfügbare Sonderzeichen überführt werden. So kann bei einer ungetrennten Behandlung beispielsweise die „tustep-analoge“ Kodierungszeichenfolge  **{^+}**  zu  **{†}**  konvertiert werden, weil  **^+**  die TUSTEP-Kodierung für  **†**  darstellt.

- d) Bei Sternchen können innerhalb eines Erfassungsprojektes Varianten mit unterschiedlicher Zackenzahl vorkommen: 5-, 6-, 7-, 8zackige Sternchen. Falls eine Unterscheidung gewünscht ist, wird so ver-

fahren: Das 6zackige Sternchen wird direkt von der Tastatur eingegeben, also nicht gesondert kodiert, die anderen jeweils explizit mit {5\*}, {6\*}, {7\*}, {8\*}.

- e) Der Tag {?VZIFU} [V=Erfassungsversion A,B,C] steht als allgemeine Marke für beliebiges Sonderzeichen / Symbol ohne spezifisch definierte Kodierung, weil diese wegen der Seltenheit bei der Erfassung sich nicht lohnt).

### 3.2 Liste der Sonderzeichen (s. Anhang 1)

Bei dem Anhang 1 handelt sich es sich nicht um eine erschöpfende, sondern um eine vorläufige Liste. Eine erschöpfende Zusammenstellung ist nur nach Abschluss der Erfassung auf der Grundlage der korrigierten Gesamtdaten mithilfe von Programmen möglich und sinnvoll.

## 4. Liste der für die Doppelerfassung definierten Standard-Kodierungen zur Abbildung typographischer Gestaltungsmerkmale (s. Anhang 2)

### 5. Allgemeines zur Erfassung der Frakturschriften

- 5.1 Falls in der Fraktur für die Großbuchstaben **I** und **J** kein Unterschied gemacht wird (das ist die Regel), wird bei der Erfassung einheitlich immer der Großbuchstabe **J** eingetippt statt **I** oder mal **I**, mal **J**. Andernfalls wird wie üblich **I** als **I**, **J** als **J** wiedergegeben.
- 5.2 In der Fraktur können Umlaute durch zwei Punkte über den Vokalen (selten) oder durch den übergesetzten Buchstaben **e** (häufig) dargestellt werden, die Umlaute mit **e** werden bei der Erfassung direkt in heutige Umlaute **ä ö ü** umgesetzt. Manchmal aber kann dieses **e** infolge der schlechten Druckqualität genau so wie ein **c** aussehen. Zur Minimierung des Korrekturaufwandes wird für diese Fälle die Anweisung formuliert, dass ein **c** über **a**, **o**, und **u** bei Frakturschriften als **e** zu lesen ist und die entsprechenden Buchstaben als Umlaute interpretiert werden.
- 5.3 Ligaturen: **s+z** als Eszett **ß** behandelt; **t+z** als **tz**.
- 5.4 Das rechts schräg nach oben gestellte Gleichheitszeichen, das in Frakturschriften üblicherweise als Silbentrennungszeichen (oder in der Funktion eines Ergänzungsstriches) verwendet wird, wird als gewöhnliches Gleichheitszeichen eingetippt. Die normalen Gleichheitszeichen, die in der Regel im Frakturdruck sehr lang sind, werden zur Unterscheidung durch doppelte Gleichheitszeichen abgebildet.
- 5.5 Wechsel zwischen Fraktur und Antiqua innerhalb eines Wortes (Zwiebelfisch): Der Wechsel wird genauso wiedergegeben wie gedruckt.

## 6. Besonderheiten

- Schriftarten: Die verschiedenen Frakturschriften werden nicht unterschieden; die Frakturgrundschrift wird nicht explizit markiert, die normale Antiqua-Schrift wird mit <A> ... </A> kodiert. Die kleineren oder größeren Schriftgrade werden in Form von <A±X> ... </A±X> oder <W±X> ... </W±X> kodiert (s. Anhang 2).
- Die Großbuchstaben der Fraktur I und J sind im Druckbild im Ggs. zu sonstigen Frakturdruckerzeugnissen deutlich unterschieden, deshalb werden sie auch so unterschieden erfasst.
- Die verschiedenen Textrahmungen werden nicht wiedergegeben, lediglich der so ausgezeichnete Texteinheit wird mit der Kodierung <BOX> ... </BOX> zusammengehalten. Siehe z.B. die Erfassungseinheit Nr. 2, Image-Seite 0024.
- Lange Gleichheitszeichen werden durch zwei normale Gleichheitszeichen wiedergegeben.
- Kopfzeile und Seitenreferenzen: Weil die gedruckten Seitenzahlen auf Zeitungsseiten für die Datenerfassung und Datenstruktur kein geeignetes Referenzsystem bilden (sie fehlen ja auch sehr oft), wird (mit Ausnahme der anfänglichen Erfassungseinheiten der Nr. 01-05 der DVD 1, bei denen das Problem noch nicht systematisch und einheitlich geregelt wurde) die folgende einheitliche Regelung durchgeführt: Neben der standardmäßig kodierten Aufnahme des Imagedateinamens (inkl. der darin enthaltenen fortlaufenden Image-Nr). in der ersten Zeile wird zur Erleichterung der Abgleich- und Korrekturarbeiten immer in einer zweiten Zeile, die durch die Anfangskennung @@1@ eingeleitet wird, die Image-Nr. aus dem Imagedateinamen (ohne die Nullen) entnommen und in Form von <SExxx> eingegeben (xxx = die Image-Nr. aus dem Imagedateinamen), s. u. Anh. 2; Falls ein Image zwei Druckseiten enthält, werden die linke und die rechte Seite durch den zusätzlichen Buchstaben a oder b, also <SExxxa> oder <SExxxb>, unterschieden. Falls die Zeitungsseite einen Kopftext (mit oder ohne Seiten-/Spaltenzahl) aufweist, wird diese Information als reine Textdaten (ohne gesonderte Kodierung der Seiten-/Spaltenzahl, falls diese in der Kopfzeile vorhanden ist) in einer dritten, durch die Kennung @@1b@ eingeleiteten Zeile aufbewahrt. Bsp. für
  - Seiten ohne Kopfzeile: Erfassungsnr. 08, Image-Nr. 0002 (= Imagedateiname: BerlinPZ0002.jpg);
  - Seiten mit Kopfzeile und Seitenzahl: Erfassungsnr. 11b, Image-Nr. 0033 (= Imagedateiname: Pfennig(2)0033.jpg);

- Images mit 2 Druckseiten, jeweils mit Kopfzeile und Seitenzahl: Erfassungsnr. 10, Image-Nr. 0004 (= Imagedateiname: Heller0004.jpg);
- Images mit 2 zweispaltigen Druckseiten, jeweils mit Kopfzeile und ohne Spaltenzahlen: Erfassungsnr. 07, Image-Nr. 0002 (= Imagedateiname: BerlinN00002.jpg);
- Images mit 2 zweispaltigen Druckseiten, jeweils mit Kopfzeile und Spaltenzahlen: Erfassungsnr. 09, Image-Nr. 0004 (= Imagedateiname: BerlinN00002.jpg).

Zu den Nrr. 01-05 der DVD 1: Die Erfassungsnummern HZ01-04 enthalten meistens gedruckte Seitenzahlen, diese wurden – nach der ersten Zeile mit dem Imagedateinamen – in der zweiten Zeile mit der Anfangskennung @@1@ in Form von <Sxxx> eingetippt (falls die Paginierung fehlt, fehlt auch eine solche Zeile); nach der obigen einheitlichen Regelung müsste also erstens @@1@ in @@1b@ geändert, <Sxxx> getilgt werden, zweitens vor dieser Zeile eine Zeile mit @@1@ und der Nr. aus dem Imagedateinamen in Form <SExxx> ergänzt werden. Bei HZ 05 haben die Zeitungssseiten keinen Kopftext mit Seitenzahlen, die hinter der Zeilenkennung @@1@ eingetippte <Sxxx> ist eine künstliche Seitennr. 1 bis 10 zur Fixierung der Seitenreihenfolge – ohne Rückgriff auf die im Imagedateinamen enthaltene Nummerierung, nach der obigen einheitlichen Regelung müsste also erstens die Nr. xxx hier durch die Nr. aus dem Imagedateinamen ersetzt werden.

– Einzelnes:

Erfassungsnr. 15a, Image-Nr. 0122, rechte Spalte unten: 3 kleine Tabellen in einer großen eingebettet, also Verschachtelung von der Tabellenkodierung <biaoge> ... </biaoge>.

Erfassungsnr. 20, Image-Nr. 0002 und ähnl. Seiten: Die Seitenzahlen (z.B. hier 22) sind durch ungewöhnliche Symbole (zwei Pfeile) statt durch lange Striche eingerahmt, diese Symbole werden beim Abtippen ignoriert.

Erfassungsnr. 20, Image-Nr. 0009 ganz oben: Die 3 Teile werden als 3 Spalten behandelt und die rechteckigen Rahmen mit <BOX> ... </BOX> kodiert, zuletzt mit typogr. Zweifelskennung versehen.

Erfassungsnr. 20, Image-Nr. 0064 linke Seite und 0066: Textverlauf senkrecht (von unten nach oben oder von oben nach unten), diese Textblöcke werden als normaler Text eingegeben, aber mit <BOX> ... </BOX> eingeklammert und mit typogr. Zweifelskennung versehen.

Anhang 1: Liste der Sonderzeichen im Proj. HZ (Zu Beispielvorkommen werden entweder die Imagedateinamen oder Referenzen in Form von Erfassungsnr. + Image-Nr. + linke/rechte Sp./S. + Zeile angegeben)

a. Nach TUSTEP-Konventionen kodierte Sonderzeichen

Sonderzeichen	TUSTEP-Kode	Beispiel-Vorkommen		Beispiel-Ausschnitt, Bemerkung
	#./	DVD1	Augsburg0007	
^	#(FA)	DVD1	Augsburg0035	
†	#(UGLS)	DVD1	Augsburg0134	
†	^+	DVD1	Augsburg0033	
≡	#(ENT)	DVD1	Augsburg0044	
▽	#(NBL)	DVD1	Augsburg0111	
♀	%++o	DVD1	Arbeitgeber(2)0011	
♂	%+o	DVD1	Arbeitgeber(2)0087	
×	^*	DVD1	Arbeitgeber(2)0080	
□ □	#(QUA)	DVD1	Auswanderer0172	Quadrate unterschiedlicher Größen nicht unterschieden
↑	#(PFO)	DVD1	Augsburg0118	
✂	#G+k#G-	DVD1	Augsburg0035	

)	#(RZG)	DVD4	17b-0038-rechts-21	Hofraum Wohnhaus ) (Kodierung für jede betroffene Zeile wiederholt)
X	#(KP)	DVD3	Mainzer journal (1) 0009	
X	#(KPG)	DVD3	Morgnblatt0063	
Φ	#G+f#G-	DVD3	Mainzer journal (1) 0107	
//// ... ///	^'^^'^'^' ...''''	DVD3	MainzerJournal (1)0014	
Π	#(PRG)	DVD1	Allgemeine Zeitung 0033	

b. An TUSTEP-Konventionen angelehnte Sonderzeichen-Kodierungen

Zeichen, Symbol	Kodierung	Beispiel-Vorkommen		Bemerkung
	{#(PFLR)}	DVD1	Augsburg0083	
	{#(WU)}	DVD3	Mainzer journal (1) 0090	
	{#(PFR)}	DVD4	Chronik0018	
	{#(PF)}	DVD1	Arbeitgeber(2)0010	

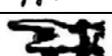
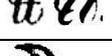
c. Selbstdefinierte Kodierungen

c1 Währungskürzel, -symbole

Zeichen, Kürzel	Kodierung	Beispiel-Vorkommen		Beispiel-Ausschnitt (und Bemerkung)
	{Sgr.}	DVD1	Auswanderer0024	
	{Sgr.}	DVD1	Auswanderer0066	pr. Gr. nur 15 Igr. zu
	{Thlr.}	DVD1	Auswanderer0024	
	{Rl.}	DVD1	Auswanderer0001	Postanstalten 1; Rb. == 2 fl. 6 W!
	{fl}	DVD1	Auswanderer0001	Postanstalten 1; Rb. == 2 fl. 6 W!
	{fl}	DVD4	20-75b-rechts-↑21	8 fl 6 W.
	{Xr.}	DVD1	Auswanderer0001	Postanstalten 1; Rb. == 2 fl. 6 W!
	{Xr}	DVD3	13-0001-a-7	6.lr (bitte prüfen, ob sachl. korrekt)
	{Xr.}	DVD4	20-75b-rechts-↑21	8 fl 6 W.
	{xf}	DVD3	Auswandererzeitung0008	1 1/3 x\$ 2 fl. 24 Xr.
	{S//}	DVD3	Auswandererzeitung0015	550 \$ ausgegeben

	{mrk.}	DVD3	Auswandererzeitung0028	
	{ß}	DVD3	Auswandererzeitung0028	
	{ggl}	DVD3	Auswandererzeitung0004	
	{gl}	DVD3	13-0001-a-5	36 36 (bitte prüfen, ob sachl. korrekt)
	{fl}	DVD3	Auswandererzeitung0008	
	{D}	DVD3	Auswandererzeitung0073	
	{Ctr.}	DVD1	Auswanderer0066	pr. Ctr. nur 15 Jgr. zu

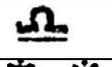
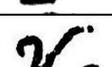
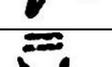
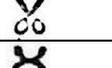
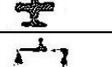
c2. Sonstiges

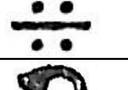
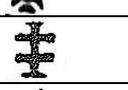
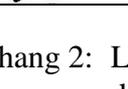
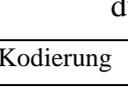
Kürzel, Symbol u.Ä.	Kodierung	Beispiel-Vorkommen		Beispiel-Ausschnitt, Bemerkung
	{et cetera}	DVD1	Weltgeschichte002	
	{sj.}	DVD1	Augsburg0016	
	{sj}	DVD1	Arbeitgeber(1)0026	
	{sj-}	DVD1	Arbeitgeber(2)0004	
	{sjh}	DVD1	Augsburg0046	
	{sj-xie}	DVD1	Arbeitgeber(2)0033	
	{SHY}	DVD1	Augsburg0093	
	{**}	DVD1	Augsburg0009	
	{./.}	DVD1	Augsburg0050	
	{shou-zuo}	DVD1	Auswanderer0082	
	{shou-you}	DVD1	Auswanderer0082	
	{#}	DVD1	Augsburg0029	
	{§}	DVD1	Auswanderer0062	
	{Numero}	DVD1	Auswanderer0009	
	{P}	DVD1	Auswanderer0003 Auswanderer0030	
	{yueya}	DVD3	MainzerJournal(2)0022	
	{ri}	DVD3	MainzerJournal(1)0071	

	{sj+}	DVD1	Bayerische Presse0013	
	{8*}	DVD1	Bayerische Presse0038	
	{7*}	DVD3	13-0006-b-↑16	
	{Numero}	DVD3	15a-0011-2	bitte prüfen, ob sachl. korrekt
		DVD3	15a-0001-7	  (bitte prüfen, ob sachl. korrekt)
	{J}	DVD3	15a-0007-9	
ten \\	<sup char="ten">\\ </sup>	DVD3	17b-0050-links-29	 ten \\
	{7*}		13-0006-b-↑16	

- d. Durchnummerierte oder nur pauschal markierte Zeichen, Symbole in der Kodierungsform „{Symbol-x}“ (x = fortlaufende Nr. oder undifferenziert als x)

Symbol	Kodierung	Beispiel-Vorkommen		Bemerkung
	{Symbol-1}	DVD1	Augsburg0039	
	{Symbol-2}	DVD1	Augsburg0111	
	{Symbol-3}	DVD1	Augsburg0002	
	{Symbol-4}	DVD1	Augsburg0002	
	{Symbol-5}	DVD1	Augsburg0005	
	{Symbol-6}	DVD1	Augsburg0006	
	{Symbol-7}	DVD1	Augsburg0007	
	{Symbol-8}	DVD1	Augsburg0009	
	{Symbol-9}	DVD1	Augsburg0015	
	{Symbol-10}	DVD1	Augsburg0021	
	{Symbol-11}	DVD1	Augsburg0023	
	{Symbol-12}	DVD1	Augsburg0036	
	{Symbol-13}	DVD1	Augsburg0029	
	{Symbol-14}	DVD1	Augsburg0032	
	{Symbol-15}	DVD1	Augsburg0052	
	{Symbol-16}	DVD1	Augsburg0059	
	{Symbol-17}	DVD1	Augsburg0055	

	{Symbol-18}	DVD1	Augsburg0059	
	{Symbol-19}	DVD1	Augsburg0066	
	{Symbol-20}	DVD1	Augsburg0064	
	{Symbol-21}	DVD1	Augsburg0127	
	{Symbol-22}	DVD1	Augsburg0068	
	{Symbol-23}	DVD1	Augsburg0069	
	{Symbol-24}	DVD1	Arbeitgeber(2)0107	
	{Symbol-25}	DVD1	Augsburg0095	
	{Symbol-26}	DVD1	Augsburg0001	
	{Symbol-27}	DVD1	Augsburg0070	
	{Symbol-28}	DVD1	Augsburg0076	
	{Symbol-29}	DVD1	Augsburg0090	
	{Symbol-30}	DVD1	Augsburg0090	
	{Symbol-31}	DVD1	Augsburg0116	
	{Symbol-32}	DVD1	Augsburg0123	
	{Symbol-32}	DVD3	Mainzer journal (1) 0009	
	{Symbol-33}	DVD3	Mainzer journal (1) 0078	
	{Symbol-34}	DVD3	Mainzer journal (1) 0121	
	{Symbol-35}	DVD3	Mainzer journal (1) 0168	
	{Symbol-36}	DVD3	MainzerJournal(1)0040	
	{Symbol-37}	DVD3	MainzerJournal(1)0061	
	{Symbol-38}	DVD3	MainzerJournal(2)0013	
	{Symbol-39}	DVD3	MainzerJournal(2)0026	
	{Symbol-40}	DVD3	MainzerJournal(2)0026	
	{Symbol-41}	DVD3	MainzerJournal(2)0076	
	{Symbol-42}	DVD1	Bayerische Presse 0032	
	{Symbol-43}	DVD1	Bayerische Presse 0158	

	{Symbol-x}	DVD3	Mainzer journal (1) 0009	
	{Symbol-x}	DVD3	Mainzer journal (1) 0078	
	{Symbol-x}	DVD3	Mainzer journal (1) 0121	
	{Symbol-x}	DVD3	Mainzer journal (1) 0168	
	{Symbol-x}	DVD3	MainzerJournal(1)0040	
	{Symbol-x}	DVD3	MainzerJournal(1)0061	
	{Symbol-x}	DVD3	MainzerJournal(2)0013	
	{Symbol-x}	DVD3	MainzerJournal(2)0026	
	{Symbol-x}	DVD3	MainzerJournal(2)0026	
	{Symbol-x}	DVD3	MainzerJournal(2)0076	

Anhang 2: Liste der für die Doppelerfassung definierten Standard-Kodierungen zur Abbildung typographischer Gestaltungsmerkmale

Kodierung	Erläuterung
@@0@ <filename: ... />	In eigener Zeile zur Aufnahme der Bilddateinamen, wenn die Textvorlage nicht in Papierform (Buch od. Kopie), sondern in Scans zur Verfügung gestellt wird und die Mitführung der Dateinamen gewünscht wird.
@@1@	Kopfzeitext/Kopfzeile; Kennung steht am Anfang der Zeile. Eventuelle Schriftgrößenunterschiede im Bereich des Kopfzeitextes werden i.d.R. ignoriert. Diese Kennung ist obligatorisch.
<Sxxx>	Seite xxx; Seitenzahlen werden immer in der Kopfzeitextzeile wiedergegeben, auch wenn sie unten auf der Seite gedruckt sind. S steht auch für Spalten, falls die Vorlage Spaltenzählung statt Seitenzählung aufweist. Nur für HZ01-04 zur Aufnahme der gedruckten Seitenzahlen verwendet (Vgl. oben 6).
<SExxx>, <SExxxa> oder <SExxxb>	Beim Erfassen ergänzte Seitenzahl xxx (E steht für Ergänzung), als xxx wird die Nr. aus dem Imagedateinamen (ohne Nullen) entnommen. Bei Images, die 2 Druckseiten abbilden, werden die linke und die rechte Seite durch den Buchstaben a oder b unterschieden: a für die linke Seite, b für die rechte. Die Unterscheidung notwendig, weil beide Druckseiten eine Imageseitennr. teilen. (Vgl. oben 6).
@@1b@	Kennung der Zeile, auf der der Kopfzeitext mit der Paginierung als Information aufbewahrt wird, falls die Zeitungseite einen Kopfzeitext mit der Paginierung haben.
@@3@	Fußzeitext (Fußnoten) Anfang (Endekennung in der Kodierung der Kopfzeitextzeile der nächsten Seite implizit enthalten); Kennung nur verwendet, falls Fußzeitext (Fußnoten) vorhanden und erfasst werden soll.
<duolan X> ... </duolan X>	Seitentext in X Spalten gesetzt; X=2,3 ...
@@	Ende einer Spalte im mehrspaltig gesetzten Bereich einer Seite
<P>-- ... </P>	Absatzzeitext 1: Einzug der ersten Zeile (--: zwei feste Blanks)
<P>... </P>	Absatzzeitext 1a: kein Einzug der ersten Zeile (auch einzeiliger Absatz ohne Einzug oder mehrzeiliger Absatz ohne Einzug, die keinem Absatzzeitext zugeordnet werden können)
<negEZ>... </negEZ>	Absatzzeitext 2: negativer Einzug der ersten Zeile

<Z> ... </Z>	Absatztyp 4: alle Zeilen des Absatzes zentriert	Achtung: Bei der Eingabe werden weiterhin zusätzlich die Direktformatierungsfunktionen von WORD für zentrierte und für rechtsbündige Ausrichtung ganzer Zeile(n) verwendet, das dient jedoch nur noch zur optischen Kontrolle. D.h. diese Codes sind überflüssig für die Abbildung der Daten und nach der Konvertierung entsprechend zu bereinigen.
<R> ... </R>	Absatztyp 5: alle Zeilen des Absatzes rechtsbündig. Außerdem kann diese Kodierung auch für rechtsbündige Ausrichtung des rechten Teils oder eines Teils einer Zeile verwendet.	
<-P>	Spalten- od. Seitenumbruch, der einen Absatz (im Haupttext oder im Fußtext) in zwei Teile trennt. Kodierung paarweise verwendet zur Erleichterung der Weiterverarbeitung der Daten: einmal am Ende des ersten Teils eines Absatzes, einmal am Anfang des zweiten Teils desselben Absatzes, z. B.: <P--> ... <-P> <-P> ... </P>	
<E> ... </E>	Einzug erster Stufe	für nicht komplexe Einzüge, die Angabe der
<E+1> ... </E+1>	Einzug zweiter Stufe	Einzugstiefe orientiert sich am linken Satzspiegelrand.
<shi> ... </shi>	Verstext. Kennung nur dann eingesetzt, wenn Vers u. Prosa gemischt vorkommen und die Prosa den Hauptteil des Textes ausmacht. Diese Kennungen bedürfen aber besonderer Kontrolle, da sie nur Anhaltspunkte für sachlich korrekte Auszeichnungen sein können.	
Einzug innerhalb von Verszeilen	Einzug erster Stufe mit 2 festen Blanks, Einzug zweiter Stufe mit 4 festen Blanks.	
<biaoge> ... </biaoge>	Tabellenanfang, -ende; Spaltenkodierung mit &tNX (s. u.); Der so markierte Tabellentext wird ohne Berücksichtigung der Absatzgrenzen zeilenweise abgetippt, vertikal verlaufende Zeilen als horizontaler Text eingegeben. Für die genaue Wiedergabe der Originaltabellenstruktur müssen also die Tabellen nach Abgleich u. Korrektur nachbearbeitet werden.	
&tNX	&t: Tabulator; N: Spalte 1, 2, 3 ...; X: Ø=linksbündig, Z=zentriert, R=rechtsbündig	
<tspb> ... </tspb>	Besonderes Layout, genaue Abbildung mit vorgegebenem Kodierungsapparat nicht möglich, Text zeilenweise abgetippt, für Layout-Abbildung vielfach feste/normale Blanks verwendet („harte Formatierung“)	
Unterstreichung in WORD	Ersatzkodierung für Sperrung, bitte bei der Konvertierung beachten. (Zur Wiedergabe der Unterstreichung werden in WORD direkt TUSTEP-Kodierungen #1+ ... #1- bzw. #2+ ... #2- benutzt.)	
<konghang>	Leerzeile(n) bzw. größerer Durchschuss; Anzahl der Leerzeilen nicht ermittelbar; als Information oft entbehrlich.	
<FT> ... </FT>	Fußtexte / Fußnoten, die innerhalb des Seitenhaupttextes statt am Seitenfuß platziert sind. (etwa wenn eine Buchseite als Abbildung auf einer Seite im Haupttext zitiert wird)	
<kongge>	Relevanter, großer Wortzwischenraum ohne erkennbare Regel.	
&duan&	Zeilenüberlauf; vor der Kennung zusätzl. 2x Tabulator	
<A±X> ... </A±X>	Antiqua: relative Schriftgröße; X =1,2,3 ...; + größer; - kleiner. Bei X=0, gilt die Kodierungsform <A> ... </A>, aber Grundschrift (d.h. die im gesamten Text am meisten verwendete Schrift [A oder F]; X=0) wird nicht explizit kodiert.	
<F±X> ... </F±X>	F=W: Fraktur Waldbaum; F=S Fraktur Schwabbacher ... [Diese Buchstabenkürzel sollen nur die verwendeten Schriften voneinander unterscheiden und müssen nicht korrekt sein im Sinne der typogr. Schriftbestimmung]; X: wie oben. Bei X=0, gilt die Kodierungsform <F> ... </F>, aber Grundschrift (d.h. die im gesamten Text am meisten verwendete Schrift [A oder F], vgl. oben; X=0) wird nicht explizit kodiert.	
<INIX>Y</INIX>	Großes Zeichen Y, das mehrere Zeilen hoch ist (i.d.R. Initialen); X=1,2,3... Zeilen	

#(IMAGE)	Abbildung, Graphisches.
<BOX> ... </BOX>	Durch eine Rahmung ausgezeichneter Textblock als Einheit
<Grafik Beginn="X" Ende="Y"><FigN> Überschrift#(IMAGE)Unterschrift</Fig N> ... </Grafik>	Abbildung und Abbildungsgruppe (mit Über-, Unterschrift), die vereinfacht wiedergegeben werden sollen: Diese werden lediglich mit der Kennung #(IMAGE) markiert, die Überschrift vor der Kennung und die Unterschrift danach werden erfasst. Jede Figureinheit wird mit der paarigen Kennung <FigN> eingeschlossen, wobei die Figuren einer Figurengruppe jeweils— in der Reihenfolge oben -unten, links-rechts — durch nummeriert werden (N). Anfang und Ende einer Figurengruppe wird mit <Grafik> gekennzeichnet, die vertikale Position in Form von Zeilennr. mit dem optionalen Attribut Beginn="X" und Ende="Y" (falls die Zeilennr. ermittelbar).
Tabulatoren in WORD	Bitte bei der Konvertierung vorerst nicht löschen, sondern prüfen, ob nützlich
{p/q}, {p-q}, {i{p/q}} oder {i{p-q}} („durch“-Zeichen: Schrägstrich oder Bindestrich je nach Druckbild)	Brüche und gemischte Zahlen: i=ganze Zahl, p=Zähler, q=Nenner; z. B. {1/2} für ½, {1{1/2}} für 1½
{?V}	Allgemeine Zweifelskennung, V=Erfassungsversion A, B, C
{?Vtypogr}	Zweifelskennung für typographische Abbildungsschwierigkeiten. V = Erfassungsversion A, B, C
{?V-NOT:ci}	Ein oder mehrere Wörter in der Vorlage wegen schlechter Qualität nicht zu entziffern und deshalb nicht erfasst
{?V-NOT:hang}	Ein oder mehrere Zeilen in der Vorlage wegen schlechter Qualität nicht zu entziffern und deshalb nicht erfasst

#### Bemerkungen:

- 1) Groß- u. Kleinschreibung der Kodierungszeichen sind ohne Bedeutung; Nicht alle vorgesehenen Kodierungen müssen auch tatsächl. zum Einsatz gekommen sein.
- 2) Jede Textzeile der Vorlage wird in WORD mit RETURN (= „Absatz“) beendet und dadurch zeilengenau abgetippt. Die Absatztypen werden mit den oben vorgesehenen expliziten Kennungen abgebildet; die Absatzformatierungsfunktion von WORD wird also nicht verwendet.
- 3) Schriftauszeichnungen wie Fett, Kursiv, Kapitälchen, Hoch- u. Tiefstellung wurden in WORD mithilfe der Formatierungsfunktionen direkt formatiert (bitte beachten: Unterstreichung ist als Ersatz für Sperrung verwendet, s. o.) und können mit praktiziertem Verfahren nach TUSTEP korrekt konvertiert werden. Die Formatierungsfunktionen von WORD für rechtsbündige Ausrichtung oder Zentrierung ganzer Zeile(n) werden aber nur für die optische Kontrolle der Eingabe benutzt, vgl. oben. Ansonsten wurden für die Abbildung der typographischen Gestaltungsmerkmale keine Formatierungsfunktionen von WORD, sondern nur die in der obigen Tabelle aufgeführten expliziten Kodierungen verwendet.