Project Specification

Medieval Words, Modern Methods

revision of August 28, 2025

Overview

This seminar is centred around your own personal project. Over the course of the term, you will select a short handwritten text or excerpt of circa 300 words, medieval or early Modern at the latest, transcribe and encode it in TEI XML, write a brief scholarly introduction, mark up your text with a range of metadata, style it in CSS, access the document tree using Python, generate a basic statistical profile of the text, and document your methods in a Jupyter notebook. The present document specifies the requirements and should be read in conjunction with three further documents: the course syllabus, which specifies what part of the work is due when; the Project Suggestions; and the Resources document.

Scholarly Introduction

As part of your project, you will write a brief scholarly introduction of 300 words or more, encoded in the relevant parts of the TEI header. It should at least have a palaeographical component describing the text's transmission (the surviving manuscript witnesses, their date and origin, your choice of manuscript) and especially that of the chosen witness. The introduction should address at least one further scholarly question of your choosing, such as authorship, date of composition, dialect, sources, metrics, or vocabulary; this information may be gleaned from existing scholarship (which should be duly cited; see TEI Guidelines § 3.12, "Bibliographic Citations and References") along with your own research. The introduction is best written after you have gained a familiarity with the text and its scholarship, and in close conjunction with your class presentation.

Encoding Specification

Mandatory Markup

Your text or excerpt should be encoded following the TEI P5 standard on the basis of the template supplied in the course repository, with at least the following features:

- A codicological account of your manuscript in the TEI header (see TEI P5 ch. 11, "Manuscript Description");
- Such basic units of markup as `<l>` (for verse) or `<p>` (for prose; see TEI P5 ch. 3, "Elements Available in All TEI Documents");
- Scribal interventions and your own editorial emendations (`<add>`, `<del>`, `<sic>`, `<corr>`, wrapped in `<choice>` as required, alongside `<surplus>` and `<supplied>` as needed; likewise documented in ch. 3 of the TEI Guidelines).

Advanced Markup

You are expected to mark up your text with at least two of the following three additional types of metadata:

- lemmatization (`<w>` with @lemma; see TEI P5 ch. 18, "Simple Analytic Mechanisms"):
- part of speech (`<w>` with @pos; likewise documented in ch. 18 of the TEI Guidelines);

- metre (`<l>` with @met, and/or `<seg>` with @met for halflines; see TEI P5 ch. 6, "Verse").

The format of your lemma and POS references is up to you; we will discuss some of the options in class.

Additional Markup

You are encouraged, but not required, to encode such further types of information as

- named entities (`<persName>`, `<placeName>`, `<name>`);
- geolocation (`<geo>`).

Styling

Your project should include a personalized adaptation of the CSS stylesheet template supplied in the course repository, and it should render appropriately.

Evaluation

Your project should include a Jupyter notebook demonstrating at least the following features, each amply documented in a markdown cell and/or code comments:

- A normalization and tokenization pipeline
- A count of tokens
- The term-to-token ratio (TTR)
- A count of at least one metadata feature (e.g. parts of speech)
- At least one graph (e.g. frequencies of different lemmata or parts of speech)