TEI BY EXAMPLE



MODULE 6: PRIMARY SOURCES

Ron Van den Branden

Edward Vanhoutte

Melissa Terras

Centre for Scholarly Editing and Document Studies (CTB) , Royal Academy of Dutch Language and Literature, Belgium, Gent, 9 July 2010

Last updated September 2020

Licensed under a Creative Commons Attribution ShareAlike 3.0 License

TABLE OF CONTENTS

1. Jeremy Bentham: manuscript JB/116/010/001 1

2. Walt Whitman: “After the Argument” 3

1. Jeremy Bentham: manuscript JB/116/010/001

This manuscript page was written by the philosopher and jurist Jeremy Bentham (1748–1832).

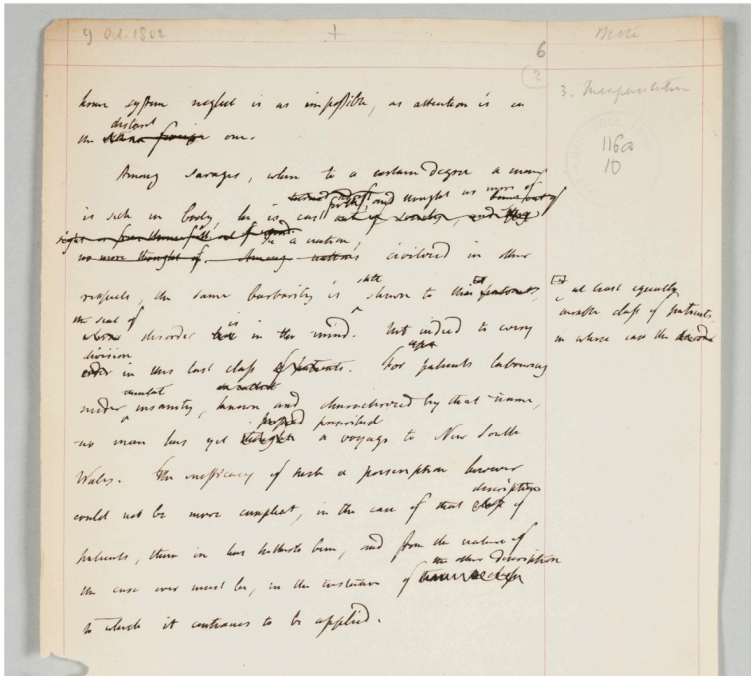

Figure 1. The manuscript page JB/116/010/001.

Since this is a prose text, the basic structural units are encoded as paragraphs (<p>), with line breaks encoded as <lb> where they occur. Marginal notes are encoded with the <note> element; the note occurring on the sixth line in this example contains a simple deletion (the final word “disorder”), which is marked with the <del> element. This manuscript contains many deletions and additions. Some are simple, such as the addition of the word “still” in the phrase “the same barbarity is still shown” on line 6. This is indicated in the encoding by wrapping the added content in a <add> element. More often, deletions and additions occur in combination, in which case the transcriber tries to reflect their order in the nesting of <del> and <add> elements. For example:

```
<add>forth, <add>
  <del>turned adrift</del>
</add> and thought no more of <add>
  <del><gap/> out of</del>
</add></add>
```

This fragment is marked as an addition. Yet, Bentham had emended this addition by adding the phrase “turned adrift” (as a second-level addition). Later, he canceled this addition by deleting it again: that’s why this phrase is encoded inside <add>, with a nesting <del>, indicating that this added text had been deleted entirely. Further, at the end of this fragment, another addition is indicated with an <add> element. Again, this entire second-level addition had been deleted. Yet, since the encoder could not decipher the deleted text anymore, this is indicated by the empty <gap> element, which signals that text was present on the manuscript, but left out from the transcription. In order to record why the transcriber had decided to omit this text, a @reason attribute could have been provided on <gap> (with a value such as "illegible").

Where text could still be transcribed, but the encoder is not certain of the reading, this reading is recorded in an <unclear> element, as is the case with the word “that,” occurring in a deletion on the last but one line.

Finally, when the encoder spotted obvious mistakes, these have been identified with the <sic> element, as is the case with the word “compleat.” The encoder could equally have provided a correction by wrapping both the incorrect form (<sic>) and correction (<corr>) inside a <choice> element:

```
<choice>
  <sic>compleat</sic>
  <corr>complete</corr>
</choice>
```

```
<text xmlns="http://www.tei-c.org/ns/1.0">
  <body>
    <div>
      <p>home system neglect is as impossible, as attention is in<lb/> the <del>other
       foreign</del> one.</p>
      <p>Among savages, when to a certain degree a man<lb/> is sick in body, he is
       cast <del>out of society, and then</del> <add>forth, <add>
        <del>turned adrift</del>
      </add> and thought no more of <add>
        <del><gap/> out of</del>
      </add></add><lb/> <del>no more thought of.</del> <add>
        <del>sight or from thenceforth out of mind.</del>
```

```
      </add> <del>Among nations</del> <add>In a nation,</add> civilized in
       other<lb/> respects, the same barbarity is <add>still</add> shown to
       this <note>at least equally unable class of patients, in whose case
       the <del>disorder</del></note> <del>patient,</del><lb/> <del>
        <gap/>
      </del> <add>the seat of</add> disorder <del>has</del> <add>is</add> in the mind.
       Not indeed to every<lb/> <del>order</del> <add>division</add> in this last
       class <del>of patients</del>. For <add>
        <unclear>upon</unclear>
      </add> patients labouring<lb/> under <add>
        <gap/>
      </add> insanity, known <add>
        <del><unclear>or</unclear> called</del>
      </add> and characterized by that name,<lb/> no man has
       yet <del>thought</del> <add><del>proposed</del> prescribed</add> a voyage to New
       South<lb/> Wales. The inefficacy of such a prescription however<lb/> could not
       be more <sic>compleat</sic>, in the case of that <del>class</del> description
       of <lb/> patients, than in has hitherto been, and from the nature of<lb/>
       the case ever must be, in the instance of <del><unclear>that</unclear> <gap/>
       class</del> <add>the other description</add><lb/> to which it continues to be
       applied.</p>
    </div>
  </body>
</text>
```

Example 1. Encoding of manuscript JB/116/010/001 by Jeremy Bentham (1802). TEI XML source available from https://transcribe-bentham.ucl.ac.uk/td/JB/116/010/001?action=edit.

2. Walt Whitman: “After the Argument”

This manuscript, featuring an early version of the poem “After the Argument,” was likely written in 1890 or early 1891, shortly before the poem’s publication.

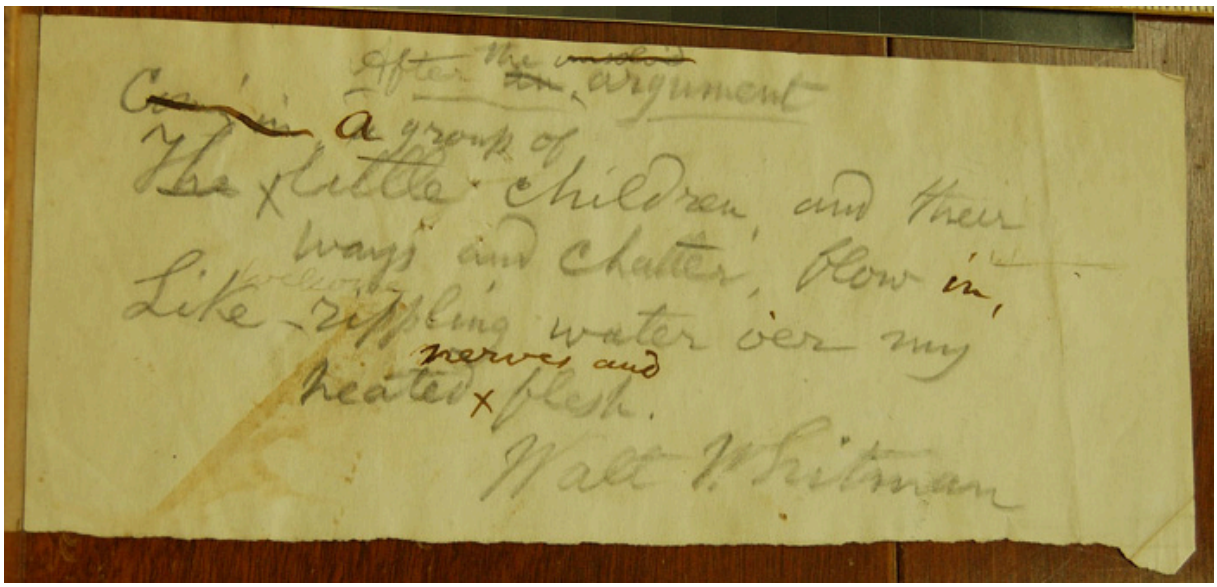

Figure 2. A facsimile of the manuscript for "After the Argument."

This example clearly illustrates how the TEI **`transcr`** module can be applied to verse texts as well. The entire poem is encoded inside `<lg type="poem">`, containing a heading (`<head>`) and two verse lines (`<l>`), in which physical line breaks have been maked with `<lb>` elements. As will be clear from the facsimile, this short manuscript features some complex editorial traces. Sequential deletions (`<del>`) and additions (`<add>`) are grouped into substitutions (`<subst>`). Moreover, inside the substitutions, the exact order of the editing interventions is specified by means of a sequence number in a `@seq` attribute, making explicit that the deletions occurred before the additions.[1]

This example illustrates nicely how additions and deletions can nest. In both cases in the example, an addition contains further deletions. The `@rend` attribute is used on `<del>` and `<add>` in order to encode the way in which deletions (`"overstrike"` or `"overwrite"`) and additions (`"insertion"`, `"overwrite"`, `"unmarked"`) have been realised on the manuscript. An additional `@place` attribute on `<add>` indicates if the additions are located `"supralinear"`, `"over"` existing text, or `"inline"`.

.....

1 The `@seq` attribute is a more advanced concept documented in chapter 11 Representation of Primary Sources of the TEI Guidelines. Notice how this explicit sequence number is not strictly needed here, as deletions logically precede additions, and only one deletion is involved.

Page breaks are indicated with <pb> elements, indicating the type of the manuscript page ("recto" or "verso") with the @type attribute. The @facs attribute points to a digital facsimile of the page indicated with the <pb> element.

```
<body xmlns="http://www.tei-c.org/ns/1.0">
  <pb facs="loc.00001.001.jpg" xml:id="leaf001r" type="recto"/>
  <lg type="poem">
    <head rend="underline" type="main-authorial">After <subst>
      <del rend="overstrike" seq="1">an</del>
      <add place="supralinear" rend="insertion" seq="2">the <del rend="overstrike">unsolv
      'd</del></add>
    </subst> argument</head>
    <l><del rend="overstrike">The</del> <add place="supralinear" rend="insertion"><del rend="overstrike
     in,</del> <subst>
      <del rend="overwrite" seq="1">a</del>
      <add place="over" rend="overwrite" seq="2">A</add>
    </subst> group of </add> little children, and their <lb/>ways and chatter,
     flow <add place="inline" rend="unmarked">in, </add> <del rend="overstrike">
      <add place="supralinear" rend="unmarked">upon me</add>
    </del></l>
    <l>Like <add place="supralinear" rend="insertion">welcome </add> rippling water
     o'er my <lb/>heated <add place="supralinear" rend="insertion">nerves and </add>
     flesh. </l>
    <closer>
      <signed>Walt Whitman</signed>
    </closer>
  </lg>
  <pb facs="loc.00001.002.jpg" xml:id="leaf001v" type="verso"/>
</body>
```

Example 2. Encoding of manuscript loc.00001 of the Walt Whitman Archive (Whitman 1890).

TEI XML source available from https://whitmanarchive.org/manuscripts/tei/loc.00001.xml.

BIBLIOGRAPHY

Bentham, Jeremy. 1802. “Manuscript JB/116/010/001.” Manuscript encoded and made available by the Transcribe Bentham project at https://transcribe-bentham.ucl.ac.uk/td/JB/116/010/001.

Whitman, Walt 1890. “After the Argument.” Manuscript encoded and made available by the Walt Whitman Archive at https://www.whitmanarchive.org/manuscripts/transcriptions/loc.00001.html.