

# Exercise 5: Transcription of Handwritten Materials

Dot Porter

March 2009

In this exercise, we'll encode the text from a manuscript in reference to that manuscript, focusing on markup that describes the text's relationship to the object on which it is inscribed.

## 1 Basic navigational encoding

The files *whitman\_27.jpg*, *whitman\_28.jpg* and *whitman\_42.jpg* (located in *documents/images*) are images of the manuscript pages. The text will be encoded in reference to these images. Open *whitman\_27.jpg* in your favorite image viewer (or in FireFox), and position the window parallel to oXygen (either horizontally or vertically). You can open the other images in this window as needed.

Go to the *documents/xml* folder and open the file *whitman\_transcription.xml*. This file contains a basic TEI wrapper including the transcription of the text from pages 27-28 and page 42. The transcription is between `<div type="ms"><ab></ab></div>`. The `div` type indicates that the contents are the contents of the manuscript; `<ab>` is "anonymous block", which can hold any "arbitrary, component-level unit of text." As these are not complete poems but rather notes, it does not make sense at this point to tag these either as poems or using paragraphs. `<div>`, however, may not have plain text content, so we must have some container for the text inside of `<div>`.

Fill in the basic header information as requested.

## 2 Marking page breaks and line breaks

Page numbers are indicated in the transcription as a number contained in square brackets, as [27]. TEI indicates page numbers using the page break (`<pb/>`) tag with the `@n` attribute indicating the specific page number. Replace those page number with `<pb/>` tags (these will always appear at the start of a page). Use the `@n` attribute to indicate the page number.

Line numbers are not indicated in the transcription, so we will have to add them. In the TEI, line numbers are indicated using the line break (`<lb/>`) tag with the `@n` attribute indicating the specific line number. You will note that the lines in the transcription do not follow the lines on the manuscript. You may wish to modify the lines, or you may keep them as they are. Just make sure that you put the `<lb/>` tags in the right places. (Note: there are 18 lines on page 27, five lines on page 28, and 15 lines on page 42.)

## 3 Marking scribal deletions and additions (basic)

On page 27 there is one instance of a scribal deletion. Whitman originally wrote "a locust", but then deleted "a" and used instead "the". He did this immediately – we can tell because "the" follows immediately on the line, and "blossoms" is plural, not singular (we would expect the "s" to be added on the end if it were added later, but it is clearly a part of the original word). So is this a deletion plus an

addition? Or is it just a deletion? It could be interpreted either way, but because “the” is clearly not a later addition we shall interpret this as a deletion standing on its own (on page 28 and later on page 42 we’ll mark several deletions + additions). To mark this deletion, insert an “a” in the relevant place at the beginning of line 12, and mark it with `<del>`. We would like to indicate how we know that this letter was deleted; we do this using the `@rend` attribute (which describes how the deletion appears). Use the value “overstrike”.

```
<lb n="12"><del rend="overstrike">a</del> the locust blossoms—
```

On page 28 there is one instance of a scribal deletion plus addition at the beginning of line 2. The deleted text is illegible (at least the workshop leader could not figure it out!), so must be indicated using a `<gap/>` tag. As the deleted text is not legible, the value for `@reason` on `gap` should be “illegible.” Again, the deleted text has been stricken out, so the value of `@rend` on `<del>` should be “overstrike”. “spirit” has been added above the line, so it should be tagged with `<add>`, with the value of the `@place` attribute as “above” (for “above the line”). As the word “spirit” is clearly a replacement for the deleted text, the two should be grouped together using the `<subst>` tag, which is described in the Guidelines as “group[ing] one or more deletions with one or more additions when the combination is to be regarded as a single intervention in the text.” Your finished encoding should look like this:

```
81 <subst>
82 <del type="overstrike">
83 <gap reason="illegible"/>
84 </del>
85 <add place="above">spirit</add>
86 </subst>
```

At this point you may wish to pretty print your code. Be aware that doing this will relocate your `<lb/>` tags; if you restructured your text to have each encoded line start with `<lb/>` your structure will be lost. It may be better for you to add hard returns in the substitution group rather than automatically pretty printing. (If you pretty-printed already and would like to return to the previous organization, press `ctrl + z`, as you would in Microsoft Word)

## 4 Viewing the transcription in a browser

`whitman_transcription.xml` is linked to the `whitman.css` file. At any point you may open the XML file in your browser and check how your code is displaying. Additions and deletions should be noticeably different from surrounding text.

## 5 Marking scribal deletions or additions (more advanced)

Before we start encoding page 42, we want to indicate that we have left out several pages between page 28 and page 42, for reasons of sampling (that is, we only need a few pages for this exercise). Insert a `<gap/>` tag before the page break for page 42, with `@reason="sampling"`.

Page 42 is rather more complex than pages 27 and 28. Before you start tagging, read through the entire page. You will notice that there are places where text has been deleted but nothing added in its place, as well as places where text has been added on its own (without an accompanying deletion). In these cases, you should use `<add>` and `<del>` on their own, without a parent `<subst>`. In other cases, text has been deleted and replaced with other text, and then that added text likewise has been deleted and replaced. Remember that `<subst>` may nest within `<del>` and `<add>`, so if you have

text that is deleted and replaced, and then that added text is deleted and replaced, you may indicate that through nesting.

The transcription we have provided contains only the final text, not any deletions. Try to provide the deleted text as you proceed encoding; if it is illegible, use `<gap reason="illegible">` as above.

Start on line 5. Mark that “how” has been deleted and replaced with “where” (above the line), and that “sound” has been added above the line between “the” and “man’s”. Your finished encoding should look like this:

```

120 | <lb n="5">the west,
121 | <subst>
122 | <del rend="overstrike">how</del>
123 | <add place="above">where</add>
124 | </subst> the <add place="above">sound</add> man's

```

Go back up to line 3. This one is a bit more complex, with nested additions and deletions. The original line “you show me how” has been deleted in two separate acts, with one part replaced with “I imagine” and the other with “where” (or is it that the whole part has been replaced with “I imagine where”? This is an editorial decision. For this exercise we’re interpreting it in two parts, but there are other possible interpretations, which would influence your tagging decisions). Then later “imagine” was deleted and replaced with “see”. This requires placing a `<subst>` (with `<add>` and `<del>`) inside the `<add>` containing “I imagine”. When you are done, your encoding should look like this (although I seem to have forgotten some attributes – add them if you can):

```

105 | <lb n="3">When
106 | <subst>
107 | <del rend="overstrike">you show me</del>
108 | <add>I
109 | <subst>
110 | <del>imagine</del>
111 | <add>see</add>
112 | </subst>
113 | </add>
114 | </subst>
115 | <subst>
116 | <del rend="overstrike">how</del>
117 | <add place="above">where</add>
118 | </subst>

```

Finish encoding page 42. If you have any questions, please don’t hesitate to ask for assistance. Remember that if you can’t read a word, replace it with `<gap/>`. Note that two full lines have been deleted at the end; you’ll need to add line breaks for these.