



Strukturierung eines TEI-Dokuments (TEI Kapitel 4)

Ulrike Henny
ulrike.henny@uni-koeln.de



Strukturierung eines TEI-Dokuments

- Worum geht es?
- Aufbau TEI-Dokument
- Struktur des Textes
 - Einzeltext, gruppierter Text, „schwebender Text“
 - Front Matter, Body, Back Matter
- Textabschnitte
 - <div>, <head>, <p>, ...
 - Titelseiten
 - Opener, Closer
- Elementübersicht
- Übungen



Worum geht es?

- Allgemeiner Aufbau eines TEI-Dokuments
- Struktur des Textes, der repräsentiert wird
 - Was sehe ich?
 - Wie verstehe ich es?
 - verschiedene Textsorten
 - Dokumente - Werke

Aufbau TEI-Dokument: *Default*

- Ein TEI-Dokument besteht aus...
 - Metadaten, die das Dokument beschreiben, das codiert wird
 - in <teiHeader>
 - dem repräsentierten Text selbst
 - in <text>

- Aufbau:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">  
  <teiHeader>  
  </teiHeader>  
  (<facsimile></facsimile>)  
  <text>  
  </text>  
</TEI>
```



Aufbau TEI-Dokument: *Korpus*

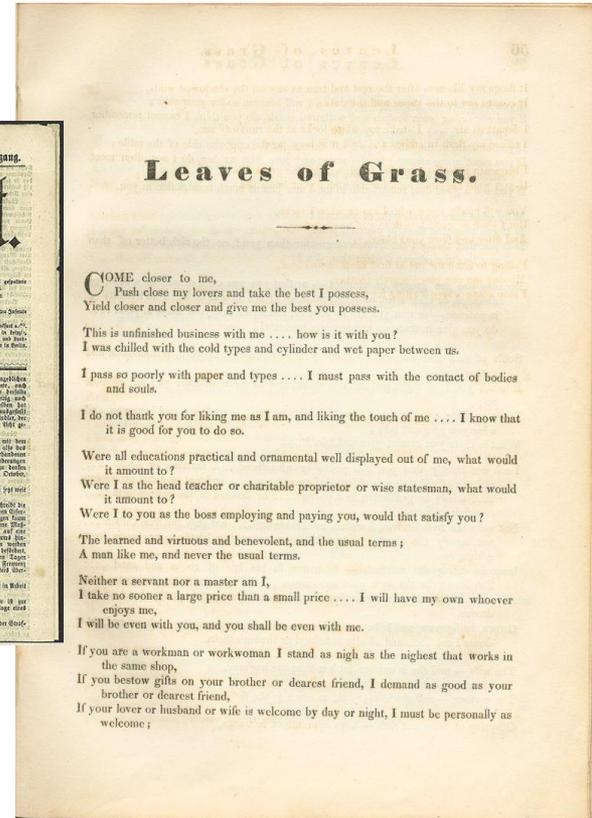
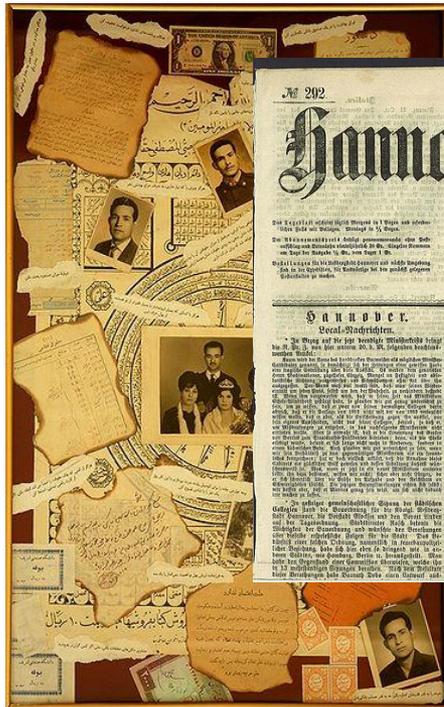
- Variante `teiCorpus`: mehrere vollständige TEI-Elemente mit eigenem Header
- Aufbau:

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">  
  <teiHeader></teiHeader>  
  <TEI>  
    <teiHeader></teiHeader>  
    <text></text>  
  </TEI>  
  <TEI>  
    <teiHeader></teiHeader>  
    <text></text>  
  </TEI>  
</teiCorpus>
```



Struktur des Textes

- Einzeltext – gruppierter Text – „schwebender Text“



Struktur des Textes: **gruppiertes Text**

- **<group>**
 - Enthält einen zusammengesetzten Text, also mehrere aufeinanderfolgende <text>-Elemente in der Gruppe
 - Die Texte werden zwar als eigenständig verstanden, bilden aber zusammen eine übergeordnete Einheit, z.B. bei Anthologien, einem Essayband, etc.
 - <group> selbst steht innerhalb von <text>
 - Schachtelung von Textgruppen möglich: <group> kann selbst wieder <group> enthalten
- Beispiel:
- **<text> ...*The adventures of Sherlock Holmes*...**
 - <group>**
 - <text>... *Adventure I* ...</text>**
 - <text> ... *Adventure II* ...</text>**
 - </group>**
 - </text>**



Struktur des Textes: „schwebender Text“

- **<floatingText>**
 - Ein bestimmter Text (jeglicher Sorte), der einen anderen Text unterbricht, der ihn umgibt
 - Kann selbst ein Einzeltext oder eine Gruppe von Texten sein
 - Kann überall vorkommen (auf allen Ebenen, beliebig tief)
 - „schwebender Text“ oder Zitat (<quote>)?
 - Ein schwebender Text hat nicht unbedingt eine externe Quelle
 - In <floatingText> kann wieder eine komplexe Textstruktur codiert werden, in <quote> nicht
- **Beispiel:**
 - **<p>**The Gentleman having finish'd his Story [...] saw lying there an old dirty rumpled Book, and found in it the following story: **</p>**
 - **<floatingText>**
 - **<body>**
 - **<p>**IN the time of the Holy War...



Struktur des Textes: Front Matter, Body, Back Matter

- Front Matter, Back Matter
 - Vor- und Nachtexte zum Text in <body>, „Paratexte“
- <TEI xmlns="http://www.tei-c.org/ns/1.0">
- <teiHeader>....</teiHeader>
- <text>
- <front>
- <body>
- <back>
- </text>
- </TEI>



Struktur des Textes: **Front Matter**

- Front Matter
 - Titelseite, Vorwort/Prolog, Widmung, ...
 - Funktionen: Identifikation, Einleitung, Aufforderung, Dank ...
 - NICHT der TEI-Header!
 - Hier wird das vorliegende Dokument transkribiert
 - Im Header dagegen werden (zusätzliche) Metadaten zur TEI-Datei notiert
- Elemente:
 - `<titlePage>`
 - `<div>`
 - `@type:`
 - `preface` | `ack` | `dedication` | `abstract` | `contents` | `frontispiece`
 - `Incipit` | `prayer`



Struktur des Textes: Front Matter

Beispiel:

```
□ <text>
  <front>
    <titlePage>
      <docTitle>
        <titlePart>Leaves<lb /> of<lb /> Grass.</titlePart><lb />
      </docTitle>
      <docImprint>Brooklyn, New York:<lb />1855.</docImprint>
    </titlePage>
  </front>
  <body>...</body>
</text>
```

Struktur des Textes: **Back Matter**

- Back Matter
 - Auch hier möglich: Titelseite, Inhaltsverzeichnis, ...
 - Anhang, Glossar, Register, Anmerkungen, etc.
 - Wird als nicht mehr zum eigentlichen Text gehörig angesehen
 - Dieselben Elemente wie in <front>

- Elemente:
 - <titlePage>
 - <div>
 - @type:
 - appendix | glossary | notes | bibliogr | index | colophon

Struktur des Textes: **Back Matter**

Beispiel:

```
□ <text>
  <body>...</body>
  <back>
    <div type="index">
      <head>Ortsregister</head>
      <list><head>A</head>
        <item>Aachen</item>
        <item>Ahlen</item>
        ...
      </list>
    </div>
  </back>
</text>
```



Textabschnitte

- Strukturelle Elemente auf einer tieferen Ebene
 - <div>, <head>, <p>, ...
 - Titelseiten
 - Opener, Closer
- flach oder hierarchisch organisiert
- Vielfalt
 - „Buch“, „Teil“, „Abschnitt“, „Kapitel“, „Brief“
 - „Akt“, „Szene“
 - „Eintrag“
 - „Faszikel“, „Relatio“, „Observatio“



Textabschnitte: `<div>`, `<head>`, `<p>`, ...

- `<div>`
 - „division“
 - Neutrales Element für Textabschnitte
 - Kann beliebig tief geschachtelt werden
 - Kann mit `@type` und `@subtype` näher bestimmt werden
 - `@n` für Nummerierung
 - `@xml:id` für Identifikation
 - Alternativ (aber nicht mehr gebräuchlich) :

`<div1>`

`<div2>`

...

`<div7>`

Textabschnitte: <div>, <head>, <p>,...

HAB Wolfenbüttel: Vorschläge für @type für verschiedene Gattungen

<http://diglib.hab.de/rules/documentation/structuralMD.xml>

• Verfügbare Wortlisten und Thesauri

Gattung/Typ	Belegung von term/@type	Präfix in div/@target	Thesaurus/Wortliste	Beispiel (ein Link auf die jeweilige tei-struct.xml befindet sich jeweils am Ende der Strukturdatenanzeige)
Strukturdaten allgemein	structure	--	structMD.xml	Kircher, Phonurgia Nova
Festkultur [Projektseite]	festival	festival_	festivalMD.xml	Weckherlin
Archäologische Funde [Projektseite]	archaeological_finds	archfunde_	archfundeMD.xml	Chassanion, Bericht Von den alten Riesen
Emblembuch [Projektseite]	emblem	emblem_	emblemMD.xml	De Bry, Proscenium vitae humanae
Kalender [Datenbank]	calendar	calendar_	calendarMD.xml	Gesprächskalender
Seuchenschriften [Datenbank]	contagion	contagion_	contagionMD.xml	Bericht von der grassierdenden Pestilenz
Flugblatt [Datenbank]	broadsheet	broadsheet_	broadsheetMD.xml	Beschreibung eines Wunder-Menschen
Gebetbuch/Gesangbuch [Datenbank]	prayer	prayer_	prayerMD.xml	Cüstrinsches Vollständiges Gebeth- Und Gesang Buch
Schneyer Codierungen für Predigten	schneyer	schneyer_	schneyerMD.xml	Sermones ..de sanctis



Textabschnitte: `<div>`, `<head>`, `<p>`, ...

- `<head>`
 - „heading“, Überschrift
 - Kann einmal oder mehrfach am Anfang eines Abschnitts vorkommen
 - `@type`, `@subtype`

- `<p>`
 - „paragraph“, Absatz
 - `@type`, `@subtype`

- `<trailer>`
 - Etwas Überschriftenähnliches am Ende eines Abschnitts



Textabschnitte: `<div>`, `<head>`, `<p>`, ...

□ Beispiel:

```
<text>
```

```
  <body>
```

```
    <div type="book" n="1">
```

```
      <head>Buch 1</head>
```

```
        <div type="chapter" n="1">
```

```
          <head>Kapitel 1</head>
```

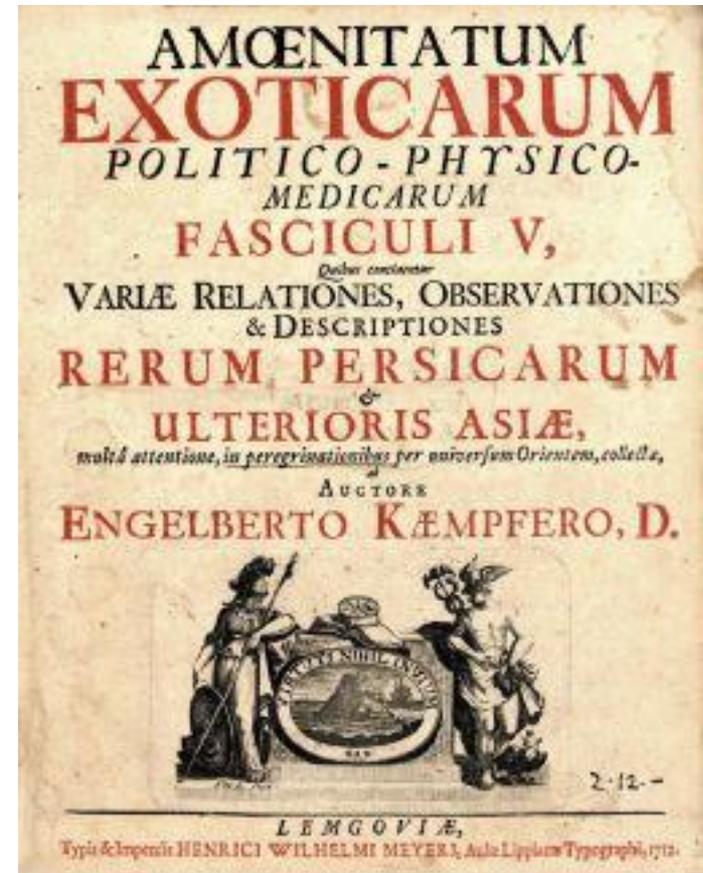
```
            <p>erster Absatz ...</p>
```

```
            <p>zweiter Absatz ...</p>
```

```
          </div> ...
```

Textabschnitte: Titelseiten

- Im Front oder Back Matter
- `<titlePage>`
 - `<docTitle>`
 - Titel des Dokuments mit allen Bestandteilen
 - `<titlePart>`
 - Teil des Werktitels
 - `@type`: main, sub, alt, short, desc
 - `<argument>`
 - Liste oder Beschreibung von Themen
 - `<byline>`
 - Zeile zur Verantwortlichkeit für das Werk
 - `<docAuthor>`
 - Autor des Dokuments/Werkes



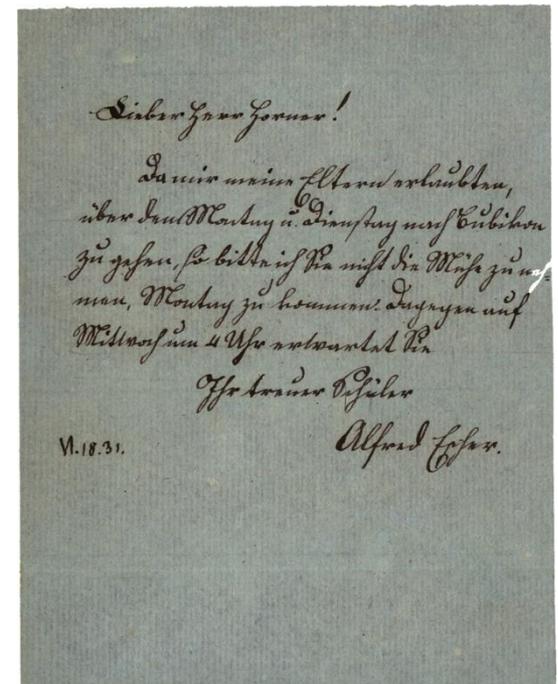


Textabschnitte: Titelseiten

- `<epigraph>`
 - Zitat
- `<imprimatur>`
 - Druckerlaubnis
- `<docEdition>`
 - Angabe zur Edition/Ausgabe/Auflage des Werkes
- `<docImprint>`
 - Druckvermerk (Zeit, Ort, Verleger)
- `<docDate>`
 - Datum des Dokuments/Werkes
- `<graphic>`
 - Ornament, Illustration, Bild

Textabschnitte: Opener & Closer

- Mehr oder weniger formelhafte Anfänge und Schlüsse
- z.B. bei Briefen
- Gruppiert in
 - <opener>
 - <closer>
- Elemente in <opener> und <closer>:
 - <byline>
 - Zeile zur Verantwortlichkeit für das Werk
 - <dateline>
 - Zeile zu Entstehungsort, -datum und -zeit des Textes





Textabschnitte: Opener & Closer

- Elemente in <opener> und <closer>
 - <salute>
 - Gruß, z.B. vor einem Vorwort oder einer Widmung oder am Ende eines Briefes
 - <signed>
 - Signatur am Ende eines Vorworts, Widmung, etc.
 - <postscript>
 - Passage nach der Signatur eines Briefes (seltener bei anderen Textsorten)
 - Oft als PS oder P.S. abgekürzt

Textabschnitte: Opener & Closer

- Beispiel:

```
<div type="letter" n="14">  
  <head>Letter XIV: Miss Clarissa Harlowe to Miss Bowe</head>  
  <opener>  
    <dateline>Thursday evening, March 2</dateline>  
  </opener>  
  <p>...</p>  
  <closer>  
    <salute>Yours more than my own, </salute>  
    <signed>Clarissa Harlowe</signed>  
  </closer>  
</div>
```



Elementübersicht

- TEI
- front, body, back
- text
- group, floatingText
- div
 - (div1, div2, div3, div4, div5, div6, div7)
 - head, p
- titlePage
 - docTitle, titlePart, argument, byline, docAuthor, epigraph, imprimatur, docEdition, docImprint, docDate, graphic
- opener, closer
 - dateline, salute, signed, postscript, trailer



Übungen

- Kopieren Sie die Text-Transkriptionen der folgenden Texte in eine neue TEI-Datei:
 1. Weber-Brief
 2. Vorlesungsverzeichnis Universität zu Köln, Sommersemester 1966

- Zeichnen Sie die Struktur der Texte mit den hier vorgestellten Elementen und Attributen aus.

Referenzen / Links

- Kapitel 4 der TEI P5 Guidelines: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>
- James Cummings, Self Study (part 3): The TEI Default Text Structure, <http://blogs.it.ox.ac.uk/jamesc/2013/01/31/self-study-part-3-the-tei-default-text-structure/>
- Thomas Stäcker, Torsten Schaßan, HAB-WDB Dokumentation 2007, <http://diglib.hab.de/rules/documentation/structuralMD.xml> (Tabelle Wortlisten und Thesauri für Strukturdaten)