

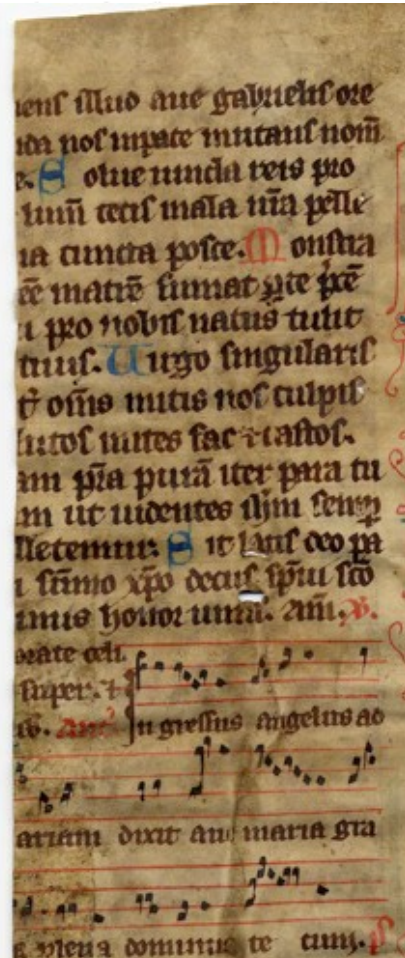
<?xml version="1.0" encoding="UTF-8" ?>

<ELEMENT docTitle ((figure | add | anchor | appSpan | addSpan | delSpan | gap | alt | altGrp | index | join | joinGrp | link | linkGrp | timeline | cb | lb | milestone | pb)*, (titlePart, (figure | anchor | appSpan | addSpan | delSpan | gap | alt | altGrp | index | join | joinGrp | link | linkGrp | timeline | cb | pb | milestone | pb)*)+) >

Source and editions

http://www.w3.org/1999/

<text>
<front>
<pb/>
<pb ed="HU" n="1" />
<lb/>
<fw type="page" p="1" />
<titlePage type="main">
<docTitle>
<lb/>
<titlePart type="main">
<lb/>
<titlePart type="de">
</docTitle>
<byline>
<lb/>
<af />
<lb/>
<docAuthor>Henricus
</byline>
<lb/>
<figure type="bar" />
<lb/>
<docDate>1865.
</titlePage>



sl:template match="//foreign
<xsl:choose>
<xsl:when test="contains
<xsl:element name="S
<xsl:attribute name="S
<xsl:apply-template
</xsl:element>
<xsl:text> <
</xsl:when>
<xsl:otherwise>
<xsl:element name="S
<xsl:attribute name="S
<xsl:apply-template
</xsl:element>
</xsl:otherwise>
</xsl:choose>
</xsl:template>

1 ?sum?ens illud aue gabrelis ore
?fun?da nos in pace mutans nomen
?eu?e. Solue uincla reis pro
4 ?fer? lumen cecis mala nostra pelle
?bon?a cuncta posce. Monstra
?te? esse matrem sumat per te precem
7 ?qu?i pro nobis natus tulit
?esse? tuus. Uirgo singularis
?in?ter omnes mitis nos culpis
10 ?so?lutos mites fac et castos.
?Uit?am presta puram iter para tu
?tu?m ut uidentes ihesum semper
13 ?co?lletemur. Sit laus deo pa
?tr?i summo christo decus spiritui sancto
imis honor unus. amen.
uersus.
16 ?R?orate celi
?de?super, et
?nu?b.
antiphona Ingressus angelus ad
19
?m?ariam dixit aue maria gra
22
?ti?a plena dominus te cum. psalmus
Magnificat.
25 oratio
<xsl:apply-templates select="//text />
</fo:flow>
</fo:page-sequence>

ra') or contains(\$style, '_a')">
]>
</xsl:text>
</xsl:template>
ter">
mt/title"/>
ter">Page : <fo:page-number/>
t" leader-length="33%"/>

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

What should a document format be like?

= "http://www.w3.org/1999/XSL/Transform" version="1.0">



- Open and well documented.
- Application independent.
- Interchangeable.
- Readable for both computers and humans
- Encoding structure, not layout.
- Make the encoding of the document explicit.

Text encoding languages

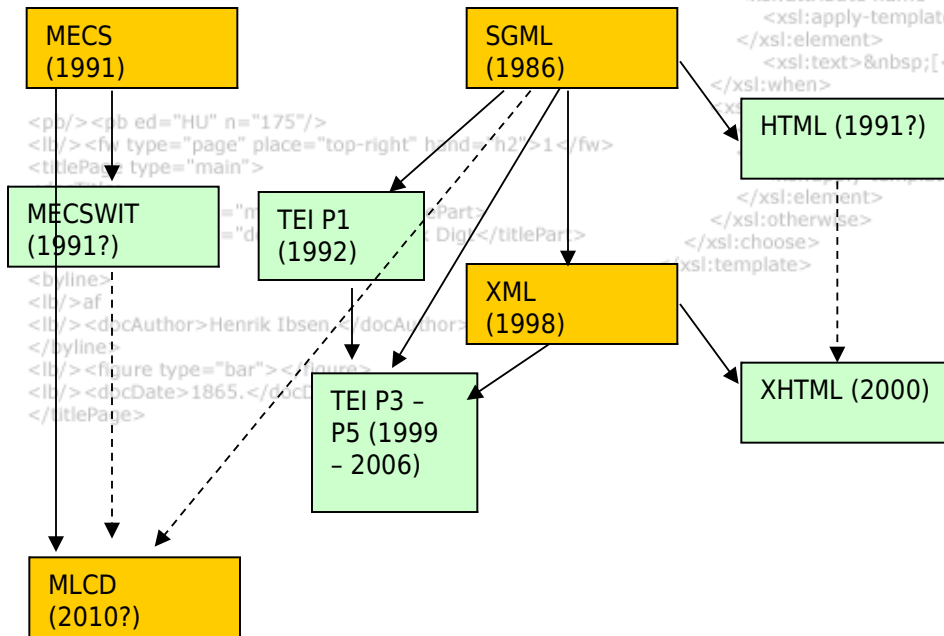
```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| altGrp | index | join | joinGrp | link | linkGrp | timeline
| cb | lb | milestone | pb)*)+) >
```

```
= "http://www.w3.org/1999/XSL/Transform" version="1.0">
```

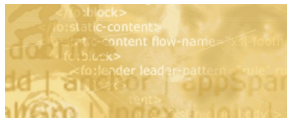
```
<xsl:template match="//foreign">
<xsl:choose>
<xsl:when test="contains($style, '_mn') or contains($style, '_a')">
<xsl:element name="SPAN">
<xsl:attribute name="style">font-style: italic;</xsl:attribute>
<xsl:apply-templates />
</xsl:element>
<xsl:when test="contains($style, '_i') or contains($style, '_a')">
<xsl:text>&nbsp;</xsl:text>
<xsl:when test="contains($style, '_u') or contains($style, '_a')">
<xsl:element name="SPAN">
<xsl:attribute name="style">font-style: italic;</xsl:attribute>
<xsl:apply-templates />
</xsl:element>
<xsl:otherwise>
<xsl:choose>
<xsl:template>
<xsl:template>
```

- Encoding grammars: SGML, MECS, XML
- Encoding semantics: HTML, HNML, MECSWIT, TEI Guidelines



```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```

```
fo:page-sequence master-reference="basic">
<fo:static-content flow-name="xsl-region-before">
<fo:block background-color="#D3D3D3" text-align="center">
<xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/title"/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-region-after">
<fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-footnote-separator">
<fo:block>
<fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
</fo:block>
</fo:static-content>
<fo:flow flow-name="xsl-region-body">
<xsl:apply-templates select="//text"/>
</fo:flow>
</fo:page-sequence>
```



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

```
<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| cb | lb | milestone | pb)*))
```

Extensible Markup Language (XML)

```
= "http://www.w3.org/1999/XSL/Transform" version="1.0">
```

Design goals:

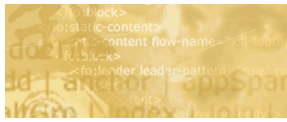
1. XML shall be straightforwardly usable over the Internet.
2. XML shall support a wide variety of applications.
3. XML shall be compatible with SGML.
4. It shall be easy to write programs which process XML documents.
5. The number of optional features in XML is to be kept to the absolute minimum, ideally zero.
6. XML documents should be human-legible and reasonably clear.
7. The XML design should be prepared quickly.
8. The design of XML shall be formal and concise.
9. XML documents shall be easy to create.
10. Terseness in XML markup is of minimal importance.

Source: *Extensible Markup Language (XML) 1.0 (Third Edition)*, W3C Recommendation 04 February 2004.

```
<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="contains($style, '_mn') or contains($style, '_n') or contains($style, '_ra') or contains($style, '_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style" value="{ $style }" />
        <xsl:apply-templates />
      </xsl:element>
    </xsl:when>
    <xsl:otherwise>
      <xsl:element name="SPAN">
        <xsl:attribute name="style">font-style: italic</xsl:attribute>
        <xsl:apply-templates />
      </xsl:element>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>
```

```
<text>
<front>
<pb/>
<lb/>
<titlePage type="main">
<docTitle>
<lb/>
<titlePart type="main">Brand.
</titlePart>
<lb/>
<titlePart type="desc">Et dramatisk Digt.
</titlePart>
</docTitle>
<byline>
<lb/>af
</byline>
<lb/>
<docAuthor>Henrik Ibsen.
</docAuthor>
</byline>
<lb/>
<docDate>1865.
</docDate>
</titlePage>
```

```
<fo:static-content flow-name="xsl-region-before">
<fo:block background-color="#D3D3D3" text-align="center">
  <xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/title"/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-region-after">
<fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-footnote-separator">
<fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
</fo:leader>
</fo:static-content>
<fo:flow flow-name="xsl-region-body">
  <xsl:apply-templates select="//text"/>
</fo:flow>
</fo:page-sequence>
```



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

```
<!ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| altGrp | index | join | joinGrp | link | linkGrp | timeline
| pb | lb | milestone | pb)*)+) >
```

DTD and Schema

```
= "http://www.w3.org/1999/XSL/Transform" version="1.0">
```

A template for the structure of a text, two main types:

- DTD - Document Type Definition

```
<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="contains($style, '_ra') or contains($style, '_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style">font-style: italic</xsl:attribute>
        <xsl:apply-templates />
      </xsl:element>
    <xsl:otherwise>
      <xsl:text>&nbsp;</xsl:text> <xsl:value-of select="attribute::lang" /> <xsl:text>]</xsl:text>
    </xsl:when>
    <xsl:otherwise>
      <xsl:element name="SPAN">
        <xsl:attribute name="style">font-style: italic</xsl:attribute>
        <xsl:apply-templates />
      </xsl:element>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>
```

- Schema:

- RELAX NG

- W3C Schema

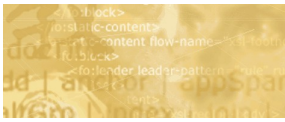
- Schematron

```
<text>
```

```
<front>
</front>
<pb/>
<lb/>
<fw type="right" hand="h2">1</fw>
<titlePage type="main">
<docTitle>
<lb/>
<titlePart type="main" hand="h1">
<lb/>
<titlePart type="de" hand="h2">
</docTitle>
<byline>
<lb/>
<docAuthor>Henrik Ibsen.</docAuthor>
</byline>
<lb/>
<figure type="bar" hand="h2">
<lb/>
<docDate>1865.</docDate>
</titlePage>
```

```
<fo:page-sequence master-reference="basic">
  <fo:static-content flow-name="xsl-region-before">
    <fo:block background-color="#D3D3D3" text-align="center">
      <xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/titl"/>
    </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-region-after">
    <fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
  </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-footnote-separator">
    <fo:block>
      <fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
    </fo:block>
  </fo:static-content>
  <fo:flow flow-name="xsl-region-body">
    <xsl:apply-templates select="//text"/>
  </fo:flow>
</fo:page-sequence>
```

```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```



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

```
<!ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| anchor | appSpan | addSpan | delSpan | gap | alt
| link | linkGrp | index | join | joinGrp | link | linkGrp | timeline
| cb | lb | milestone | pb)*)+ >
```

Well formed and Valid

```
= "http://www.w3.org/1999/XSL/Transform" version="1.0">
```

- A document has to be well formed to be called a XML document:

- Has only one root element.

- Has no open tags.

- All tags nest.

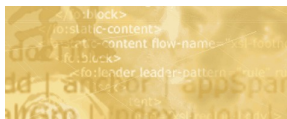
- A document which follows the rules of a schema is said to be valid.

```
<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="not($lang) or ($lang = 'en') or contains($style, '_ra') or contains($style, '_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style">font-style: italic</xsl:attribute>
        <xsl:apply-templates />
      </xsl:element>
    <xsl:otherwise>
      <xsl:text>&nbsp;</xsl:text> <xsl:value-of select="attribute::lang" /> <xsl:text></xsl:text>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>
```

```
<text>
<front>
<pb/> <pb ed="HU" n="175" />
<lb/> <fw type="page" place="top-right" hand="h2">1 </fw>
<titlePage type="main">
<docTitle>
<lb/> <titlePart type="main">Brand. </titlePart>
<lb/> <titlePart type="de" />
</docTitle>
<byline>
<lb/> af
<lb/> <docAuthor>Henrik Ibsen. </docAuthor>
</byline>
<lb/> <figure type="text" />
<lb/> <docDate>1866. </docDate>
</titlePage>
```

```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```

```
<fo:page-sequence master="page-sequence" basic-
font-family="serif" font-size="10pt" font-weight="normal"
font-style="normal" color="black">
  <fo:block background-color="#D3D3D3" text-align="center">
    <xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/titl"/>
  </fo:block>
  <fo:static-content>
    <fo:static-content flow-name="xsl-region-after">
      <fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
    </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-footnote-separator">
    <fo:block>
      <fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
    </fo:block>
  </fo:static-content>
  <fo:flow flow-name="xsl-region-body">
    <xsl:apply-templates select="//text"/>
  </fo:flow>
</fo:page-sequence>
```

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

```
<!ELEMENT docTitle  
((figure | add | anchor | appSpan | addSpan | delSpan |  
gap | alt | altGrp | index | join | joinGrp | link | linkGrp  
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure  
| add | anchor | appSpan | addSpan | delSpan | gap | alt  
| altGrp | index | join | joinGrp | link | linkGrp | timeline  
| cb | lb | milestone | pb)*)+) >
```

TEI versions

```
= "http://www.w3.org/1999/XSL/Transform" version="1.0">
```

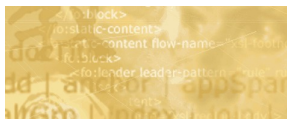
- P1 (1992)
- P3 (1999)
- P4 (2003)
- P5 (2006, and still in development)

```
<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="contains($style, 'italic') or contains($style, '_ra') or contains($style, '_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style">
          <xsl:apply-templates />
        </xsl:attribute>
        <xsl:text>&nbsp;</xsl:text>
      </xsl:when>
      <xsl:otherwise>
        <xsl:element name="SPAN">
          <xsl:attribute name="style">font-style: italic</xsl:attribute>
          <xsl:apply-templates />
        </xsl:element>
      </xsl:otherwise>
    </xsl:choose>
  </xsl:template>
```

- Do not be afraid of new versions of the TEI, old texts will still be valid according to the old version.
- New TEI versions are a new set of “spelling rules” which only applies to texts that are referring to them.

```
<fo:page-sequence number="1">
  <fo:static-content flow-name="xsl-region-before">
    <fo:block background-color="#D3D3D3" text-align="center">
      <xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/ttitle"/>
    </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-region-after">
    <fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
    </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-footnote-separator">
    <fo:block>
      <fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
    </fo:block>
  </fo:static-content>
  <fo:flow flow-name="xsl-region-body">
    <xsl:apply-templates select="//text"/>
  </fo:flow>
</fo:page-sequence>
```

```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```

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

TEI on Speech and Corpora

"http://www.w3.org/1999/XSL/Transform" version="1.0">

```
<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| link | linkGrp | timeline
| cb | lb | milestone | pb)*)+) >
```

11 Transcriptions of Speech

10.1 General Considerations and Overview

10.1.1 Divisions

10.2 Elements Unique to Spoken Texts

10.2.1 Utterances

10.2.2 Pause

10.2.3 Vocal, Kinesic, Event

10.2.4 Writing

10.2.5 Temporal Information

10.2.6 Shifts

10.2.7 Formal Definition

10.3 Elements Defined Elsewhere

10.3.1 Segmentation

10.3.2 Synchronization and Overlap

10.3.3 Regularization of Word Forms

10.3.4 Prosody

10.3.5 Speech Management

10.3.6 Analytic Coding

- TEI P5 Chapter 11: Transcriptions of Speech
- TEI P5 Chapter 15: Simple Analytic Mechanisms
- TEI P5 Chapter 18: Transcription of Primary Sources
- TEI P5 Chapter 23: Language Corpora

```
<xsl:template match="//foreign">
```

```
<xsl:choose>
```

```
<xsl:when test="contains($style,'_ra') or contains($style,'_ra') or contains($style,'_a')">
```

```
ime="SPAN">
```

```
ame="style">
```

```
mplates />
```

```
bsp;[</xsl:template match="attribute:class" /><xsl:text>]</xsl:text>
```

```
ime="SPAN">
```

```
ame="style">
```

```
mplates />
```

```
<fo:page-sequence master-reference="basic">
```

```
<fo:static-content flow-name="xsl-region-before">
```

```
<fo:table border="1" color="#D3D3D3" text-align="center">
```

```
<tr><td style="background-color:#D3D3D3">
```

```
</fo:block>
```

```
</fo:static-content>
```

```
<fo:static-content flow-name="xsl-region-after">
```

```
<fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
```

```
</fo:block>
```

```
</fo:static-content>
```

```
<fo:static-content flow-name="xsl-footnote-separator">
```

```
<fo:block>
```

```
<fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
```

```
</fo:block>
```

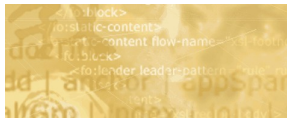
```
</fo:static-content>
```

```
<fo:flow flow-name="xsl-region-body">
```

```
<xsl:apply-templates select="//text"/>
```

```
</fo:flow>
```

```
</fo:page-sequence>
```

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

```
<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| altGrp | index | join | joinGrp | link | linkGrp | timeline
| cb | lb | milestone | pb)*, ) >
```

Elements Unique to Spoken Texts in TEI

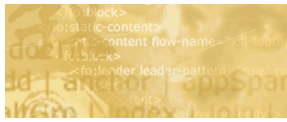
- `<u>` utterance.
- `<pause>` a pause.
- `<vocal>` vocalized semi-lexical.
- `<kinesic>` any communicative phenomenon, for example a gesture, frown, etc.
- `<event>` any phenomenon or occurrence, for example incidental noises or other events affecting communication.
- `<writing>` a passage of written text revealed to participants in the course of a spoken text.
- `<shift>` marks the point at which some paralinguistic feature of a series of utterances by any one speaker changes.

```

<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="contains($style, '_mn') or contains($style, '_n') or contains($style, '_ra') or contains($style, '_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style">font-style: italic</xsl:attribute>
        <xsl:apply-templates />
      </xsl:element>
    <xsl:text>&nbsp;</xsl:text> <xsl:value-of select="attribute::lang" /> <xsl:text></xsl:text>
  </xsl:choose>
</xsl:template>
<xsl:element name="SPAN">
  <xsl:attribute name="style">font-style: italic</xsl:attribute>
  <xsl:apply-templates />
</xsl:element>
</xsl:choose>
</xsl:template>
<fo:static-content flow-name="xsl-region-before">
  <fo:block background-color="#D3D3D3" text-align="center">
    <fo:static-content>
      <fo:static-content flow-name="xsl-region-after">
        <fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
      </fo:block>
    </fo:static-content>
  </fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-region-body">
  <xsl:apply-templates select="//text"/>
</fo:flow>
</fo:page-sequence>

```

Source: TEI P5, chapter 11 Transcriptions of Speech)



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

```
<!ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| altGrp | index | join | joinGrp | link | linkGrp | timeline
| cb | lb | milestone | pb)*)+) >
```

Sample: Coding of schwa

```
"http://www.w3.org/1999/XSL/Transform" version="1.0">
```

<div>

<u who="A">Heter hun Ronja Langangen?</u>

<u who="B">Nei hun heter Sonja Langang<c rend="vowel"
ana="phon.schwa">e</c>n</u>

</div>

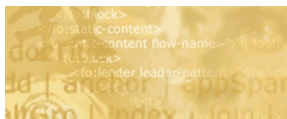
<u who="A">Heter hun Nelly Dalen?</u>

<u who="B">Nei hun heter Molly Dal<c rend="empty"
ana="phon.schwa">e</c>n</u>

</div>

```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```

```
<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="contains($style, '_mn') or contains($style, '_n') or contains($style, '_ra') or contains($style, '_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style">font-style: italic</xsl:attribute>
      </xsl:element>
    <xsl:otherwise>
      <xsl:text>&nbsp;</xsl:text> <xsl:value-of select="attribute::lang" /> <xsl:text></xsl:text>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>
```



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

Sample : ...more on schwa

```
<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| altGrp | index | join | joinGrp | link | linkGrp | timeline
| cb | lb | milestone | pb)*+)) >
```

= "http://www.w3.org/1999/XSL/Transform" version="1.0">

<u who="B">Nei hun heter Molly Dal<c rend="empty" ana="phon.schwa">e</c>n</u>

```
<xsl:template match="//foreign">
<xsl:when test="contains($style, '_mn') or contains($style, '_n') or contains($style, '_ra') or contains($style, '_a')">
<xsl:element name="SPAN">
<xsl:attribute name="style">font-style: italic</xsl:attribute>
<xsl:apply-templates />
</xsl:element>
```

<text> <front> <u who="B">Nei hun heter Molly Dal<del rend="empty" ana="phon.schwa">en</u>

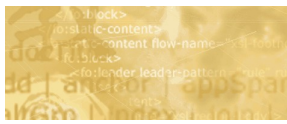
```
<pb/><pb ed="HU" n="175"/>
<lb/><fw type="page" place="top-right" hand="h2">1</fw>
<titlePage type="main">
<docTitle>
<lb/><titlePart type="main">Dalen</titlePart>
<lb/><titlePart type="dear">Et dramatisk Digt</titlePart>
</docTitle>
<byline>
<lb/>af
<lb/><docAuthor>Henrik Ibsen.</docAuthor>
</byline>
```

<u who="B">Nei hun heter Molly Dal<choice><reg>e</reg><orig ana="phon.schwa"></orig></choice>n</u>

```
<fo:page-sequence master-reference="basic">
<fo:static-content flow-name="xsl-region-before">
<fo:block background-color="#D3D3D3" text-align="center">
<xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm1/title"/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-region-after">
<fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
</fo:block>
</fo:static-content flow-name="xsl-footnote-separator">
<fo:block>
<fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
</fo:block>
</fo:static-content>
<fo:flow flow-name="xsl-region-body">
<xsl:apply-templates select="//text"/>
</fo:flow>
</fo:page-sequence>
```

<u who="B">Nei hun heter Molly <reg ana="phon.schwa">Dalen</reg></u>

<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">



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

Sample: Prosodic features

```
<!ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| altGrp | index | join | joinGrp | link | linkGrp | timeline
| cb | lb | milestone | pb)*)+) >
```

```
= "http://www.w3.org/1999/XSL/Transform" version="1.0">
```

```
<!ENTITY Ir "?">
intonation -->
```

```
<!-- low rise
```

• Might be encoded using entities.

```
<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="contains($style,'_a') or contains($style,'_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style">
          <xsl:apply-templates />
        </xsl:attribute>
      </xsl:element>
    <xsl:otherwise>
      <xsl:text>&nbsp;</xsl:text> <xsl:value-of select="attribute::lang" /> <xsl:text>] </xsl:text>
    </xsl:otherwise>
  </xsl:choose>
  <xsl:element name="SPAN">
    <xsl:attribute name="style">font-style: italic</xsl:attribute>
  </xsl:element>
</xsl:template>
```

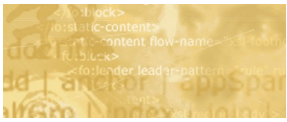
```
<!ENTITY rf "!">
intonation -->
```

```
<!-- rise fall
```

```
<u who="person2">hvilket
da&lr;</u>
```

```
<fo:page-sequence master-reference="basic">
  <fo:static-content flow-name="xsl-region-before">
    <fo:block background-color="#D3D3D3" text-align="center">
      <xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/ttitle"/>
    </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-region-after">
    <fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
  </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-footnote-separator">
    <fo:block>
      <fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
    </fo:block>
  </fo:static-content>
  <fo:flow flow-name="xsl-region-body">
    <xsl:apply-templates select="//text"/>
  </fo:flow>
</fo:page-sequence>
```

```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```



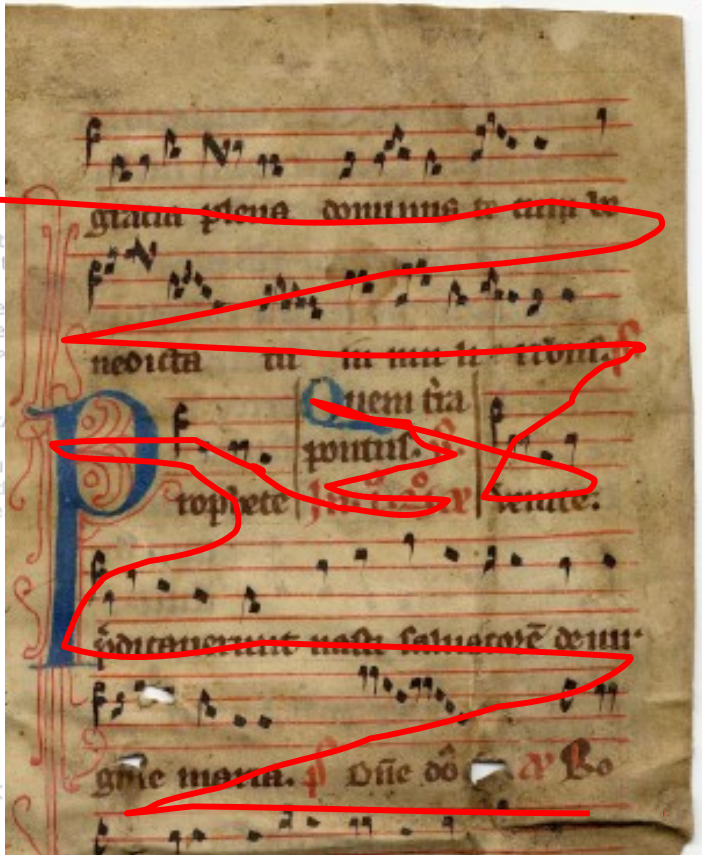
<?xml version="1.0" encoding="UTF-8" ?>

<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | ch | lb | milestone | pb)*, (titlePart, (figure
| table | anchor | appSpan | addSpan | delSpan | gap | alt
| altGrp | index | join | joinGrp | link | linkGrp | timeline
| ch | lb | milestone | pb)*)+>

Problems with text encoding

"http://www.w3.org/1999/XSL/Transform" version="1.0">

<text>
<front>
<pb/><pb
<lb/><fw t
<titlePage t
<docTitle>
<lb/><title
<lb/><title
</docTitle>
<byline>
<lb/>af
<lb/><doc
</byline>
<lb/><figu
<lb/><doc
</titlePage

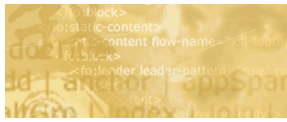


"h="//foreign">

t="contains(\$style
nt name="SPAN">
ite name="style">
ly-templates />
ent>
t> </xsl:tr
><xsl:value-of select="attribute::lang" /><xsl:text>
>>
nt name="SPAN">
ite name="style">
ly-templates />
ent>
se>

- Texts are more complex than one might think.
- Text encoding can be a both a very philosophic and a very prosaic task.
- Time consuming.

<fo:page-sequence master-reference="basic">
<fo:static-content flow-name="xsl-region-before">
<fo:block background-color="#D3D3D3" text-align="center">
<xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/titl"/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-region-after">
<fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-footnote-separator">
<fo:block>
<fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
</fo:block>
</fo:static-content>
<fo:flow flow-name="xsl-region-body">
<xsl:apply-templates select="//text"/>
</fo:flow>
</fo:page-sequence>



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

<ELEMENT docTitle ((figure | add | anchor | appSpan | addSpan | delSpan | gap | alt | altGrp | index | join | joinGrp | link | linkGrp | timeline | cb | lb | milestone | pb)*, (titlePart, (figure | add | anchor | appSpan | addSpan | delSpan | gap | alt | altGrp | index | join | joinGrp | link | linkGrp | timeline | cb | lb | milestone | pb)*)+) >

... and more problems

= "http://www.w3.org/1999/XSL/Transform" version="1.0">

”This *is a problem* to encode in XML”.

<text>
<front>

<pb/><pb ed="HU" n="175"/>
<lb/><fw type="page" place="top-right" hand="h2">1</fw>
<titlePage>
<docTitle>
<lb/><titlePart type="main">Brand.</titlePart>
<lb/><titlePart type="text">Et dramatisk Digt</titlePart>
</docTitle>
<byline>
<lb/>af
<lb/><docAuthor>Henrik Ibsen.</docAuthor>
</byline>
<lb/><figure type="bar" /></figure>
<lb/><docDate>1865.</docDate>
</titlePage>

<u>This **is a** *problem* **to** encode in XML</u>.

<xsl:template match="//foreign">

<xsl:choose>

<xsl:when test="contains(\$style, '_n') or contains(\$style, '_ra') or contains(\$style, '_a')">

<xsl:element name="SPAN">

<xsl:attribute name="style">font-style: italic</xsl:attribute>

<xsl:apply-templates />

</xsl:element>

<xsl:text> </xsl:text><xsl:value-of select="attribute::lang" /><xsl:text></xsl:text>

</xsl:when>

<xsl:otherwise>

<xsl:element name="SPAN">

<xsl:attribute name="style">font-style: italic</xsl:attribute>

<xsl:apply-templates />

</xsl:element>

</xsl:otherwise>

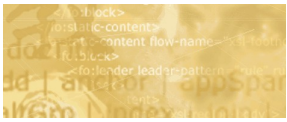
</xsl:choose>

</xsl:template>

- Overlap
- Discontinuous elements
- Alternative element orderings

<fo:page-sequence master-reference="basic">
<fo:static-content flow-name="xsl-region-before">
<fo:block background-color="#D3D3D3" text-align="center">
<xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm/ttitle"/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-region-after">
<fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
</fo:block>
</fo:static-content>
<fo:static-content flow-name="xsl-footnote-separator">
<fo:block>
<fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
</fo:block>
</fo:static-content>
<fo:flow flow-name="xsl-region-body">
<xsl:apply-templates select="//text"/>
</fo:flow>
</fo:page-sequence>

<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">



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

Questions to be solved

```
<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| link | linkGrp | index | join | joinGrp | link | linkGrp | timeline
| cb | lb | milestone | pb)*)+) >
```

```
"http://www.w3.org/1999/XSL/Transform" version="1.0">
```

```
<xsl:template match="//foreign">
  <xsl:choose>
    <xsl:when test="contains($style, 'italic') or contains($style, '_a')">
      <xsl:element name="SPAN">
        <xsl:attribute name="style">font-style: italic</xsl:attribute>
        <xsl:apply-templates />
      </xsl:element>
    <xsl:otherwise>
      <xsl:text>&nbsp;</xsl:text><xsl:value-of select="attribute::lang" /><xsl:text></xsl:text>
    </xsl:otherwise>
  </xsl:choose>
  <xsl:element name="SPAN">
    <xsl:attribute name="style">font-style: italic</xsl:attribute>
    <xsl:apply-templates />
  </xsl:element>
</xsl:template>
```

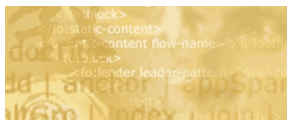
```
<text>
<front>
</front>
</text>
<pb/><pb ed="1" />
</pb/><fw ty="text" />
</fw/>
<titlePage>
<docTitl>
</docTitl>
</titlePage>
<byline>
</byline>
</titlePage>
```



- What should be encoded?
- Separate tiers?
- Should a standard encoding schema be used?
- Should the text be interchangeable?
- Is there a community that has to decide what schema to be used?

```
<fo:page-sequence master-reference="basic">
  <fo:static-content flow-name="xsl-region-before">
    <fo:page-header>
      <xsl:value-of select="/TEI.2/teiHeader/fileDesc/titleStm|title"/>
    </fo:page-header>
  </fo:static-content>
  <fo:static-content flow-name="xsl-region-after">
    <fo:page-footer>
      <fo:background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
    </fo:background-color>
  </fo:static-content>
  <fo:static-content flow-name="xsl-footnote-separator">
    <fo:block>
      <fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
    </fo:block>
  </fo:static-content>
  <fo:flow flow-name="xsl-region-body">
    <xsl:apply-templates select="//text"/>
  </fo:flow>
</fo:page-sequence>
```

```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```



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

```
<ELEMENT docTitle
((figure | add | anchor | appSpan | addSpan | delSpan |
gap | alt | altGrp | index | join | joinGrp | link | linkGrp
| timeline | cb | lb | milestone | pb)*, (titlePart, (figure
| add | anchor | appSpan | addSpan | delSpan | gap | alt
| timeline | cb | lb | milestone | pb)*)+) >
```

Links and references

```
= "http://www.w3.org/1999/XSL/Transform" version="1.0">
```

- *TEI P5 - Guidelines for Electronic Text Encoding and Interchange*, edited by C.M. Sperberg-McQueen and Lou Burnard, 2005 (<http://www.tei-c.org/release/doc/tei-p5-doc/html/>).

```
<xsl:template match="//foreign">
  <xsl:element name="SPAN" style="{style, '_ra'} or contains($style, '_a')">
    <xsl:attribute name="style" value="{style, '_ra'} or contains($style, '_a')"/>
  </xsl:element>
</xsl:template>
<xsl:otherwise>
  <xsl:element name="SPAN" style="font-style: italic"/>
  <xsl:apply-templates />
</xsl:otherwise>
</xsl:template>
```

- ROMA, <http://tei.oucs.ox.ac.uk/Roma/>

- XML, <http://www.w3.org/XML/>

- Markup Language for Complex Documents (MLCD), <http://teksttek.aksis.uib.no/projects/mlcd>

```
<fo:page-sequence master-reference="basic">
  <fo:static-content flow-name="xsl-region-before">
    <fo:block text-align="center">
      <fo:table border="1" style="width:100%; border-collapse: collapse;">
        <tr>
          <td style="text-align:center; padding: 5px;>Title</td>
          <td style="text-align:center; padding: 5px;>Desc</td>
        </tr>
      </fo:table>
    </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-region-after">
    <fo:block background-color="#D3D3D3" text-align="center">Page : <fo:page-number/>
  </fo:block>
  </fo:static-content>
  <fo:static-content flow-name="xsl-footnote-separator">
    <fo:block>
      <fo:leader leader-pattern="rule" rule-thickness="0.5pt" leader-length="33%"/>
    </fo:block>
  </fo:static-content>
  <fo:flow flow-name="xsl-region-body">
    <xsl:apply-templates select="//text"/>
  </fo:flow>
</fo:page-sequence>
```

```
<!DOCTYPE TEI.2 SYSTEM "ibsen-xml.dtd">
```