TEI@Oxford

July 2009

The transcription of primary sources

Which features of a primary source might one want to include in a transcription?

- variant letter forms
- page layout
- orthography
- capitalisation
- word division
- punctuation
- abbreviations
- additions and deletions
- errors and omissions

Elements defined for transcription

Defined in 'core' module: `<abbr>`, `<add>`, `<choice>`, `<corr>`, `<del>`, `<expan>`, `<gap>`, `<sic>`

Defined in 'transcr' module: `<addSpan>`, `<am>`, `<damage>`, `<damageSpan>`, `<delSpan>`, `<ex>`, `<facsimile>`, `<fw>`, `<handNotes>`, `<handShift>`, `<restore>` `<space>`, `<subst>`, `<supplied>`, `<surface>`, `<zone>`

Using `<choice>` in tanscriptions

- `<choice>` (groups alternative editorial encodings)
- Abbreviation:
    - `<abbr>` (abbreviated form)
    - `<expan>` (expanded form)
- Errors:
    - `<sic>` (apparent error)
    - `<corr>` (corrected error)
- Regularisation/normalisation:
    - `<orig>` (original form)
    - `<reg>` (regularised form)

Structure and layout

By *structure* is meant the division of the work into its constituent parts, by *layout* the arrangement of the text on the page.

The text of the work and the physical object carrying that text have separate structural hierarchies, both of which need ideally to be encoded. For the former the <div> element can be used for the largest structural divisions in prose texts, with a *@type* attribute to specify the nature of the division, "chapter", "section" etc. Paragraphs within these divisions can be tagged using <p>. Verse texts can be marked up using the tags <l> (for "line") and <lg> (for "line-group", i.e. a group of lines functioning as a formal unit), again with a *@type* attribute to identify the type of unit, e.g. "stanza", "couplet". Lines and line-groups can also be numbered and identified using the *@n* and *@xml:id* attributes. For the structure of the physical document, empty 'milestone' elements can be used, <pb/>, <cb/> and <lb/>, for page-, column- and line-boundaries respectively, which can also be numbered and provided with a *@xml:id*.

Abbreviations

Abbreviations are highly characteristic of manuscript materials of all kinds. It is customary to distinguish between several types:

Suspensions the first letter or letters of the word are written, generally followed by a point or other marker: for example e.g. for exempla gratia

Contractions both the first and last letters are written, often with some mark of abbreviation such as a superscript stroke, or, less commonly, a point: e.g. Mr. for Mister

Brevigraphs Special signs or tittels, such as the Tironian *nota* used for 'et' or the letter p with a barred tail commonly used for per.

Superscript letters Superscript letters (vowels or consonants) are often used to indicate various kinds of contraction: e.g. w followed by superscript ch for which.

Abbreviation and expansion

An abbreviation may be transcribed in two ways:

- One may choose to give the unexpanded abbreviation, transcribing it simply as a particular sequence of letters or marks on the page: thus, a 'p with a bar through the descender' or an 'a with a macron'
- One may also interpret or 'expand' the abbreviation, supplying the letter or letters it is seen as standing for: thus, 'per', 'an'

The TEI allows one to provide both the abbreviated and expanded forms.

Encoding abbreviations

The TEI proposes two levels of encoding:

- the whole of an abbreviated word and the whole of its expansion can be encoded using <abbr> and <expan>
- the mark or sign used to indicate the suppression of one or more letters, and the letters supplied in the process of expansion can be encoded using <am> and <ex>

Here too one may also use both levels simultaneously.

<am> and <ex>

Using these elements, from the 'transcr' module, the transcriber may indicate the status of the individual letters or signs within both the abbreviation and the expansion.

- <ex> (editorial expansion) contains a sequence of letters added by an editor or transcriber when expanding an abbreviation.
- <am> (abbreviation marker) contains a sequence of letters or signs present in an abbreviation which are omitted or replaced in the expanded form of the abbreviation.

Previously, people have re-purposed existing elements such as `<hi>` and `<supplied>` to mark individual letters/signs in abbreviations and expansions. The new P5 elements `<am>` and `<ex>` are the TEI's attempt to support this desire.

TEI

A simple example

The Icelandic word 'hann' ('he') is frequently written in medieval manuscripts as the letter h with a horizontal stroke or bar (Unicode character 0305, functionally similar to the modern tilde). It looks like this:

Encoding abbreviations 1

Depending on editorial policy, we might represent this in any one of the following ways:

```
<abbr>h&amp;#x305;</abbr> or
<expan>hann</expan>
```

```
 h<am>&amp;#x305;</am>or h
<ex>ann</ex>
```

```
<abbr>h<am>&amp;#x305;</am>
</abbr> or
<expan>h<ex>ann</ex>
</expan>
```

TEI

Using `<choice>`

Any of these pairs can be wrapped in `<choice>` tags:

```
 h<choice>
 <am>&amp;#x305;</am>
 <ex>ann</ex>
</choice>
```

```
<choice>
 <abbr>h<am>&amp;#x305;</am>
 </abbr>
 <expan>h<ex>ann</ex>
 </expan>
</choice>
```

Classifying abbreviations

The @*type* attribute on <abbr> allows us to distinguish between different types of abbreviation, either for statistical analysis or if they are to be rendered in varying ways, as is sometimes the case with suspensions, the expansions of which are placed in round brackets, while other types are italicised:

```xml
<choice>
 <abbr type="susp">k<am>.</am>
 </abbr>
 <expan>k<ex>onungr</ex>
 </expan>
</choice>
<choice>
 <abbr type="tittel">ml<am>¯</am>i</abbr>
 <expan>m<ex>æl</ex>l<ex>t</ex>i</expan>
</choice>
```

These can then be made to display as follows:

*k(onungr) mæll**t**i*

As elsewhere, the @*resp* and @*cert* attributes can also be used to indicate who is responsible for an expansion, and the degree of

Corrections and emendations

The `<sic>` element can be used to indicate that the reading of the manuscript is erroneous or nonsensical, while `<corr>` (correction) can be used to provide what in the editor's opinion is the correct reading:

```
<sic>giorit</sic>
```

```
<corr>giorir</corr>
```

The two may be combined within a `<choice>` element:

```
<choice>
 <sic>giorit</sic>
 <corr>giorir</corr>
</choice>
```

Normalisation/regularisation

Source texts rarely use modern normalised orthography. For retrieval and other processing reasons, such information may be useful in a transcription. The <reg> (regularized) element is available used to mark a normalised form, while the <orig> (original) element indicates a non-standard spelling. These elements can optionally be grouped as alternatives using the <choice> element.

TEI

Normalisation/regularisation (example)
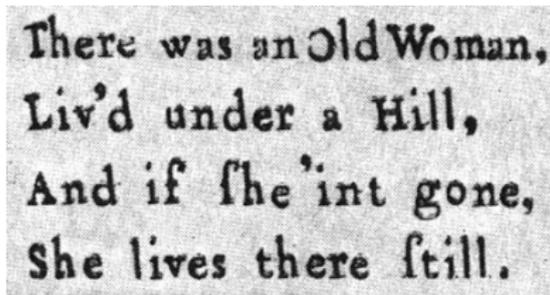
There was an Old Woman,
Liv'd under a Hill,
And if she 'int gone,
She lives there still.

```
<lg>
 <l>There was an Old Woman,</l>
 <l>
  <choice>
   <orig>Liv'd</orig>
   <reg>Lived</reg>
  </choice> under a hill,</l>
 <l>And if she <choice>
   <orig>'int</orig>
   <reg>isn't</reg>
  </choice> gone,</l>
 <l>She lives there still.</l>
</lg>
```

Additions, deletions and substitutions

Alterations made to the text, whether by the scribe or in some later hand, can be encoded using <add> (addition) or <del> (deletion).

- <add> (addition) contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector.
- <del> (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, annotator, or corrector.

Where the addition and deletion are regarded as a single *substitution*, they can be grouped together using the <subst> (substitution) element .

<add> and <del> Examples

```
by
the ancient Dutch navigators
<del rend="strikethrough" hand="#WI">of these
waters</del> the Tappaan Zee, and where they
always <add hand="#WI" place="supralinear">prudently</add>
shortened sail ...
<handNote xml:id="WI">Washington Irving holograph</handNote>
```

<addSpan> and <delSpan>

These two elements delimit a span of text by pointing mechanisms rather than by enclosing it. This is useful if an addition or deletion overlaps another span of text.

@*spanTo* indicates the end of a span initiated by the element bearing this attribute.

```
<addSpan spanTo="#id4"/>
<!-- added text -->
<anchor xml:id="id4"/>
```

Substitutions

<subst> (substitution) groups one or more deletions with one or more additions when the combination is to be regarded as a single intervention in the text. Examples:

- one word/letter written over another
- one word/letter deleted, replaced by another written above it by the same hand at one time
- one word/letter deleted, replaced by a different hand some other time
- a long chain of substitutions on the one stretch of text, with uncertainty as to the order of substitution and as to which of many possible readings should be preferred

<subst> examples

```
<l>
 <delSpan rend="verticalStrike" spanTo="#delend02"/> Tis
moonlight <subst>
  <del>upon</del>
  <add>over</add>
 </subst> Oman's sky
</l>
<l>Her isles of pearl look lovelily<anchor xml:id="delend02"/>
</l>
```

Deletion precedes addition by default, but this may be over-ridden
by means of the *@seq* attribute indicating sequence.

```
 One must have lived longer with <subst>
 <del seq="1">this</del>
 <del seq="2">
  <add seq="1">such a</add>
 </del>
 <add seq="2">a</add>
</subst> system, to appreciate its advantages.
```

TEI

Another `<subst>` example


```
<l>And towards our distant rest began to trudge,</l>
<l>
 <subst>
  <del>Helping the worst amongst us</del>
  <add>Dragging the worst amongt us</add>
 </subst>, who'd no boots
</l>
<l>But limped on, blood-shod. All went lame; <subst>
  <del>half-</del>
  <add>all</add>
 </subst> blind;</l>
<l>Drunk with fatigue ; deaf even to the hoots</l>
<l>Of tired, outstripped <del>fif</del> five-nines that dropped
behind.</l>
```

TEI

Cancellation of deletions and other markings

`<restore>` indicates restoration of text to an earlier state by cancellation of an editorial or authorial marking or instruction.

If in the line 'For I hate this my body' from D.H. Lawrence's poem *Eloi, Eloi, Lama Sabachthani?*, the 'my' was first deleted then restored by writing 'stet' in the margin, this might be encoded thus:

```
For I hate this
<restore hand="#dhl" type="marginalStetNote">
 <del>my</del>
</restore> body
```

Text omitted from or supplied in the transcription

Where a word has been supplied by the editor, `<supplied>` can be used. It is customary to distinguish between text now illegible or lost through damage but assumed originally to have been in the manuscript (which in some editorial traditions is printed in square brackets), and text assumed to have been inadvertently omitted by the scribe (printed in angle brackets). This distinction is indicated in the mark-up through the use of the @*reason* attribute:

```
…Dragging the worst
among<supplied reason="omitted">s</supplied>t us…
```

Metadata for supplied text

Attributes *@resp* and *@cert* can be used here as elsewhere. A *@source* attribute is also available to indicate that another witness supports the reconstruction:

```
<p>ath þeir <supplied reason="omitted" source="AM02-
152" cert="high">mundu</supplied> sundr
ganga</p>
```

When missing text cannot be confidently reconstructed, the <gap> element should be used. Its *@reason* attribute explains the reason for the omission and its *@extent* and *@unit* attributes indicate its presumed size.

```
<gap reason="damage" extent="7" unit="chars"/>
```

Other uses of <gap>

The <gap> element can also be used where material present and legible has been omitted in a transcription, whether for editorial reasons or as part of sampling practice.

```
<div rend="slide">
 <head>Lectio x.</head>
 <p> Hic itaque paterfamilias ad excolendam
<gap extent="20" unit="words" reason="not
transcribed" resp="#DC"/> congregare non desistit.
 </p>
</div>
```

Damage and illegibility

Use `<unclear>` if the text has been rendered partly illegible by deletion or damage so that the text can be read but without perfect confidence.

Use the *@reason* attribute to state the cause (damage, deletion etc.) of the uncertainty in transcription and the *@cert* attribute to indicate the confidence in the transcription.

```
 shore of the <unclear reason="damage" cert="medium">the Hudson,
at</unclear> that broad
```

The `<damage>` element should be used to indicate areas of damage affecting the text, but normally where at least some of the text can be read with confidence. The attributes *@agent* and *@extent* indicate the cause and extent of the damage respectively.

Headers, footers, catchwords and so on

<fw> (forme work) contains a running head (e.g. a header, footer), catchword or similar material appearing on the current page.

```
<fw place="top-centre" type="head">Poëms.</fw>
<fw place="top-right" type="pageno">29</fw>
<fw place="bot-centre" type="sig">E3</fw>
<fw place="bot-right" type="catch">TEMPLE</fw>
```

`<handNote>` and `<handShift>`

The `<handNote>` element is used to provide information about each hand distinguished within the encoded document.

- When the 'transcr' module is used, the element `<handNotes>` is available, within the `<profileDesc>` element of the Header, to hold one or more `<handNote>` elements. (brief)
- When the 'msdescription' module is included, the `<handDesc>` element also becomes available as part of a structured manuscript description. (more robust)

It is possible to use the two elements together if, for example, the `<handDesc>` element contains a single summary describing all the hands discursively, while the `<handNotes>` element gives specific details of each.

<handShift>

<handShift> marks the beginning of a sequence of text written in
a new hand, or the beginning of a scribal stint.

```
<l>When wolde the cat dwelle in his ynne</l>
<handShift medium="greenish-ink"/>
<l>And if the cattes skynne be slyk
<handShift medium="black-ink"/> and gaye</l>


<handNotes>
 <handNote xml:id="h1" script="copperplate">Carefully written with
   regular descenders</handNote>
 <handNote xml:id="h2" medium="pencil">Unschooled
scrawl</handNote>
</handNotes>
```

<handShift> example

```
<handShift new="#h1" resp="#das"/>... and that
good Order Decency and regular worship may be once
more introduced and Established in this Parish
according to the Rules and Ceremonies of the
Church of England and as under a good
Consciencious and sober Curate there would and
ought to be <handShift new="#h2" resp="#das"/> and
for that purpose the parishioners pray
```

```
<add place="supra" hand="#WJ" cert="medium"> But</add>
<choice>
 <sic>One</sic>
 <corr resp="#FB" cert="high">one</corr>
</choice> must have lived ...
<!-- elsewhere -->
<respStmt xml:id="FB">
 <resp>editorial changes</resp>
 <name>Fredson Bowers</name>
</respStmt>
<respStmt xml:id="WJ">
 <resp>authorial changes</resp>
 <name>William James</name>
</respStmt>
```

TEI