

Talk 8: Transcription and Primary Sources

Dot Porter

July 2009

Transcription of primary sources

In most cases when creating an electronic edition, the primary source from which it originates may have certain textual phenomena. For example:

- Text which is abbreviated and/or expanded
- Editorial correction and conjecture
- Scribal additions, deletions, substitutions, or restorations
- Text which an editor supplies or intentionally omits
- Text which is damaged or illegible

Transcribable features

Which features of a primary sources 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
- regularizations

Elements defined for transcription work

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

<choice> reminder

- <choice> (groups alternative editorial encodings)
- Abbreviation:
 - <abbr> (abbreviated form)
 - <expan> (expanded form)
- Errors:
 - <sic> (apparent error)
 - <corr> (corrected error)
- Regularization:
 - <orig> (original form)
 - <reg> (regularized form)

Abbreviation

Abbreviations are highly characteristic of manuscript materials of all kinds. Western MSS traditionally distinguish:

- 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 first and last letters are written, generally with some other mark of abbreviation such as a superscript stroke, or, less commonly, a point or points: e.g. Mr. for Mister
- Brevigraphs** Special signs or tittels, such as the Tironian *nota* used for 'et', the letter p with a barred tail commonly used for per, the letter c with a circumflex used for cum (ĉ) etc.
- Superscripts** 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

A manuscript abbreviation may be viewed in two ways:

- One may transcribe it as a particular sequence of letters or marks upon the page: thus, a 'p with a bar through the descender', a 'superscript hook', a 'macron'
- One may also interpret the abbreviation in terms of the letter or letters it is seen as standing for: thus, 'per', 're', 'n'

Both of these views may be recorded simultaneously by the TEI

Two Levels of Encoding Abbreviations

TEI proposes two levels of encoding:

- the whole of an abbreviated word and the whole of its expansion: `<abbr>` and `<expan>`
- abbreviatory signs or characters and the 'invisible' characters they imply: `<am>` and `<ex>`

<ex> and <am>

Using these elements, from the 'transcr' module, a 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 recommended tags for standardizing this markup.

Brevigraph Example

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



Encoding abbreviations

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

`<abbr>h̅</abbr>` or
`<abbr>h̄</abbr>`

`<expan>hann</expan>`

h`<am>̅</am>`

h`<ex>ann</ex>`

`<abbr>h<am>̅</am></abbr>`

`<expan>h<ex>ann</ex></expan>`

<abbr> and <expan> Examples

eu<g ref="#er">er</g>y <g ref="#per">per</g>sone that loketh after heven
hath a place in this ladder

```
<abbr>ev<g ref="#er">er</g>y</abbr>
<abbr>
  <g ref="#per">per</g>sone
</abbr> ...
```

```
<expan>every</expan>
<expan>persone</expan> ...
```

```
<choice>
  <abbr>ev<g ref="#er">er</g>y</abbr>
  <expan>every</expan>
</choice>
```

Brevigraph Alternative

We could also indicate multiple alternatives (at either level) by using the `<choice>` element

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

or

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

Classifying abbreviations

The `@type` attribute on `<abbr>` allows us to provide alternative renderings for the same markup in different contexts.

```
<choice>
  <abbr type="susp">k<am>&amp; #x307; </am>
</abbr>
  <expan>k<ex>onungr </ex>
</expan>
</choice>
<choice>
  <abbr type="tittel">ml<am>&amp; #x305; </am>i</abbr>
  <expan>m<ex>æll</ex>l<ex>t</ex>i</expan>
</choice>
```

k(onungr) mællti

As elsewhere, the `@resp` and `@cert` attributes can also be used to indicate who is responsible for an expansion, and the degree of certainty attached to it.

More Abbreviation and Expansion Examples (1)

`<abbr>SPQR</abbr>`

or

`<expan>senatus populusque romanorum</expan>`

More Abbreviation and Expansion Examples (2)

or

```
<choice>  
  <abbr>SPQR</abbr>  
  <expansion>senatus populusque romanorum</expansion>  
</choice>
```

or

```
<choice>  
  <abbr>SPQR</abbr>  
  <choice>  
    <expansion>senatus populusque romanorum</expansion>  
    <expansion>Sebastian Patrick Quintus Rahtz</expansion>  
  </choice>  
</choice>
```


Earlier Examples Now With <am> and <ex>

```
<abbr>ev<am>
  <g ref="#er" />
</am>y</abbr>
<abbr>
  <am>
    <g ref="#per" />
  </am>sone
</abbr> ...
```

or

```
<expan>ev<ex>er</ex>y</expan>
<expan>
  <ex>per</ex>sone
</expan> ...
```

And one more...

```
ev<choice>
  <am>
    <g ref="#er" />
  </am>
  <ex>er</ex>
</choice>y <choice>
  <am>
    <g ref="#per" />
  </am>
  <ex>per</ex>
</choice>sone ...
```

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>
```

Alternatively, they may be combined within a `<choice>` element, thus allowing the possibility of providing multiple corrections:

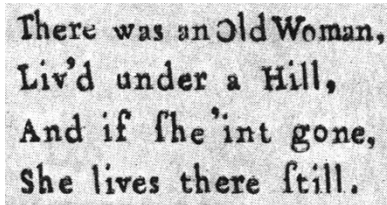
```
<choice>
  <sic>giorit</sic>
  <corr cert="high">giorir</corr>
  <corr cert="low">gioret</corr>
</choice>
```

Correction/Conjecture Examples

Nos autem iam ostendimus quod nutrimentum
et **<choice>**
 <sic>angues**</sic>**
 <corr>augens**</corr>**
</choice>.

Normalization

Source texts rarely use modern normalized orthography. For retrieval and other processing reasons, such information may be useful in a transcription. The `<r eg>` (regularized) element is available used to mark a normalized form; the `<or ig>` (original) element to indicate a non-standard spelling. These elements can optionally be grouped as alternatives using the `<choice>` element:



There was an Old Woman,
Liv'd under a Hill,
And if she 'int gone,
She lives there still.

Normalization example

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

- <add> (addition) or (deletion) are used for evident alterations in the source
- a combined addition and deletion may be marked using <subst> (substitution)

And towards our ^{distant} rest began to rudge,
~~Helping the worst amongst us~~ ^{Dragging the worst amongst us}, who'd no boots all
 But limped on, blood-shod. All went lame; half-blind;
 Drunk with fatigue; deaf even to the hoots
 Of tires, outstripped ~~five~~ ^{five} five-miles that dropped behind.

Additions and Deletions

- `<add>` (addition) contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector.
- `<addSpan />` (added span of text) marks the beginning of a longer sequence of text added by an author, scribe, annotator or corrector (see also `add`).
- `` (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.
- `<delSpan />` (deleted span of text) marks the beginning of a longer sequence of text deleted, marked as deleted, or otherwise signaled as superfluous or spurious by an author, scribe, annotator, or corrector.

<add> and Examples

by the ancient Dutch navigators ~~rend="strikethrough" hand="#WI">of these waters~~ the Tappaan Zee, and where they always ~~hand="#WI" place="supralinear">prudently~~ shortened sail ...

~~handNote xml:id="WI">Washington Irving holograph~~

<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 t> (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

Substitutions

<subst t> (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

Substitutions

<subst t> (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

Substitutions

<subst t> (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 lovely<anchor xml:id="delend02"/>
</l>

```

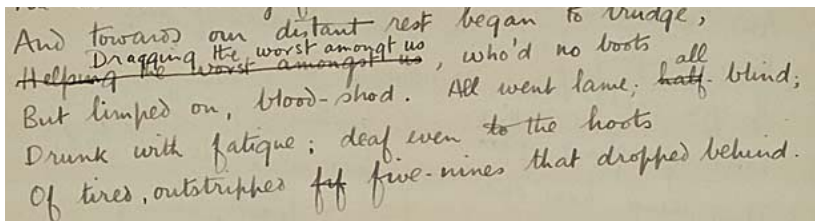
Deletion precedes addition, but this may be over-ridden by means of the @seq attribute indicating sequence.

```

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.

```

Another <subst> example



<l>And towards our distant rest began to trudge, </l>

<l>

<subst>

Helping the worst amongst us

<add>Dragging the worst amongst us</add>

</subst>, who'd no boots

</l>

<l>But limped on, blood-shod. All went lame; <subst>

<del status="shortEnd">half-

<add>all</add>

</subst> blind; </l>

<l>Drunk with fatigue ; deaf even to the hoots</l>

<l>Of tired, outstripped fuf five-nines that dropped behind. </l>

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.

by the ancient Dutch navigators `<restore hand="#WI2">`
`<del rend="strikethrough" hand="#WI2">`of these waters``
`</restore>` the Tappaan Zee, and where they always
`<add hand="#WI2" place="supralinear">`prudently`</add>` shortened sail ...
`<handNote xml:id="WI2">`Washington Irving
holograph`</handNote>`

<restore>

If in 'For I hate this my body', the *my* was first deleted then restored by writing 'stet' in the margin. This may be encoded:

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

Text Omitted from or Supplied in the Transcription

- `<gap>` indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible or inaudible.
- `<suppl ied>` signifies text supplied by the transcriber or editor for any reason, typically because the original cannot be read because of physical damage or loss to the original.

Supplied text

Sometimes, a transcript may need to include words not visibly present in the source:

- because the carrier has been damaged or is barely legible
- because of (assumed) scribal error

The `<supplied>` element is provided for use in either situations; the `@reason` attribute is used to distinguish them.

```
...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 peir <supplied reason="omitted" source="AM02-152">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="cm"/>
```

<gap> and <supplied> examples

expansion `<gap reason="illegible" agent="water"/>` river denominated

expansion `<supplied reason="illegible" source="#SH1862">`of
the`</supplied>`river denominated

<gap> Example

```
<div>
  <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>
```

More <suplied>

Where the transcriber considers that one or more words have been erroneously omitted in the original source and corrects this omission, the <suplied> element should be used in preference to <corr>.

by the ancient Dutch navigators
<suplied>of</suplied> the Tappan Zee

<supplied> Example

```
<p>Oblatus est  
<supplied reason="omitted" resp="#DC"> quia ipse voluit</supplied>.  
</p>
```

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

<damage>, <space>, and <unclear> Example

Revelabunt caeli iniquitatem Judae et <damage agent="rubbing"/> consurget et
<space/> manifestum erit peccatum ipsius in die furoris
do<unclear agent="rubbing" resp="#JC">mini</unclear> cum eis qui dixerunt
domino deo recede a nobis scientiam viarum tuarum nolumus

Damage and Illegibility

Use `<damage>` if the text can be read with perfect confidence

```

<p>
<!-- ... -->
  <pb n="5r"/>
  <damageSpan agent="rubbing" extent="whole leaf" spanTo="#damageEnd"/>
</p>
<p> .... </p>
<p> .... <pb n="5v" xml:id="damageEnd"/>
</p>

```

Disjoint Damage

IN the bosom **<damage group="1">o</damage>**f one of those spa**<lb n="2"/>**cious
 coves wh**<damage group="1">ich inde</damage>**nt the eastern **<lb n="3"/>**shore of
 the **<damage group="1">Hudson, at </damage>**that broad **<lb n="4"/>**expansion
<damage group="1">of the r</damage>iver denominated **<lb n="5"/>**by the
 ancie**<damage>**nt**</damage>** Dutch navigators

<fw>

<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>
```


<handShift> Example

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

```
<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
```

@hand, @resp, @cert

```

<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>

```