

## Allgemeine Erfassungsanweisungen für Werke des Projekts *Die Schule von Salamanca*

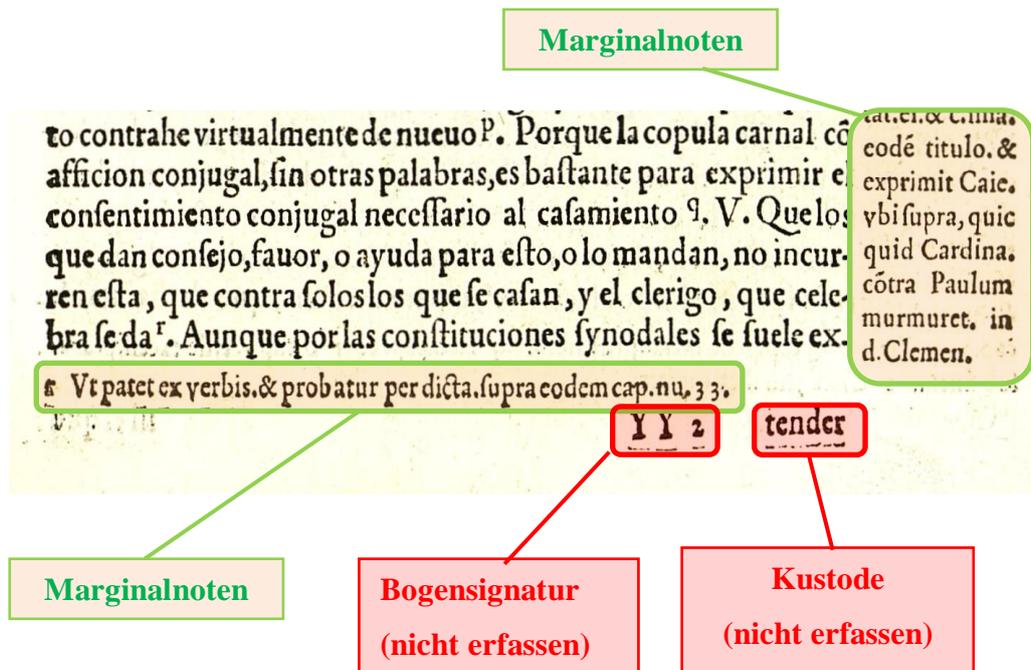
Stand: 31.01.2024

### I. Allgemeine Definitionen: Was (nicht) zu erfassen ist

#### Hauptbereich und Marginalbereich des Textes

Der 'Hauptbereich' ist der innerste (d.h. selbst nicht wieder einen anderen Textbereich umfließende), ggf. mehrspaltige Textblock; er wird komplett erfasst. Der 'Marginalbereich' ist der Bereich vom äußeren Seitenrand bis zum Haupttext. Er umfließt den Hauptbereich oben und unten sowie innen und außen und beinhaltet Seitenzahlen, Kustoden, Noten, Kolumnentitel und Marginalnoten. Davon werden nicht erfasst: Seiten-/Foliozahlen, Kopfzeilen, Kolumnentitel, Bogensignaturen, Kustoden. Erfasst werden die Marginalnoten, die in der Regel an den Seiten und unterhalb des Textes stehen.

Hier als generelles Beispiel ein Seitenende (nicht werkspezifisch):



## Handschriftliche Elemente

Handschriftliche Notizen, Anmerkungen, Symbole, Zeichnungen und Anstreichungen werden nicht erfasst. Für den Umgang mit (möglicherweise handschriftlich) durchgestrichenen Passagen siehe Abschnitt III.4.a „Nicht oder schwer lesbare Stellen“.

## II. Zeichenkodierung

### **Allgemeines**

Textzeichen werden grundsätzlich und so weit wie möglich als Unicode-Zeichen erfasst. Dies gilt auch für Digraphen und Zeichen mit diakritischen Marken (z.B. „æ“, „œ“, „é“, „ñ“, „ä“, „ç“) sowie für nicht-lateinische Zeichen (etwa griechische oder hebräische Zeichen, siehe hierzu auch Punkt III.3.m). Sind Zeichen nicht als Unicode-Zeichen darstellbar, so sollen sie in NCR-Notation (d.h. in der hexadezimalen Schreibweise von xml/html entities) ausgedrückt werden; wie (besondere) Zeichen erfasst werden, kann dabei generell der [Transkriptionsübersicht der ungewöhnlichen Zeichen](#) entnommen werden.

Kann ein Zeichen nicht eindeutig mithilfe der „Transkriptionsübersicht“ erfasst werden, so wird anstelle des entsprechenden Zeichens das leere TEI-Element `<g sameAs="..."/>` gesetzt, wobei der Wert von `@sameAs` auf einen entsprechenden Eintrag in einer externen Sonderzeichen-Dokumentationsdatei (XML) verweist, in welcher ggf. entsprechende Symbolgrafiken verlinkt sind. Sollte sich bereits vor oder im Laufe der Erfassung eines Textes abzeichnen, dass ein darin häufig benutztes Zeichen nicht in der „Transkriptionsübersicht“ vorhanden ist, soll das Projekt bezüglich der richtigen Kodierung und der Anpassung der „Transkriptionsübersicht“ kontaktiert werden.

### **Umgang mit besonderen Buchstaben und Zeichen**

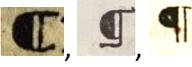
#### „S“ und Ligaturen

Rundes S („S“/„s“), Schaft-S („ſ“) und die f-s-Ligatur („ß“) werden als solche erfasst, unabhängig von benachbarten Zeichen. Ligaturen mit „s“ oder „ſ“ werden dementsprechend in aufgelöster Form erfasst (siehe auch folgenden Absatz).

Nur die Ligaturen „Æ“/„æ“ und „Œ“/„œ“ werden als solche erfasst. Andere Ligaturen (z.B. „ij“, „ft“, „st“, „fl“) werden in aufgelöster Form erfasst (z.B. „ij“, „ft“, „st“, „fl“).

Zur Kodierung von Ligaturen (und Zeichen generell) siehe auch die [Transkriptionsübersicht der ungewöhnlichen Zeichen](#)

#### Absatzzeichen

Zeichen wie , die einen (sinnhaften) Absatzbeginn markieren (auch innerhalb des typographischen Absatzes), sollen als „¶“ (Latin-1 Supplement Code Block, bzw. `&#x00B6`; in NCR-Notation) erfasst werden.

„I“ vs. „J“ und „V“ vs. „U“

„V“/„v“ und „I“/„i“ werden als solche erfasst (d.h. nicht etwa normalisiert als „U“/„u“ oder „J“/„j“).

#### Anführungsstriche

Anführungsstriche werden grundsätzlich nicht mittels (Unicode-)Zeichen wiedergegeben, sondern mithilfe des TEI-Elements <q> ausgedrückt und durch dieses ersetzt; siehe hierzu unbedingt III.3.h „Text in Anführungsstrichen“.

#### **Leerzeichen**

Die Verwendung von Leerzeichen (nach Satzzeichen) wird, soweit möglich, stillschweigend an moderne Gepflogenheiten angepasst, so dass die korrekte Zusammen- und Getrennschreibung, anders als im Druckbild der Vorlage, deutlich werden. Fehlende Leerzeichen nach Satzzeichen werden also beispielsweise ergänzt, während überzählige Leerzeichen (sofern der erhöhte Zwischenraum nicht bedeutungstragend ist) reduziert werden.

#### **Trennungszeichen und Gleichheitszeichen**

Trennstriche (als Markierungen von Worttrennungen) werden in normalisierter Form als moderner Bindestrich „-“ erfasst (siehe hierzu in jedem Fall Punkt III.3.a „Zeilenumbrüche (und Worttrennungen)“). Dies gilt etwa auch für das rechts schräg nach oben gestellte Gleichheitszeichen, das vor allem in Fraktuschriften vorkommt; normale Gleichheitszeichen (also nicht als Trennungsstrich benutzte G., die im Frakturdruck sehr lang sein können) sollen aber als „=“ erfasst werden.

#### **Fraktur- und Antiquaschrift**

Die Benutzung von Fraktur- oder Antiquaschrift, oder der Wechsel zwischen beiden Schriftsystemen, wird in den Texten nicht als solche(r) ausgezeichnet.

### III. Allgemeine Textauszeichnung

Die Textauszeichnung soll dem **TEI Tite**-Schema folgen (siehe [http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei\\_tite.doc.html](http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_tite.doc.html)), einem speziell für Transkriptionen spezifizierten Schema der *Text Encoding Initiative*. Das ‚Endprodukt‘ der Transkription des Werkes sollte somit in Form von **TEI Tite**-XML vorliegen. Da **TEI Tite** keinen XML-Header erlaubt (<text> ist das Wurzelement), sind bibliographische o.ä. Metadaten für die Transkription und Textauszeichnung irrelevant. Für die genauen Spezifikationen von **TEI Tite** kann auch dessen [Dokumentation](#) konsultiert werden. Für Fragen zur finalen Edition und zur allgemeinen Orientierung bei der Textauszeichnung können auch die [Editionsrichtlinien des SvSal-Projekts](#) hilfreich sein, die jedoch nicht als Erfassungsanweisungen (etwa im Hinblick auf Zeichenkodierung) anzusehen sind.

Auszuzeichnende Entitäten und die Arten der Auszeichnung werden im Folgenden genauer erläutert.

#### 1. Typografische Merkmale

Für typografische Besonderheiten stehen folgende Elemente zur Verfügung (das

Präfix "t:" gibt ggf. den TEI-Tite-Namensraum an, es handelt sich bei diesen Elementen also um Tite-Elemente):

- Fett: <t:b>...</ t:b>
- Kursiv: < t:i>...</ t:i>
- Tiefgestellt: < t:sub>...</ t:sub>
- Hochgestellt: < t:sup>...</ t:sup>
- Kapitälchen: < t:smcap>...</ t:smcap>
- Initialen: <hi rend="init">...</hi>; ist eine Initiale nicht erkennbar, so wird ein leeres <hi rend="init"/> gesetzt
- Sperrsatz (erhöhter Abstand zwischen den Buchstaben): <hi rend="sp">...</hi>
- Recte: innerhalb eines kursiven Textes recte ("gerade" bzw. nicht kursiv) gesetzter Text wird mit <hi rend="recte">...</hi> ausgezeichnet
- größere Lücken im Text innerhalb einer Zeile (nur im Hauptbereich des Textes): <seg rend="gap"/>

## 2. Textausrichtung und Schriftgröße

Textblöcke, die zentriert oder rechtsbündig stehen UND keine Überschriften oder Verstehtext sind, sollen mittels <hi rend="center"> bzw. <hi rend="right"> entsprechend ausgezeichnet werden. Textblöcke, die keiner Position (links, mittig, rechts) eindeutig zugeordnet werden können, werden der relativ am nächsten liegenden Position zugeordnet.

Um eine Trennung zwischen Textausrichtung/Typografie und „konzeptuellen“ Elementen (z.B. Paragraphen) herzustellen, soll die Textausrichtung stets mittels <hi>-Tags erfasst werden und nicht innerhalb konzeptueller Elemente: also etwa <p><hi rend="right">...</hi></p> (anstatt <p rend="right">...</p>).

Die Schriftgröße (bzw. Unterschiede in derselben im Laufe eines Texts) wird nicht erfasst (abgesehen von dem Sonderfall der Initialen, s.o.).

## 3. Strukturelle Elemente

### a. Zeilenbeginn (und Worttrennungen)

Der Beginn jeder neuen Zeile (im Haupt- und im Marginalbereich des Textes) wird mit <lb/> ausgezeichnet. Beginnen zugleich strukturelle bzw. konzeptuelle Annotationen (<p>; <p rend="hX">; <table>, <row>, <cell>; <foreign>; <list>, <head>, <item>, <ref>; <lg>, <l>; <note>; <q>; <title>, <titlePage>, <titlePart>; <unclear>), so soll <lb/> möglichst innerhalb des entsprechenden strukturellen Elements gesetzt werden. Im Falle eines Spaltenformats werden die Zeilenanfänge für jede Spalte ausgezeichnet. Auch im Falle unlesbarer, zeilenübergreifender Textstellen ist die Markierung von Zeilenanfängen per <lb/> durchzuführen, wenn diese erkennbar oder genau abschätzbar sind; im Falle von grafischen Elementen wie Illustrationen

oder Ornamenten wird jedoch keine Auszeichnung von (impliziten) Zeilenanfängen durchgeführt.

Geht mit dem Zeilenumbruch eine explizit, d.h. durch Trennungszeichen, markierte Worttrennung einher, so bekommt `<lb/>` ein Attribut `@type` mit dem Wert "nb" (für "no break"), also: `<lb type="nb"/>`; das Trennzeichen wird in normalisierter Form (d.h. als moderner Trennstrich „-“) direkt (!) vor dem `<lb type="nb"/>` miterfasst. Geht mit dem Zeilenbeginn bei einer (expliziten) Worttrennung zugleich ein Spalten- (`<cb/>`) und/oder Seitenbeginn (`<pb.../>`) einher, so wird dieser nur im Element für den Zeilenbeginn (`<lb.../>`) in obiger Form annotiert (und nicht etwa in `<cb/>` oder `<pb/>`). Hierbei (wie auch sonst) ist die richtige Reihenfolge der Elemente zu beachten, also: 1.) `<pb.../>`, 2.) `<cb/>`, 3.) `<lb .../>`.

#### **b. Spaltenbeginn und -format**

Der Beginn jeder einzelnen Spalte wird durch das Element `<cb/>` markiert, wobei eine eventuelle Nummerierung der Spalten im (nur in diesem Fall vorhandenen) `@n`-Attribut des jeweiligen `<cb/>` angegeben wird. In jedem Fall (außer in Marginalnoten, siehe III.3.f) bekommt `<cb/>` eine eindeutige ID in `@xml:id` (deren Wert frei gewählt werden kann). Beispiel:

```
<cb n="2" xml:id="W0002_cb123"/>
```

Bei der Änderung des Spaltenformats im Lauf des Textes (auch etwa für spaltenübergreifende Überschriften) wird ein `<t:colShift cols="..."/>` gesetzt, wobei `@cols` die Anzahl der Spalten pro Seite im folgenden Textabschnitt angibt (es hat also mindestens den Wert "1"). Bei mehrspaltigem Format im folgenden Text wird `<t:colShift .../>` direkt vor dem ersten `<cb/>` gesetzt.

#### **c. Seitenbeginn**

Der Beginn jeder Seite wird durch `<pb n="..." facs="..." xml:id="..."/>` markiert, wobei `@n` die Seitenzahl (falls es im Werk eine durchgängige Seitenpaginierung gibt) und `@facs` den Dateinamen des zugehörigen Faksimiles angibt. (Zu Seitenzahlen und deren Normalisierung siehe evtl. auch die besonderen Erfassungsanweisungen für das jeweilige Werk.) Zusätzlich (außer in Marginalnoten, siehe III.3.f) soll jedes `<pb/>` eine eindeutige ID in `@xml:id` haben, deren Wert sich z.B. an jenen von `@facs` anlehnen kann. Beispielsweise:

```
<pb n="41" facs="W0002-0060" xml:id="W0002-0060"/>
```

Nicht zu erfassende Seiten (etwa die leeren Seiten zu Beginn und Ende eines Drucks) werden jeweils als `<pb/>` mit leerem `@n` repräsentiert, also bspw. `<pb n="" facs="W0096-A-0002"/>`.

**Für die Positionierung von Umbruchselementen (<pb/>, <t:colShift/>, <cb/>, <lb/>, etc.) gilt allgemein folgende Regel:**

Fällt ein Umbruch mit dem Beginn „konzeptueller“ Textelemente (z.B. <head>, <p>, <div?>, <table>, <row>, <cell>; <foreign>; <list>, <head>, <item>, <ref>; <lg>, <l>; <note>; <q>; <title>, <titlePage>, <titlePart>; <unclear>) zusammen, dann soll das Umbruchselement (<pb/>, <t:colShift/>,<cb/>,<lb/>) als erstes Kindelement (child) des ersten Mixed-Content-Elements gesetzt werden. Beispiele:

<div1><head><pb/>... , nicht: <pb/><div><head>...

<table><row><cell><lb/>... , nicht: <table><row><lb/><cell>...

...</item><item><cb/>... , nicht: ...</item><cb/><item>...

<p><t:colShift cols="2"/><cb/><hi rend="init">...

Treten mehrere Umbruchs- und/oder andere leere Elemente gemeinsam auf, dann soll die folgende Reihenfolge eingehalten werden:

**<pb/><t:colShift/><cb/><lb/><milestone/>** (<figure/> und <t:ornament/>

sollen dabei entweder vor oder nach Seiten- bzw. Spaltenumbrüchen gesetzt werden, je nachdem, auf welcher Seite sie im Original vorkommen. Auf jeden Fall sollten sie in einer Sequenz mit anderen Umbruchselementen vorkommen, falls diese an der jeweiligen Stelle vorhanden sind.)

#### d. Überschriften

Überschriften *außerhalb des Hauptteils* eines Bandes (d.h. etwa solche in der Titelei oder im Anhang) werden grundsätzlich als <p rend="hx">...</p> erfasst.

Überschriften *im Hauptteil* werden grundsätzlich als <p rend="h?">...</p> erfasst, wobei das „?“ in @rend für eine Zahl steht, die die relative Überschriftengröße angibt – analog etwa zur Repräsentation von Überschriften in HTML mit <h1>, <h2> usw. Beispielsweise werden in einem Hauptteil Überschriften mit der größten Schrift als <p rend="h1"> annotiert, die nächstkleineren mit @rend="h2“. Kann keine klare Zuordnung getroffen werden, so wird der Attributwert "hx" vergeben, also <p rend="hx">...</p>.

Treten mehrere Überschriften (etwa die Überschrift für ein Buch und die für dessen erstes Unterkapitel) gemeinsam auf, so ist auf die getrennte Auszeichnung der Überschriften zu achten.

Überschriften in Listen werden innerhalb von <list> ausgezeichnet und als

<head>...</head> erfasst.

Überschriften auf Titelseiten werden mit den entsprechenden Tags erfasst (siehe hierzu Punkt j. „Titelseite(n)“).

Bei werkspezifischen Fragen, was in welcher Weise als Überschrift anzusehen ist, siehe ggf. auch die besonderen Erfassungsanweisungen für das jeweilige Werk.

**e. Paragraphen**

Die typographischen Absatz-Unterteilungen im Hauptbereich, markiert in der Quelle durch vertikale Abstände, Erstzeileneinzüge, verkürzte (also nicht bis in den Blocksatz ausgetriebene) Zeilenenden o.ä., werden als <p>-Paragraphen ausgezeichnet.

**f. Marginalnoten und Querverweise**

Falls vorhanden, werden Marginalnoten (Text im Marginalbereich, s.o.) per <note>-Element mit dem Attribut @type="margin" festgehalten. Zusätzlich erhält die Note ein @xml:id. Dabei sind unterschiedliche Fälle möglich: Die Note kann in der Vorlage durch ein Zeichen (z.B. einen hochgestellten Buchstaben oder ein Sternchen) identifiziert werden; dieses kann an der genauen Position, an der die Note verankert ist, stehen, am Beginn der Note selbst oder an beiden Positionen.

Ist die genaue Position, an der die Note im Hauptbereich des Textes verankert ist, ersichtlich (etwa ein hochgestellter Buchstabe), so wird die Note (vollständig) auch an genau dieser Stelle erfasst und unterbricht somit im TEI-XML den Haupttext. In diesem Fall wird als Wert von @n in <note> das Verankerungszeichen (z.B. ein Buchstabe) festgehalten; das Zeichen selbst (im Text) wird dann durch den <note>-Tag ersetzt (da es in note/@n "aufgelöst" wurde).

Ansonsten – falls es kein referenzierendes Zeichen im Hauptbereich gibt – wird die Note (vollständig) am Ende der Zeile erfasst, auf deren Höhe sie beginnt (oder, falls sie auf der Höhe zwischen zwei Zeilen steht, am Ende der ersten Zeile) und erhält das Attribut @rend="noRef". „Am Ende der Zeile“ heißt hierbei: direkt vor dem <lb/>, mit dem die folgende Zeile beginnt; das gilt auch für Fälle, bei denen mit dem Zeilenumbruch eine Worttrennung einher geht (<lb type="nb"/>, s.o.).

Wenn in der Note ein Seiten- oder Spaltenwechsel erfolgt, verweist dieser mit dem @sameAs-Attribut auf die xml:id des Seitenwechsels im Haupttext (<pb/> bzw. <cb/> wird in diesem Fall also sowohl im Haupttext als auch in der <note> gesetzt). Zum Beispiel (<pb> innerhalb von <note> verweist auf <pb> im Haupttext):

<pb sameAs="#W0066-0005"/>

(Achtung: im Falle von @sameAs sollte <pb> – anders als im Haupttext – nicht über @facs verfügen.)

Oder (<cb> innerhalb von <note> verweist auf <cb> im Haupttext):

<cb sameAs="#W0066-cb123"/>

Es ist zu beachten, dass <cb/> bzw. <pb/> hier, im Gegensatz zu <pb/> / <cb/> im Haupttext, kein Attribut @xml:id haben.

(Allgemeiner Hinweis: Wenn Text im Marginalbereich nur aus einer Ziffer oder einzelnen Symbolen besteht, handelt es sich um Absatz- oder Artikelnummern, die nicht als Marginalnoten, sondern mittels spezifischer Elemente (z.B. <milestone.../>) aufgenommen werden; siehe hierzu Punkt III.3.n und ggf. die besonderen Erfassungsanweisungen für das jeweilige Werk.)

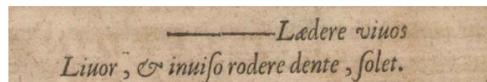
#### g. Verstext

Verstext (meist an Einrückung und Linksbündigkeit zu erkennen) wird insgesamt in einem <lg> (pro Verstext-Block, also z.B. pro Gedicht oder Verszitat) erfasst, mögliche Strophen (falls identifizierbar) werden als einzelne <lg> (unterhalb des Gesamt-<lg> für den Verstext-Block) und die einzelnen Verszeilen jeweils als <l> annotiert.

Dabei gilt grundsätzlich jede typografische Zeile auch als Verszeile, es sei denn, die folgende Zeile ist gegenüber der aktuellen Zeile eingerückt; in letzterem Fall handelt es sich um einen Zeilenumbruch (<lb/>) innerhalb der Verszeile, also z.B.:

<l>[erster Teil der Verszeile]<lb/>[zweiter Teil der Verszeile]</l>

Treten innerhalb (z.B. am Anfang) einer Verszeile durchgezogene Linien auf, wie in folgendem Beispiel, so werden diese durch <seg rend="line-in-verse"/> markiert und ersetzt:



<l><seg rend="line-in-verse"/>Laedere ...</l>

Die Textausrichtung ist bei Verstext nicht gesondert zu annotieren.

#### h. Text in Anführungsstrichen

Mit Anführungsstrichen markierter Text wird entweder mit <q>...</q> annotiert, falls Anfang und Ende des Textes eindeutig bestimmt werden können. Ist dies nicht der Fall, so werden am Ort der Anführungsstriche leere <q/> gesetzt. Im Original vorhandene Anführungsstriche entfallen dabei. In jedem Fall werden also öffnende Anführungsstriche durch <q> ersetzt (wenn es zugehörige schließende Anführungsstriche gibt); schließende Anführungsstriche werden durch </q> ersetzt (wenn es zugehörige öffnende

Anführungsstriche gibt); und alle anderen Anführungsstriche werden durch <q/> ersetzt.

**i. Ornamente, Grafiken und Illustrationen**

Ornamente (d.h. als strukturelle Marker fungierende grafische Elemente, wie z.B. langgezogene Striche) werden jeweils durch ein leeres <t:ornament/> erfasst. Alle weiteren grafischen Elemente werden nicht transkribiert, sondern lediglich jeweils durch ein leeres <figure/>-Element markiert. Das Setzen von (impliziten) Zeilenanfängen als <lb .../> entfällt dabei.

**j. Titelseite(n)**

Spezielle Textelemente der Titelseiten können mit den dafür in TEI Tite zur Verfügung stehenden Elementen (<titlePage>, <titlePart>, <byline>, <docAuthor> etc.) ausgezeichnet werden. Ansonsten können entsprechende Elemente der Einfachheit halber auch mit den "üblichen" Elementen (v.a. <p> und <p rend="hx">) ausgezeichnet werden.

**k. Grobstrukturierung**

Die Titelseite(n) eines Werkes wird/werden insgesamt mit <front>...</front> umschlossen; der gesamte darauf folgende Text (bis zum Ende des Buches) wird der Einfachheit halber insgesamt mit <body><div1>...</div1></body> umschlossen. (D.h. es darf im Werk genau ein <body>-Tag und ein <div1>-Tag, in direkter Aufeinanderfolge, geben. Die genauere Unterteilung (auch mittels <back> und strukturelle Tiefenannotation mittels <div> erfolgt in der späteren Redaktionsarbeit.)

**l. Mehrbändige Werke**

Bei mehrbändigen Werken wird jeder Band einzeln (d.h. wie eine Monographie) erfasst.

**m. Textstellen mit nicht-lateinischen Schriftzeichen**

Bei Textpassagen in anderen Sprachen als der Hauptsprache des Werkes werden diese Passagen mit einem <foreign xml:lang="...">...</foreign> markiert, wobei der Wert von @xml:lang eine entsprechende Abkürzung der Sprache ist. Dies ist grundsätzlich für Textpassagen in nicht-lateinischen Sprachen/Schriftsystemen wie Griechisch (<foreign xml:lang="grc">) und Hebräisch (<foreign xml:lang="heb">) durchzuführen. Dabei sollen Zeichen, die eindeutig solchen im entsprechenden Unicode-Block zugeordnet werden können, mittels der geeigneten Zeichen erfasst werden; falls die Zuordnung schwer fällt oder unmöglich ist, sollen Textstellen als <unclear reason="foreign"/> markiert werden, wobei ein <unclear/> für mehrere Zeichen bis zu einer ganzen Zeile stehen kann.

Der Unicodeblock für das Griechische ist "Greek and Coptic" (Codepoints

U+0370 bis U+03FF):

[https://de.wikipedia.org/wiki/Unicodeblock Griechisch und Koptisch](https://de.wikipedia.org/wiki/Unicodeblock_Griechisch_und_Koptisch)

Für Hebräisch ist es der Unicodeblock "Hebrew" (Codepoints 0590 bis 05FF):

[https://de.wikipedia.org/wiki/Unicodeblock Hebr%C3%A4isch](https://de.wikipedia.org/wiki/Unicodeblock_Hebr%C3%A4isch)

Bei Passagen in anderen Sprachen/Schriftsystemen als den hier genannten (sofern ohne weiteres identifizierbar) sollte mit dem Projekt Rücksprache gehalten werden.

Bei sehr häufigen und kleinteiligen fremdsprachlichen Textpassagen (Daumenregel: auf mehr als 15 Seiten jeweils durchschnittlich mehr als 10 Zeilen) innerhalb eines Werkes ist ebenfalls Rücksprache mit dem Projekt über die genaue Art der Annotation zu halten.

#### **n. Inhaltliche Markierungen im Text und Marginalziffern**

Inhaltliche Umbrüche sind oft durch Ziffern im Marginalbereich und/oder bestimmte Zeichen im Hauptbereich des Textes (z.B. Kreuzchen oder Asteriske) markiert. Diese Markierungen werden stets mittels des <milestone>-Elements erfasst, wobei folgende drei Fälle zu unterscheiden sind:

- 1.) Eine Marginalziffer referenziert eindeutig ein Zeichen (z.B. ein Kreuzchen) im Hauptbereich des Textes, indem sie auf derselben Höhe wie dieses (oder unweit davon) im Marginalbereich auftritt. In diesem Fall werden die Marginalziffer UND das markierende Zeichen im Hauptbereich in einem (leeren) <milestone>-Element zusammengefasst:  
<milestone n="[Marginalziffer]" unit="x" rend="[Zeichen]"/>  
Erlaubte Werte für das [Zeichen] in @rend sind: "dagger", "asterisk".  
Der <milestone>-Tag soll genau an der Stelle des markierenden Zeichens im Text gesetzt werden und das markierende Zeichen ersetzen, wobei etwaige Leerzeichen um das Zeichen erhalten bleiben.
- 2.) Eine Marginalziffer kann nicht eindeutig einem referenzierten Zeichen im Hauptbereich zugeordnet werden. In diesem Fall wird ein Tag <milestone n="[Marginalziffer]" unit="x"/> am Ende der Zeile, auf deren Höhe die Marginalnote auftritt (oder am Ende der ersten, falls zwischen zwei Zeilen), gesetzt.
- 3.) Ein den Umbruch markierendes Zeichen im Hauptbereich, z.B. ein Kreuzchen oder ein Asterisk, kommt ohne Marginalziffer vor oder kann keiner Ziffer eindeutig zugewiesen werden. In diesem Fall wird ein Tag <milestone unit="x" rend="[Zeichen]"/> genau an der Stelle des Zeichens im Text gesetzt und ersetzt dieses, wobei etwaige Leerzeichen um das Zeichen erhalten bleiben. Erlaubte Werte für das [Zeichen] in @rend sind: "dagger", "asterisk".

#### o. Tabellen

Tabellen bzw. tabellenartige Textformatierungen werden als <table> (pro Tabelle) annotiert. Innerhalb einer Tabelle werden die Texteinheiten zeilenweise erfasst: als mehrere "Zellen" (<cell>) pro Zeile (<row>). Spaltentitel können mithilfe des Attributs @role="label" im betreffenden <cell> markiert werden.

```
<table>
  <row>
    <cell role="label">Titel Spalte 1</cell>
    <cell role="label">Titel Spalte 2</cell>
  </row>
  <row>
    <cell>Spalte 1, Zeile 1</cell>
    <cell>Spalte 2, Zeile 1</cell>
  </row>
  <row>
    ...
  </row>
</table>
```

Bei unregelmäßigen Formaten, in denen einzelne Zellen oder Zeilen sich über mehrere Zeilen und/oder Spalten erstrecken, kann die Anzahl der Zeilen/Spalten, über die sich eine Zelle/Zeile erstreckt, mittels der Attribute @cols und @rows angegeben werden (siehe [http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei\\_tite.doc.html#tablelists](http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_tite.doc.html#tablelists)).

#### 4. Umgang mit nicht oder schwer lesbaren Stellen sowie mit Zweifelsfällen

##### a. Nicht oder schwer lesbare Stellen

Bei völligem Textverlust – unabhängig von der Ursache – wird der genaue Ort bzw. Bereich des Verlustes mit einem (möglicherweise leeren) <unclear>-Element markiert. Alle gesetzten <unclear>-Tags sollen grundsätzlich eine Verantwortlichkeitsangabe mittels @resp mit dem Wert "#TL" erhalten, also: <unclear resp="#TL" ...>...

Bei zeilen-, spalten- oder seitenübergreifendem Textverlust muss die Strukturauszeichnung unbedingt beibehalten werden, indem innerhalb von <unclear> die entsprechenden Tags gesetzt werden. Ein Textverlust von genau drei Zeilen wäre etwa folgendermaßen auszuzeichnen: <unclear resp="#TL"><lb/><lb/><lb/></unclear>

Ist der Text noch erkennbar, aber nicht eindeutig transkribierbar, so wird er ebenfalls mit einem <unclear>-Element ausgezeichnet, wobei innerhalb des

Elements der transkribierte Text steht. Auch in diesem Fall ist die Zeilenzählung per `<lb/>` selbstverständlich weiterzuführen.

**b. Zweifelsfälle**

Ergeben sich bei der Transkription bzw. Textauszeichnung Zweifelsfälle, etwa wenn ein Textphänomen nicht durch die hier beschriebenen Tags erfasst werden kann oder sich Fragen hinsichtlich der „richtigen“ Annotation ergeben, so sind sie mit einem leeren `<gap resp="#TL"/>` (auch wenn semantisch nicht unbedingt ganz passend) als Hinweis für die spätere Redaktionsarbeit zu markieren. Bei durchgängig oder häufig in einem Werk auftauchenden Zweifelsfällen ist das Projekt hinsichtlich der richtigen Annotationsweise zu kontaktieren.