Representing Textual Variation (2nd Draft)

Author: Hugh Cayless

There are many possible motivations for constructing a presentation of a text that takes into account the history and sources of the text. The form of a critical edition may vary greatly depending on what it is intended to show. For some editions, the motivation might be to recover or at least approximate the text as written by the author. For others, it might be to show how the same work was republished, re-edited and reinterpreted over space and time. An edition might present a single witness or many; it might take account of modern editors' attempts to emend the text, or attempt to represent as closely as possible only the evidence of copies.

What these all have in common is an interest in showing how the text as presented came to exist. The methods and procedures used towards this goal include the collation of manuscripts to assess patterns of variation, the establishment of ancestral relationships between manuscripts so that it is possible to determine lines of influence, the assessment of other evidence, such as quotations and commentary in authors who had access to copies no longer extant, and editorial conjectures based on an intimate knowledge of the language and style of the author. In printed critical editions, the critical apparatus is a set of formally-structured notes that describe the ways the editor has applied these tools to produce the text(s) of the edition.

TEI is capable of modeling any and all of the steps involved in the creation of a critical edition. It can encode witness texts, collation tables, base texts, variant readings, editorial interventions, and a variety of other components of a critical edition. A TEI critical edition might make all of the decisions that go into the publication of a traditional printed critical edition, including the establishment of a base text and the exclusion of variants or emendations deemed not to be significant; it might stop short of that point, deferring those choices to the user interface, or it might add sufficient hooks to generate a traditional-looking edition, but not itself exclude any variants. It might encode a collation table and allow the user interface to provide the means to assemble editions from it. Or it might simply present entire witnesses, upon which collation software can be run in a variety of ways.

TEI employs two methods for modeling textual variation. The first, called "parallel segmentation" relies on forking the text stream where variant readings occur and putting them (notionally) in the same position in the text. The text stream "re-joins" when the variation is complete and proceeds as a single stream. The second records variation from the base text outside the body of that text and attaches the variant readings using one of two methods: "location-referenced"—pointers into the text, and "double-endpoint-attachment"—milestone elements marking the beginning and end of the variance. Parallel segmentation is, on the whole, easier to implement for

someone working directly on the XML, and is the most popular method for modeling variation using TEI. But the standoff methods make it easier to manage cases where variants overlap or nest in complex ways.

Although its structures are named for components found in traditional printed critical apparatuses, the TEI's critical apparatus tools (app/lem/rdg) provide more functionality and are not by design intended to be used to represent printed apparatuses as they appear on the page. TEI can certainly reproduce the typographic features of a printed apparatus, but such a reproduction would be better achieved using other structures, such as specialized notes anchored to lines or parts of the main text.

Proposed changes

- ~~Change the content model of <lem> and <rdg> to permit them to include any structural level up to and including .~~
- Provide clear examples for encoding features like interpolation, omission, and transposition.
- Make it clear that the critical apparatus elements are not to be used to directly encode a printed apparatus, although the latter could certainly be re-encoded as a TEI-style apparatus.
- Clarify that <wit> is a type of inline annotation whose contents are extra-textual and meant to provide information beyond attributing a particular reading to a particular witness or bibliographic source.
- Note that if you are encoding an edition that contains an apparatus, and wish to preserve all the typographic attributes of that apparatus, rather than to convert it into a TEI textual variance model, it will be best to treat the entries as <note>s.
- ~~Add an element <secl> (seclusit—secluded) to mark text bracketed by an editor as interpolated or out of place but probably genuine, though without a secure location. This function doesn't seem to be handled anywhere by the Guidelines.~~
- Determine whether other types of apparatus (e.g. quotations of the text to be found in other works) can be represented adequately by TEI structures (like <note>) and provide examples of how to encode them.
- Changes to the definitions for app. crit elements:

  app

  Indicates a locus of variation in the text. Contains an optional lemma and usually one or more readings or notes on the relevant passage.

  l**em**

  (lemma) if used in an app, represents the reading to be printed by default in the text.

  Note: lem is most useful if the parallel segmentation method is being employed. It is likely to be redundant in a standoff apparatus, since its parent <app> will be linked to a base text.

rdg

(reading) contains a reading from a witness or a conjectured emendation from an editor. Whereas <lem> contains a preferred (or at least a default) reading, <rdg> holds a possible or otherwise noteworthy variant.

rdgGrp

[no change suggested]

wit

(witness note) a note containing information about the derivation of the preceding rdg or lem, such as a list of manuscript sigla, references to bibliography, or arguments supporting an emendation.
Note: wit provides additional information beyond the pointers to sources in @wit, @source, and @resp.

Note on att.witnessed and att.responsibiilty:

For both <lem> and <rdg> the attributes @wit, @resp, and @source should be used to indicate the derivation of the text they contain. @wit points to one or more witnesses declared in the header, @source to an identifier for a publication from which the content derives (e.g. a suggested emendation in an article), @resp to an identifier for an encoder/editor who bears responsibility for the tag or to a respStmt which makes clear the nature of the responsibility. In the absence of @wit and/or @source, @resp implies the referenced person has proposed the content as an emendation.

Notes

1. I'm sensitive to Peter Robinson's suggestions about addressing the “variant graph” format and the automated collation tooling that is available, but I’m not sure such discussion of tools belongs in the Guidelines. I suspect that parallel segmentation is actually very close, even perhaps isomorphic to variant graphs, but that’s possibly a research question rather than something that belongs here.
2. Lots of people didn’t like the preamble to the first draft (and to be fair, it was a provocation) but I think there needs to be some discussion of what the Critical Apparatus chapter is meant to enable. The preamble isn’t meant to be a direct replacement for the introduction to the Guidelines chapter, but a means to get us onto ground solid enough to provide a basis for discussion. It might eventually turn into a chapter intro, but it isn’t there yet.
3. Peter Robinson’s discussion of the origins and intent of the Critical Apparatus structures cleared a good deal of the fog in which we had been operating. I think most of the people currently working on this were under the impression that app/lem/rdg/wit had been intended to be able to represent apparatuses as they appear in print. Since this is not the case, I have substantially altered the proposal from the first draft.

SIG MEETING NOTES:

Next steps:

- Set up working plan (Dot and Hugh)
- Redraft proposal, taking above comments into account.
- Need proposals for how to implement these changes (what would we need to do to the TEI ODD to make this happen?)
- Create an ODD to make the modifications, share it widely before sharing with the TEI Council
- Get a group together in a room to write the ODD together (TEI Meeting in Lyon, perhaps one meeting before - attached to Keystone DH Conference in Philadelphia?)
    - We need to get money to pay for the meeting (request funds from TEI consortium)
    - Need to be strategic about who we include in this group (various disciplines, concerns)
    - Importance of outreach, testing

Issues:

- No way to differentiate between output of automatic collation and critical apparatus
- We are limited by the limitations of processing for standoff markup.
- Elena's suggestion: <app> element vs. <collation> element vs. <var> element, both with <rdg> and <lem> inside
- Crowdsource examples (omission, transposition, etc.); include lots of images