

Transcribing a historical source

Lou Burnard Consulting

July 2014

1 The Robert Graves Project

In this and subsequent exercises, we will use the TEI to mark up an important literary source document: the diaries of the English writer Robert Graves. Four years of this diary exist, along with much other material, at the University of Victoria, by whose kind permission we are using them in this tutorial. Copyright in the materials is owned by the Robert Graves Trust, which does however make them available for scholarly and educational purposes.

In your **Work** folder there is a folder called **Graves** which contains digital images of a few pages from the diary for 1938. You may like to look at them quickly to see the kind of material you will be working on. (The University of Victoria also makes the whole set available from its web site)

In this first session, the aim is to make a transcription which covers:

- the written text and its organization into entries
- some editorially significant aspects such as deletions or additions
- detailed indexing of proper names of persons and places
- provision of explanatory notes etc.

We will begin with the entry for 26 Nov 1938, which you'll find in the file 19381126_1378.jpg.

2 Structural tagging of a diary entry

- Open oxyGen and make a new XML document. Click the **Customize** button to specify the schema you will be using. Then select the `gravesDoc.rnc` schema from your **Work** folder
- This schema provides all the elements you'll need in the following exercises. It was made from the ODD file `gravesDoc.odd` which is also available for the curious.
- Leave the header alone for the moment. Inside the `<body>`, add first a `<pb>` element, to indicate that a new page starts here. Supply the page number (**1378**) as the value of the `@n` attribute, and the name of the image file corresponding with this page as the value of the `@facs` attribute.
- Next, add a `<div>` element to the body of the `<text>` element. Give it a `@type` attribute with the value **entry**
- Add a `<head>` element at the start of the entry. Put a `<date>` element inside it. This should have the content **Nov 26 Sat.**
- You may wish to provide a normalised date such as **1938-11-26** as the value of the `@when` attribute of the `<date>`. What are the benefits of doing that?
- Close examination suggests that the letter **N** date was written over the word **To**, the first word of the next section. Add a `` element containing the word **To** if you wish.
- Insert a `<p>` element after the `<head>` to contain the text of the document.

Your document should now look like this:

```
<text>
  <body>
    <pb n="1378" facs="19381126_1378.jpg" />
    <div type="entry">
      <head><del>To</del><date when="1938-11-26">Nov 26 Sat.</date>
      <p></p>
    </div>
  </body>
</text>
```

To save you having to type too much, we've thoughtfully prepared a plain text transcription of this page in the file 19381126_1378.txt. You can cut and paste from this file, or insert the whole of it directly into your encoded version using the File -> Insert File command on the Document menu, as you did before.

Whichever method you adopt, you'll need to do something about the ampersands in the text. An XML document cannot contain an ampersand directly: it has to be represented indirectly by means of the short cut `&`. But you probably knew that already.

3 Editorial markup

Here are some interventions you may wish to mark in the text:

- There are a few places where Graves has inserted a word or a part of a word above the line, for example the letters **in** in **preliminary** in the third line. The textual transcription we prepared for you uses a slash to indicate these: you should of course remove that and replace it by the appropriate TEI element: `<add>`
- There is one place where the word **we** is substituted for the word **I**: you could tag that with `<subst>`, or simply record that the `<add>` follows a deletion
- At the start of the last line of the document there is a hard to read deletion: you can tag this with `<unclear>`

You may be wondering what to do about the way the text suddenly changes orientation. In a textual transcription like this one it seems unimportant, and so we don't suggest you do anything. You are not even recording the line breaks, after all. You may however wish to consider how you might make a documentary-style transcription, recording the lines and zones which constitute the document.

4 Semantic markup

This document is full of names! Go through it carefully marking up the names of people with `<persName>` and the names of places with `<placeName>`. You can do this quite quickly as follows:

- highlight the characters which you want to tag
- type CTRL-E to display the menu of available tags
- type 'pe' (for `<persName>`) or 'pl' (for `<placeName>`) and then press RETURN

Next you need to disambiguate the names. **rue Mirò Mesnil** for example is the same place as **Miro Mesnil**, just as **M. Priour** and **Priour** are both ways of referring to the same person. To do this, and for other reasons, you need to allocate unique codes to each unique person or place you identify. In a real life project, you would probably use special software for this purpose, but for the moment proceed as follows:

- Give each `<persName>` and each `<placeName>` *@ref* attribute. The value of this should be a pointer, starting with a # and followed by any code you care to make up. You could use sequential numbering, (`pers01`, `perso2`, `place01`, `place02` etc), or you could follow our suggestion of using two initial letters taken from the name in question, followed by two digits.
- For example, we give ‘Laura’ the code `la01` wherever she appears, and we give the two different versions of the place name ‘Miromesnil’ both the same code `mm02`

We will use these codes as identifiers for `<person>` or `<place>` elements which we will add to the TEI Header, documenting in one place the named entities found throughout the diary. For example, ‘Laura’ refers to Laura Riding, the American poet (you can read about her on wikipedia at http://en.wikipedia.org/wiki/Laura_Riding) who lived with Graves from 1929 to 1936, and ‘M Priour’ appears to have been the proprietor of a cafe in Montauban de Bretagne, where Graves and his companions were living at this time.

- Create a `<listPerson>` element inside the `<sourceDesc>` of your TEI Header
- Inside it, create `<person>` elements like these

```
<listPerson>
  <person xml:id="la01">
    <persName>Laura Riding</persName>
  </person>
  <person xml:id="pr01">
    <persName>M Priour</persName>
  </person>
</listPerson>
```

- Similarly, create a `<listPlace>` inside the `<sourceDesc>` of your TEI Header, following the `<listPerson>`
- Inside it, add `<place>` elements for the various places referred to:

```
<listPlace>
  <place xml:id="pa01">
    <placeName>Paris</placeName>
  </place>
  <place xml:id="mm01">
    <placeName>rue de Miromesnil</placeName>
  </place>
</listPlace>
```

Try to create at least minimal `<person>` and `<place>` elements for each of the different people and places whose names you have tagged. Of course, it’s up to you how much time you spend researching these named entities and transferring the information you find into TEI form! Use elements such as `<birth>`, `<death>`, `<occupation>`, `<event>` to record for example these facts about Laura Riding, which we have copied from Wikipedia:

```
Laura (Riding) Jackson
Born 16 January 1901, New York City
Died 2 September 1991 (aged 90), Wabasso, Florida, U.S.
Occupation Poet, critic, essayist, novelist
```

```
Nationality American
Spouse(s) Louis R. Gottschalk (1920–1925); Schuyler B. Jackson (1941–1968)(his
death)
```

Don't forget to save the file you have created! You will continue to work on it in the next couple of exercises.

5 And finally

If you get bored with named entities, feel free to add some more diary entries to your file. The folder `Work/Graves` contains images and basic transcriptions of a few more for you to work on....

Alternatively, you might like to experiment with a documentary style encoding for this page. You will need to add a `<sourceDoc>` containing a `<surface>`, within which you identify (at least) two `<zone>` elements...