TEI Encoding Project

In this project, you will work in pairs to encode short passages from the 1688 edition of *Oroonoko* using the [Text Encoding Initiative](#) (TEI) markup language. You'll use TEI to mark the important features of your passages, including document structures, dialogue and quotation, references to persons and places, and distinctive or emphatic language. You will also use the encoding to apply a set of interpretive categories collectively identified by the class.

Downloading Oxygen

For this project, you will need to download the Oxygen XML editor from this link:
https://www.oxygenxml.com/xml_editor/download_oxygenxml_editor.html
Select the version that matches your computer and follow the steps to download and install Oxygen. When you are prompted, paste the license key from the email circulated before class.

Interpretation Tags: this is the document to add all of the interpretation tags for our class discussion.
- https://docs.google.com/document/d/1YN52sxzVRPQWsCg1TxY8Fc1UBlB8rltE21k4EDeJNc4/edit?usp=sharing

Basic steps for encoding

- Download the exercise folder from here:
  https://www.dropbox.com/sh/nnpbrcbygpg6l2o/AADhHED-GVaL7xWzjtWmXSCIa?dl=0
- If you have a PC, you will need to "extract all" the files in the folder; if you have a Mac, just double-click the zip file to expand it and its contents.
- Launch Oxygen and hit `control-` or `command-O` to open a file. Navigate to the folder you just downloaded and open the "template.xml" file.
- If you want to rename the file, you can go to `"file"` and then `"save as"` to give it a new name (when you and your partner begin working together, you will want to rename your file with your last names in this format: "LastName_LastName.xml"). You can hit `control-` or `command-S` to save as you work.
- To practice encoding individually, there are a few commands that it is useful to know:
  - Type a `<` to insert a new element; you will get a dropdown list with all the elements that are permitted in that part of the document. You can narrow this list down by beginning to type the name of the element that you want. Just make sure to select your element from the list (by double-clicking on the element you want or hitting return with that element selected), so that you will get the right markup.

- - With your cursor inside of the element's start tag, just after the name of the element but before the closing `>` character, hit the spacebar. This will give you a dropdown list with the attributes that are allowed on that element. As with inserting elements, you can type the name of the attribute you want to insert, and you should make sure to let Oxygen fill in the markup for you.
    - If you want to surround existing text with element start and end tags, select that text and type `control-` or `command-E`; this will also give you a dropdown that will function in the same way.
- When you are ready to begin working on your project, start by filling in the missing "metadata" at the top of the document in the <teiHeader> element. Delete each comment and fill in the page numbers for your passage, as well as your names. You will also want to rename the file you will be submitting as described above.
- You should then locate the relevant section in the file to begin your encoding; this will be in either the <front> or the <body>, depending on whether you are doing the dedicatory letter or the main body of the text. The comments will tell you where to start your encoding; locate the right section, delete the comment, and then begin encoding inside of that <p> (paragraph) element. When you need a new paragraph, you can insert one as described above.
- Make sure to save often! You will also want to be careful about errors; if you get a red bar or red underlining, hit `control/command-shift-W` to check the well-formedness of your document. If your document is well-formed, you can then check validity by hitting `control/command-shift-V`. Well-formedness errors are more serious, and should be resolved right away. Validity errors should also be addressed, but they are less serious.
- As you encode, you can consult the list of elements at the end of this document; if you find something you want to encode that's not listed, try the linked element list below or even the TEI Guidelines.
- When you are ready to mark the interpretive analyses in your text, go to the end and and follow the instructions in the comments there to set up your <interp>s and give them @xml:ids. Then, you can use the @ana attribute and the <seg> element to apply those interpretations to your text. Surround the text you want to mark with <seg> and then add an @ana attribute. As long as you have set up the <interp>s, you will see their values in a dropdown when you add @ana.

Links and web materials

- **Slides for this project:**
  https://docs.google.com/presentation/d/1ien310eBx25Lzm_QcbYoxpQSS6-Td1RSgL-_jsIgGRc/edit#slide=id.g5190ea1b66_1_5
- **Slides: Introduction to XML:**
  https://www.wwp.neu.edu/outreach/seminars/_current/presentations/xml_intro/xml_newIntro_00.xhtml

- **Slides: Basics of Prose Markup:**
  https://www.wwp.neu.edu/outreach/seminars/_current/presentations/basic_encoding/basic_encoding_00.xhtml
- **Element list (see below for the specific elements you are most likely to need):**
  https://wwp.northeastern.edu/outreach/seminars/_current/handouts/elementList.xhtml
- **Oxygen cheat sheet:**
  https://github.com/NEU-DSG/wwp-public-code-share/blob/fulltext/docs/oxygen-cheat-sheet.md
- **Unicode quick guide:**
  https://dsg.neu.edu/wiki/Unicode_Quick_Guide
- **TEI Guidelines:**
  https://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html
- **Digital editions of *Oroonoko* from EEBO-TCP**
  - Chicago Textual Optics Lab:
    https://artflsrv03.uchicago.edu/philologic4/eebo/navigate/21025/
  - Oxford Text Archive:
    https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/A27305

Templates and examples

- **Exercise folder** (download this entire folder and open the "template.xml" file from Oxygen to get started; please wait until right before class to download the folder, since it is still being finalized):
  https://www.dropbox.com/sh/nnpbrcbygpg6l2o/AADhHED-GVaL7xWzjtWmXSCla?dl=0
- **Oroonoko sections:**
  https://www.dropbox.com/s/5gl2fejjfxgtv98/Oroonoko_Sections.pdf?dl=0
- **Mary Jones sample:**
  - **Document link:**
    https://www.dropbox.com/s/gdr04tsfzjtmdvt/JonesMiscellanies.pdf?dl=0
  - **Encoding:** https://www.dropbox.com/s/994jlbbcmvj07pu/jones.sample.xml?dl=0
- **Oroonoko sample:**
  - **Document link:**
    https://www.dropbox.com/s/81zpg244g9u919f/oroonoko_sample_excerpt.pdf?dl=0
  - **Encoding:**
    https://www.dropbox.com/s/mh1kz71sbl2dqkl/oroonoko_sample.xml?dl=0

Common Elements

<body>

Contains the main body of the document, not including front matter and back matter.

<date>

Used to encode dates. The `@when` attribute can be used to encode a regularized form of the date.

<distinct>

Used for linguistically distinct words (e.g. dialect words, regionally accented words).

<div>

A division of a text: for instance, an act, a chapter, a section, a poem, a letter… Use the `@type` attribute to indicate what kind of division.

<emph>

Used to encode linguistically (as opposed to just typographically) emphasized words or phrases.

<foreign>

Used for foreign-language words when no other element (e.g. `<quote>`) is already present.

<front>

Contains the front matter of the document, if any: title pages, tables of contents, introductory essays, and so forth.

<fw>

"Forme work," i.e., page numbers, signature marks, and catchwords. Use the `@type` attribute to indicate what kind of forme work.

<head>

The heading of a division: contains words and phrase-level encoding.

\<hi\>

Used to encode words or phrases which are highlighted for reasons which the encoder either does not know or chooses not to analyze.

\<interp\>

Used to establish interpretative annotations which can be linked to a span of text with the `@ana` attribute

\<lb\>

An empty element which marks a typographical line break.

\<name\>

Used to encode all kinds of names, i.e. proper nouns and noun-phrases.

\<opener\>

This element groups together the elements that appear at the start of a letter or similar document: the date and place of writing (using `<dateLine>`, and the salutation to the person being addressed (using `<salute>`).

\<orgName\>

Used to encode the names of organizations.

\<p\>

A prose paragraph: contains words and phrase-level encoding.

\<pb\>

An empty element which marks the break between one page and another.

`<persName>`

Used to encode the names of persons.

`<placeName>`

Used to encode the names of places.

`<quote>`

Used to encode quotations from other sources.

`<ref>`

A reference. Use with `@target` to link out to web resources.

`<rs>`

Used to encode all kinds of references to people, places, and things; i.e., nouns and noun phrases. If you want to distinguish between different categories of entity being referred to, you can use the `@type` attribute.

`<said>`

Passages spoken aloud or thought, e.g. by a character in a novel.

`<seg>`

General-purpose phrase-level segment: can be used with `@ana` to associate interpretive annotations with particular sections of text.

`<salute>`

Used within `<opener>` and `<closer>` to encode the salutation to the person being addressed (e.g. Dear Sir, or I remain faithfully yours…).

`<signed>`

Used within `<closer>` to encode the signature or name of the person writing.