

Towards a New Edition of the Scholia to Euripides¹

Donald J. Mastronarde, U. of California, Berkeley

It has long been acknowledged that a new and more complete edition of the scholia to Euripides would be desirable. The older scholia on nine select plays are extant in manuscripts dating from the 11th to 13th and 14th centuries and were presumably compiled for the most part no later than the 9th century. They are represented in Eduard Schwartz's edition of 1887-1891,² which was excellent for its time, but deliberately omitted some material as too late or too trivial and is not wholly reliable in conveying the presence or absence of lemmata and the articulation of some particular notes. It is also possible that Schwartz's overall dismissal of the *recentiores* carrying old scholia is unjustified, since modern editors of the plays themselves have argued for the value of some *recentiores* scorned in the nineteenth and early twentieth centuries. Only large-scale collation of the *recentiores* will improve our understanding of this issue.

The scholia from the Palaeologan era, especially from the late 13th and early 14th centuries, cover only the triad plays, which were prominent in the Byzantine curriculum for the well educated individuals who read drama, and which were so frequently copied that we have between 100 and 300 manuscripts of these three plays earlier than 1600 (not all of these, fortunately, with scholia). Many of these notes are printed in the editions of

¹ This working paper was initially prepared for the seminar "Critical Editions in the 21st Century" organized by Cynthia Damon for the Annual Meeting of the American Philological Association in Chicago, Jan. 4, 2008. This version differs from the paper distributed Nov. 29, 2007, in that it includes some corrections and changes of mind resulting from further work on the sample XML file and discussion with others. In making my first approaches to this project, I have benefited from the advice of Greg Crane, Thomas Elliott, and Sharon Goetz. I have not yet been able to follow up on all their suggestions, and I am alone responsible for the opinions expressed in this paper.

² *Scholia in Euripidem*. Collegit, recensuit, edidit Eduardus Schwartz, Berlin 1887-1891.

scholia that preceded Schwartz, notably in W. Dindorf's edition of 1863.³ The Palaeologan scholia are principally by Manuel Moschopoulos and Thomas Magister; there are a few notes apparently ascribed to Maximus Planudes,⁴ and Demetrios Triklinios produced an annotated edition of the triad that compiled Moschopouleian and Thoman notes (and distinguished their origins) and added his own notes on metrical structure and metrical analysis. Most of these notes are in Dindorf, but his versions are not based on what we know are the best available witnesses (including Triklinios' own manuscript). The Triklinian metrical scholia, edited by Lorena de Faveri, were published in 2002 in a fairly obscure monograph series, and if your library does not already own it, it is no longer possible to obtain copies of this work.⁵

If one were doing a new conventional edition in printed form, one might plan a multivolume work that separates the old scholia from the Palaeologan, on the lines of the edition of the Aristophanes scholia of Koster, Holwerda and others.⁶ But the separation of *vetera* and *recentiora* into separate volumes is inconvenient and misleading, as it may obscure the relation of *recentiora* to *vetera*. It is better to keep the scholia together in one sequence, as in Ole Smith's unfinished edition of Aeschylus,⁷ even though a static print layout tends to leave the older elements visually lost in a mass of Palaeologan glosses and

³ *Scholia graeca in Euripidis tragoedias ex codicibus aucta et emendata*, edidit Gulielmus Dindorfius, Oxford 1863.

⁴ See the important work of Hans-Christian Günther, *The Manuscripts and the Transmission of the Palaeologan Scholia on the Euripidean Triad* (Hermes-Einzelschriften 68) Stuttgart 1995, which I reviewed in *Classical Review* 47 (1997) 23-25.

⁵ Lorena de Faveri, *Die metrischen Trikliniusscholien zur byzantinischen Trias des Euripides* (M & P Schriftenreihe für Wissenschaft und Forschung. Drama: Beiheft 18, Stuttgart 2002).

⁶ W. J. W. Koster, D. Holwerda, et al., ed., *Scholia in Aristophanem*. Groningen 1960-.

⁷ *Scholia graeca in Aeschylum quae exstant omnia*, ed. Ole Langwitz Smith, Leipzig 1976-1982.

paraphrases. A digital edition, however, offers the possibility of customizable display, at least if the digital edition is more than a simple text file. What is needed is an edition that makes use of XML (Extensible Markup Language), which has emerged as an important platform-independent standard for many kinds of data, textual and other, in modern computing and information technology.⁸ With an XML edition, the whole corpus of scholia could be one digital text capable of being displayed in all its detail, but one would have the option of color-coding different kinds of scholia being displayed simultaneously; or the display could be limited to a certain subset of the scholia, on the basis of the information incorporated in the XML tags and their attributes. Likewise, apparatus criticus, apparatus of testimonia and *loci similes*, and other notes could be displayed or not at will. And an XML edition would also be capable of being transformed into a PDF and printed as a book.

In theory, one can create any document-design one wishes in XML by declaring its structure in an associated document. Up until recently, the associated document would have been a DTD or Document Type Definition/Declaration. But the TEI (Text Encoding Initiative)⁹ has emerged as widely-used framework for digital editions and is recommended by the NEH to those applying for scholarly editions grants for digital editions.¹⁰ Originally based on SGML (Standard Generalized Markup Language), TEI has now (in 2007) reached a version 5 that is expressed in terms of XML. There are two versions, TEI Lite, a limited subset suitable for simpler texts but insufficient for an

⁸ <http://www.w3.org/XML/>

⁹ <http://www.tei-c.org/index.xml>

¹⁰ <http://www.neh.gov/grants/guidelines/editions.html> - [howto](#)

edition as complex as needed for the Euripides scholia, and the full TEI P5.¹¹ DTDs are apparently on their way out, now being replaced by documents in some dialect of a schema language that describes the components and structures of the particular TEI XML document. TEI 5 contains a vast array of possible tags and attributes, but also allows for customization where needed. Although information technologies are always evolving, it seems best for a project that will take a number of years to be completed to adhere as closely as possible to what is at present a leading-edge framework. TEI P5 is also the basis of EpiDoc, a collaborative project to define the structure of and foster the creation of a corpus of epigraphic documents in TEI XML.¹²

The TEI in its earlier versions has been used in some pioneering textual projects, some of which are discussed in the 2006 MLA publication *Electronic Textual Editing*.¹³ Many of these deal with much larger corpora or aim to produce complete transcriptions of primary documents, and they involve large teams and significant grant funding. I have no intention of producing transcriptions of manuscripts, and this project has been in the back of my mind for many years as a solo effort (more on this later). I hope to use extremely basic tools and work on my Mac, although at some point I will have to decide on the method of delivery and archiving, which will require some institutional partner.

To prepare for this working paper, I decided to make a specimen out of a short passage of *Phoenissae* (lines 202-238), since for the mss of this play I already have many microfilms and facsimiles from the days of my work on the textual tradition and the

¹¹<http://www.tei-c.org/Guidelines/>

¹²<http://epidoc.sourceforge.net/>

¹³ Lou Burnard, Katherine O'Brien O'Keefe, and John Unsworth, eds., *Electronic textual editing*, New York 2006.

Teubner edition.¹⁴ This is not something I actually intended to begin work on in 2007, but the invitation to participate in this workshop changed my plans, and what I have done so far is to be regarded as a partial proof of concept (it is also, I find, the most effective way for me to familiarize myself with a new technology). I began with Schwartz's text of the older scholia, kindly provided to me in Unicode encoding by the TLG,¹⁵ and I entered the apparent Palaeologan (Moschopoulean and Thoman) scholia from Dindorf and entered the Triklinian metrical scholia from De Faveri's edition. I then collated a few handy Palaeologan mss (not all the ones that I will end up collating for the actual edition) in order to revise and add to the items from Dindorf, and checked the Triklinian scholia against the readable faithful copy Ta (since the microfilm of T is exceptionally hard to read and I don't have access to a good microfilm reader at home and have not yet digitized any microfilms, as I probably shall do). For the Triklinian notes I also added explicit annotation of the relevant colon of Greek text as he wrote it in his autograph copy, along with a metrical scheme of longs and shorts. This gave me a quick and dirty untagged edition in a Word document, with a few notations of textual variation for the Palaeologan witnesses (that is, I have done no collating of the older scholia yet, nor did I even copy Schwartz's apparatus as a stopgap). Then I worked on tagging this same sample as XML. I have not yet dealt with every kind of tagging that may be needed, but I have gotten far enough in November and December 2007 to produce the current working paper for discussion and feedback.

¹⁴ Donald J. Mastronarde, J. M. Bremer, *The Textual Tradition of Euripides' Phoinissai* (University of California Publications: Classical Studies, vol. 27) Berkeley 1982; *Euripides. Phoenissae* [Bibliotheca scriptorum Graecorum et Romanorum Teubneriana] Leipzig 1988.

¹⁵ Thesaurus Linguae Graecae: <http://www.tlg.uci.edu/>

Before starting the tagging process, I did some preliminary reading about digital editing in the humanities, about XML and XSLT (Extensible Stylesheet Language Transformations,¹⁶ a language for transforming XML into various other forms for presentation, for instance into HTML for display within a standard browser), and about TEI P5. I soon discovered I needed to read up on the schema language as well. I am using the XML editor Oxygen,¹⁷ available at a very reasonable price for an academic license. The latest version is 9.1, and it takes account of TEI P5 and allows for very easy validation of compliance with TEI markup. Using the Roma tool (linked on the TEI web site)¹⁸ I was able to generate a schema document to attach to my sample XML document. With this tool I specified the modules of P5 that I wanted to use (in my latest effort I took the approach of removing unwanted modules from a complete set) and then generated a schema in Relax NG (Regular Language for XML Next Generation)¹⁹ and downloaded it to the same directory as my sample. With Oxygen I then associated the schema document with the sample, which produces a processing instruction in the opening of the XML document, after the document type declaration and before the <TEI> tag. Oxygen is a Java-based application, so can be a little sluggish, and the Java Virtual Machine settings need to adjusted to assign more memory so that a large complex document can be handled.²⁰ If this is already a problem with my sample, I wonder whether it will be

¹⁶ <http://www.w3.org/Style/XSL/>

¹⁷ <http://www.oxygenxml.com/>

¹⁸ <http://tei.oucs.ox.ac.uk/Roma/>

¹⁹ <http://relaxng.org/>

²⁰ This adjustment is made by editing the hidden file Oxygen/Contents/Info.plist (hidden files are revealed by control-clicking or right-clicking on the Oxygen application icon and selecting the command Show Package Contents. Near the end of the file you will find the key element VMOptions, and you may double or triple the three numbers in the string element.

possible to work effectively with a much larger document. Another minor annoyance is that the identification of errors detected in validation is on rare occasions not exact; but through experience one soon learns the types of syntax errors that one should look for if the error notice doesn't initially make sense—some arise from bad typing, and others arise from automatic features of the program that one needs to pay attention to. In the sample I have introduced only one non-TEI tag so far, `<schLemma>`, and because it is used in almost every scholion the number of errors detected in validation from a pure TEI schema constantly mounted. Finally, I took the plunge and tried defining this element in my Roma-generated Relax NG schema by copying the definition of the element `<seg>` and adjusting it slightly. I don't yet fully understand all the details of Relax NG, and no doubt the schema document will need further attention in the future. But this simple addition worked, and now the sample document validates except for one metrical symbol.

The first step I took in creating an actual XML document was to imagine a tentative form of the edition at both its macro level and its micro level. At the macro level, the scholia could be treated under TEI guidelines either as a single text, with the annotation of the nine plays in separate subdivisions within the `<body>` element of the `<text>` element of one `<TEI>` element; or as a corpus of separate texts, where the outermost element is `<teiCorpus>` and each play's annotation would fall in a separate `<TEI>` subdivision (namely in the `<body>` of the `<text>` element). For now, I am adopting the former approach. But it may be that I will need to adopt a third approach, namely creating a series of separate documents for various parts of the work, since large files can be problematic to work with both in the editor itself and when it comes to processing and display.

Every TEI document has a `<teiHeader>` and a `<text>`. The header contains various metadata about the edition, including many optional items such as title, author, and declaration of languages used. Since most of the content of the edition will be in ancient Greek, it will apparently be most convenient to define the TEI attribute “xml:lang” as “grc”—the accepted international language code for ancient Greek to 1453 (different from the code “el” used for modern Greek).²¹ This means that portions of the edition that are in English or Latin will have to have their language explicitly marked as “en” or “la” while the tags surrounding the Greek will contain no language attribute, “grc” being the default. There is no need to add the available subtag “polyton” (i.e., to use “grc-polyton”). I am assuming that when I incorporate metrical scansions using Unicode metrical symbols to explain the colon descriptions in Triklinios’ scholia (see below), the language need not be specified as “en.” Although only a few apparatus criticus notations have been created so far, the declaration and tagging involved is clear: each witness has to be identified by a unique xml:id attribute. At a later stage these could be declared somewhere in the front matter in a discussion of the witnesses used, but for now they are declared in the alternative location, within a `<sourceDesc>` element within `<fileDesc>` in `<teiHeader>`. A second subsection of `<teiHeader>` called `<encodingDesc>` currently contains a `<refsDecl>` in the form of narrative description of how the essential items in the document are identified and referred to (see below). Another part of `<encodingDesc>` is `<metDecl>`, which is used to declare the metrical symbols (`<metSym>`) that are used in the `met` attribute of notes containing metrical schemes. Oxygen accepts both U+2012 (longum) and U+0361 (combining mark placed over two shorts that result from

²¹ <http://www.iana.org/assignments/language-subtag-registry>

resolution) as values for <metSym>; but it balks at U+23D1 (breve symbol). It turned out to be illegal to use a character with this code point as the value of the element metSym. I continue to use the character in giving the value of the @met attribute of the <l> tag (for line of verse), but have deceptively declared the value of the relevant metSym element as “sh”. The only other header part so far is the <langUsage> section within the <profileDesc> tag: this declares that English, Latin, and ancient Greek are used in the document.

The content of the edition is in the <text> element that follows the <teiHeader> element. This is divided in <front>, <body>, and <back>. I have ignored <front> and <back> for now, but the former would contain the editor’s preface and other front matter, while the latter would contain at least the indexes. The <body> element contains the meat of the edition. This is subdivided into a hierarchical series of numbered division elements, <div1>, <div2>, etc. Since the goal is to capture all the annotation that has come down from late antiquity or been produced in the middle ages, the top level, <div1>, will serve not only for the sections that have the type “subdivisionByPlay” but also for the section with type “preliminaryTexts” containing the scholarly material that precedes all nine (or three) plays, the Life of Euripides and the expanded version of that life by Thomas Magister. The <div1> element for a particular play will have two parts, a <div2> of type “hypothesesis” and another <div2> of type “scholia.” With the “hypothesesis” type, the next subdivision <div3> would have one of several different types: so far I would distinguish (1) the hypothesesis ascribed to Aristophanes of Byzantium; (2) the epitomes derived from the ancient “Tales of Euripides” sometimes associated with the name of Dikaiarchos; (3) epitomes as redacted in the Moschopoulean

edition; (4) Thomas Magister's more expansive hypothesais; (5) miscellaneous prefatory material, including anonymous mythographic notes, anonymous critical judgments, poems or oracles, and the short essay περὶ τοῦ εἰδώλου that Moschopoulos placed before *Hecuba*.

The subdivisions within <div2> of type "scholia" will contain the individual scholia themselves as <div3> units. To facilitate filtering of what is or is not displayed for users with different interests, I set out to provide a rough classification of the scholia according to origin and type. As to origin, the possibilities are vetera, Planudes (though rarely), Moschopoulos, Thomas, Triklinios, and anonymous recentiora. Schwartz considered some of the scholia intermingled with the scholia vetera in B (which he dated two centuries too late) and V to be "late" (perhaps he thought they were no earlier than the 12th century whereas the "true" vetera go back at least to the 9th or 10th) and stigmatized them with an obelus, but I do not think this distinction is useful. As to type, annotation is of three general types. There are, first, simple glosses of single words or a short phrase. Some of these are intended to give a simpler equivalent for a rare item of vocabulary, others to indicate that in a particular line a common word has a special or usual meaning, and some seem to be instead part of an educational exercise of paraphrase by synonyms (akin to the practice in some Palaeologan manuscripts of adding the article in the appropriate case and gender above almost every noun to exercise skills in declension and vocabulary). Secondly, there are paraphrasing scholia, which explicate a half-line, a line, or more with reordered syntax and simpler language. These too reflect educational practice, whether the paraphrase represents how the teacher is to explicate the text or how the student is to recite to show his understanding. Thirdly, there are exegetic

notes, ranging from simple matters like identification of proper names or clarification of mythographic background to more complex matters like characterization, plot construction, staging, and discussion of textual variants or debated articulations of the text.

By combining the six origins with the three types, we could have at least 18 potential types. But I have found that in addition to a simple gloss category, it will be useful to have a separate category for two annotations that may or may not all be recorded: the “article gloss” (mentioned above, a way of indicating the gender and number of a noun) and the “eta gloss” (eta written above a Doric alpha in a lyric line to indicate the normal form of the word). And separate categories are also needed for Triklinian metrical notes and for anonymous metrical notes. In an earlier version of this paper, I wrongly said that the TEI tag <div> lacked an attribute “subtype” to go with the attribute “type.” Since subtype does exist, I now use @type for source (vet, pla, mos, tho, tri, anon) and @subtype for kind (exeg, gloss, para, metr, etaGloss, artGloss). The utility of the tagging is that one will be able to create display options to see only Thoman scholia, or only old exegetic scholia, or only short glosses, or conversely to suppress such types so that they do not crowd out the kind of note that one may really want to see.

When we get down to the micro level of the individual scholia, one of the first issues to face is how to deal with the reference system. Even without dealing with XML and TEI, any editor would have to develop a scheme. It is traditional to use the line numbers of the poetic text. Scholia that pertain to a whole passage have a reference expressed as a range: for instance, Triklinios’ first long scholion on the metrical analysis of the parodos of *Phoenissae* pertains to lines 202-238 as one astrographic stanza (because it

was not until Canter in the late 16th century that these lines were recognized as a strophe, antistrophe, and exceptionally placed epode). Scholia that pertain to an individual line have the number of that line. When items that differ significantly in content are found in different witnesses, they are treated as distinct scholia, and if they pertain to the same line or same passage, one should append a lower case letter to the line number, as is done in Drachmann's Pindar scholia²² or Erbse's *Iliad* scholia.²³ It is more problematic how to treat items that have coalesced in the tradition and are attested in the same witnesses, but are either marked as distinct by the use of ἄλλως or are more easily understood if separated from each other. Drachmann and Erbse treat a note introduced by ἄλλως as a separate item, progressing to the next letter suffix (4b, 4c, or the like), while Wilson in his scholia to *Acharnians*²⁴ treats an ἄλλως item as a subdivision of the same number and suffix, adding a further level of subdivision with (i) and (ii). For now, I prefer the former method.

It will be necessary to be able to make cross-reference to each individual scholion, and for this reason, each one will need to be assigned a unique xml:id property within its <div3> tag. These should clearly be based on the traditional line-number reference system used for editions of scholia of poetry, but it is in fact illegal in an xml:id attribute to begin the string with a number, although one can use a number with an underscore prefixed to it; moreover, the same line numbers will occur in many plays, so if all the plays are treated in the same XML document, one cannot obtain unique xml:id strings

²² *Scholia vetera in Pindari carmina*. Recensuit A.B. Drachmann, Leipzig 1903-1927.

²³ *Scholia Graeca in Homeri Iliadem (scholia vetera)*. Recensuit Hartmut Erbse, Berlin 1969-1988.

²⁴ *Scholia in Aristophanem*. pars 1:1B: schol. in Acharn. ed. N. G. Wilson, Groningen 1975.

using only the numbers. Thus I propose to build the xml:id out of the first two letters of the Latin version of the play title plus the line number that is either the only line to which the scholion refers or the first line of the passage to which it refers. To insure the correct sorting of these strings, if that is ever needed, the numbers will be provided with leading zeros so that all line numbers consist of four digits. (Simplicity of sorting order is also the reason why I use only the number of the first line covered by a scholion and not two numbers reflecting the range of lines, if more than one line is explained.) Then I will add a dot and a two-digit “decimal” suffix starting from 01. If, for example, I identify nine separate scholia that apply to the first line of the parodos of *Phoenissae* or to a passage that starts with that line, the xml:id would range from Ph0202.01 through Ph0202.09. One problem that may arise is that the number of scholia that belong to a particular line will be unstable during the course of the work on any individual play. As I collate more witnesses, I may find additional items that do not match any of the items already given an xml:id. It is not certain that if I just assign the next available suffix number (e.g., Ph0202.10) the sequence will match the order in which I would choose to have the items appear to the user. At an early stage I could revise the xml:id value for several adjacent notes to adjust for the added item. But once any part of the edition is made public and thus capable of being cited by users, it would be most unfortunate to change any xml:id values (also, such a change would also entail searching and replacing any cross-references to any xml:id that has to be changed). Does it really matter if the xml:id values are “out of order” when a new <div3> is placed within an existing sequence to contain a new item? Perhaps not. If it does matter, use of the n attribute might be of service to

record the correct ordering in a small set of adjacent items. This is a question which may not be answered until I am farther along in the work.

Each scholion item will have a core of Greek text, from one word to a long paragraph; but each item also needs to be displayed or printed with a human-readable reference number (typically in bold). If some information about the range of a note's reference were separately encoded in a custom attribute (or even in the *n* attribute), one could write XSLT instructions to create and format the human-readable number from the value of the attribute. Initially, I was concerned whether the data I entered could be recovered in a straightforward way if at some point it appears that an XML edition will not work, so I preferred to have the line number reference as well as the text of the scholion available outside the tags. This is a needless fear based on my long familiarity with conventional texts, and in a revision I decided to show confidence in the technology and aim for a more concise result. It is most economical if the line number is entered just once, as an attribute value rather than as content, so in the revised XML file I turned these into @*n* attributes of the <div3> tag. Likewise, each item of scholiastic annotation will be followed by a list of sigla of the attesting witnesses, as is traditional in modern editions of scholia. Again, my first inclination is that I would like this information to be expressed outside the tags, as content. Yet, if the sigla were all simple, this information too could be most economically expressed as an attribute of the <div3>. Since the sigla are not simple, however, it seems less likely that I will take this route.

If both the number and the list of witnesses are expressed as content, then a normal scholion will have a tripartite structure: number in bold, Greek text, sigla (mostly in Roman letters). I briefly considered using the <head> and <trailer> elements, available

in any division, but I think this would be an unwarranted stretching of the usage, and might lead to formatting difficulties. I started out by treating the reference number, Greek text, and list of sigla as part of a single paragraph (or in a multi-paragraph scholion, the reference number is part of the first paragraph and the sigla are the end of the last paragraph). Reference number and sigla were at first both set off within the <seg> tag, used for any segment of interest to the editor; the type attribute was then used to distinguish the different uses (e.g., “schNo” and “witnesses”). In revision, however, the <seg> for the line number has been eliminated.

The next problem to consider is the lemma. Even in a printed edition, one has to decide how to deal with the supplying of lemmata. Erbse, for instance, supplies a lemma for every single item in his *Iliad* scholia. If a lemma is used in even one of the attesting manuscripts, then the lemma is printed in expanded type and separated from the annotation itself by a dicolon. If no witness has a lemma, Erbse supplies one, and both the supplied lemma and the dicolon are enclosed in the angle brackets used by editors to indicate editorial additions (the originator of the supplement is noted in the apparatus). Erbse adds a lemma even when the annotation applies to a pair of lines. For the Pindar scholia, Drachmann prints a lemma (also in expanded type and followed by a dicolon) only if it appears in at least one of the witnesses (omission by other witnesses is noted in the app. crit.); otherwise there is no lemma. In the scholia to *Acharnians*, Wilson shows a lemma in expanded type with dicolon if the lemma is in the manuscripts. If the manuscripts have no lemma and he feels one is needed, then he prints a lemma (in type with regular spacing) followed by a closing square bracket (the same usage is found, e.g., in Smith’s Aeschylus scholia). But for many annotations, Wilson allows the line number

reference to speak for itself, and no lemma is added. I can see arguments in favor of all these approaches, but on the whole would prefer in display or printing the explicit use of angle brackets for supplied lemmata and the consistent use of the dicolon to separate the lemma from the annotation. On the other hand, I would prefer to follow Drachmann and Wilson in allowing some items to be without a lemma: my inclination would be to supply a lemma for every annotation that applies to a single word or a short phrase, but not to do so when an annotation applies to a whole line or several lines. Such flexibility may, however, be inappropriate in a digital text (since there is no concern about how much paper we are consuming). On the other hand, adding lemmata would hardly be necessary if each scholion could be equipped with a pointer to the exact word or phrase in an online edition of Euripides using a canonical reference. That will certainly be possible some day, but this is not a feature I am going to worry about in the early stages.

As for tagging the lemmata, TEI P5 does not provide a tag for this purpose: the TEI <lem> element is something distinct, meant for the lemma (= accepted reading) cited within a critical apparatus notation <app>, and I think it would be better not to redefine the same tag to allow its use for the lemmata of scholia. One could use a <seg> element with a type “schLemma,” but I have preferred to create a new element instead, since TEI is extensible. As I mentioned earlier, I successfully defined <schLemma> in my schema document. I give this new element an attribute “source” with the possible values “mss” (for a lemma found in manuscript sources) or “sign” (for a lemma implied by a reference symbol or numeral present both next to the lemma-word and next to the annotation) or “ed” (for a lemma supplied by the editor). One then needs to decide whether to make the angle brackets explicit in content when the lemma is an editorial addition or is implicit

because based on a sign or whether to rely on a processing instruction to enclose a lemma with this source attribute within angle brackets. Again, at first I was too hesitant and I placed angle brackets in content, again from my perhaps superstitious attachment to old-fashioned human readability. In revision, I decided they should be omitted and supplied by the XSLT processing from the source attribute. When the brackets were still present in the sample, I, unlike Erbse, chose to leave the dicolon outside the closing angle bracket, because I do not intend to be very detailed in collating for punctuation or annotating punctuation choices in the apparatus (except for the few cases where the choice makes a real difference). In the revision, the dicolon too is omitted and left to be displayed through a processing instruction. By the way, the angle brackets used in the initial sample were not the less than and greater than symbols (U+003C and U+003E) that are used to enclose XML tags, but the mathematical angle brackets (U+27E8, U+27E9) recommended by TLG for this editorial function and included in GreekKeys fonts, but so far not in standard system fonts. There was thus no need to use the entities < and > (as one would have to do with U+003C and U+003E in an XML document), and the content gained thereby in human readability. But now that the brackets are left to the processing, this little detail no longer applies. Since <schLemma> has been defined without difficulty, one might consider whether the witness list could also be defined, rather than treated as a <seg> element with a particular type. This may well be advantageous, especially if I decide to use <seg> instead to deal with the problem of modified sigla (discussed below).

In the item sch. Ph. 202a, the scholiast makes a loose reference to a passage later in the text with “at any rate this is what the chorus tells Polyneices.” The reference is to

lines 280-285, and it is usual for the editors of scholia to add such a reference in parentheses. I am not sure whether the <note> tag is appropriate for this, since this seems to be intended for notes in the margin, footnotes, endnotes, and the like; also, since I much prefer the inline placement of such references to the alternative of display of the references in the secondary apparatus, using <note> would involve the need of some anchor for proper processing. Again the <seg> tag seems flexible enough for this, so I initially used this with the type specified as “refText.” This allowed the same content to be included without additional detail in markup for anchoring. Later, I was advised that the <bibl> tag by itself could be used for such a reference. Editors of printed editions seem to use the minimum reference possible, and so here might use just the line numbers; in the digital edition it is best to be explicit and give both the author’s name and the title along with the line reference. These can be enclosed with the <bibl> tag.²⁵ To distinguish these from <bibl> items that refer to modern publications, I give the @type “internal” when the reference is to lines within the same play, and “anc” for references to ancient texts. Then a type such as “mod” could be used elsewhere in the document when modern bibliography is cited. Inside the <bibl> tag the details can be expressed with the optional <author> and <title> elements; I do not initially find a tag to apply to the line numbers of the reference, but have been told that the <ref> tag is appropriate. At some point it would be desirable to provide these citations with canonical references to allow pointing to online instances of the cited texts.

²⁵ At the final stage of writing this paper I have benefited from conversation with Dr. Sharon Goetz of the Mark Twain Papers & Project at The Bancroft Library at UC Berkeley (<http://bancroft.berkeley.edu/MTP/>). She has a long history of involvement with SGML and TEI and has worked extensively with TEI P4 for the Mark Twain Papers as well as for other editions. See The Mark Twain Project Online, <http://www.marktwainproject.org/> .

One area of flexibility opened up by a digital edition is the possibility of adding translations of the more interesting exegetic scholia. It would not be a worthwhile expenditure of effort to translate every gloss or paraphrase, but some of the exegetic notes would be useful to those interested in literary and rhetorical theory, ancient literary criticism, the ancient genre of problems and solutions, and similar topics, and the availability of translations would assist those who want to get used to reading Greek scholia. It seems odd to me there is no tag for a translation of a given passage into another language already provided in TEI (unless I have missed something). (In dictionaries, <term> and <gloss> are used, so one might consider stretching these elements and treating the Greek paragraph as the <term> and the translation as the <gloss>.) So perhaps one should consider extending TEI with such a tag. I considered adding, optionally, an additional paragraph within the <div3> after the paragraph containing the scholion in Greek: the content of this paragraph would have been a segment identified as of type "transl" and with its language declared as English (since the paragraph itself cannot have a type attribute; or perhaps each such paragraph could have an xml:id beginning with "t" so that processing instructions could have distinguished the translations from the originals). In the end, however, I decided to use the <note> tag, with type "transl"; it should be easy to process the inclusion or exclusion of translations in a display. I have been advised, however, that if <note> is used, there should be an explicit pointer or anchor for the note to the relevant text, because without it problems might arise in processing (such pointers are not yet in the sample).

The TEI gives considerable flexibility as to how to deal with the apparatus criticus, yet the included tags do not seem quite adequate for the conventions usually

followed in an edition of a classical author. First, it is hard to decide which placement of the apparatus element to adopt without some experimentation. The <app> element can be placed in line directly after the relevant word, but since the extent of most scholia is not very great, it may improve the legibility of the XML document to use the alternative inline placement at the end of the relevant scholion. There is no canonical line numbering for the scholia to make a reference by line number feasible. As a matter of its standard practice, the TLG encodes the line structure of Schwartz's edition, but this is not the sort of line-reference system that can be used in my proposed edition. To deal with reference in a longer note, one could consider dividing up into sentences, each provided with the <s> tag and the number attribute n="1" etc. Then, if a number attribute is also supplied for the full note, in the <div3> tag, one could use the location (loc) attribute in the <app> tag to refer precisely to the relevant sentence: if the <div3> has n="202b" and the sentence has n="3", then loc="202b 3" would provide the tie-in in the <app> tag. But if one adopts this method for references in the apparatus within longer notes, this does not imply that one has to do the same for every single item in the apparatus.

A second issue is that an expert apparatus criticus in a printed Greek edition takes advantage of a long tradition of editorial practice about the modification of sigla to indicate hands, corrections, and doubts of various kinds, and in editing scholia one also wishes to record location of the note: in Schwartz, for instance, a plain siglum means that the note is found in the body of scholia surrounding the text (this body may be in some combination of upper, side, and lower margins); a gloss above a word in the text is marked with a superscript g; a note or gloss in the margin between the text and the block of scholia has a superscript i (for *intermarginalia*). More recent editors use superscripts

like s and marg for the same purposes, and in some editions we find superscript uv for “ut videtur,” or a dot under a symbol for a manuscript group to indicate “all those in the group except for the one or two cited for a different reading at this point.” As mentioned earlier, each manuscript source needs to be listed in the front matter or the header of the TEI document in an element called <listWit>: this element is given an xml:id that is the siglum of the manuscript, and that xml:id is referred to when a witness is cited in an apparatus entry. There is a provision for recording different hands as an attribute (@hand) of the lemma reading or variant reading, but no provision for noting a supralinear correction or gloss or *graphetai*-notation or a marginal correction, which classical editors would do with symbols like A^s A^γ or A^{marg}. (I did not think that the element “witDetail” as described in P5 would serve the purpose, but this may bear further consideration.) Perhaps it will be necessary to extend the TEI schema by adding a modifier attribute. For the moment, one can stick to the TEI schema by defining a different <witness> in the <listWit> for A^s or A^{marg} as opposed to A (e.g., with xml:id’s of “A” and “A_s” and “A_marg”), but this is very cumbersome. There is a further problem for the list of witnesses which is to be placed at the end of each scholion. An XML file cannot contain anything but plain text, so to produce an entry like Xa^sXb^sT^s one needs either to adopt an alternative expression like Xa_sXb_sT_s or to use the cumbersome expansion “Xa<seg type=“witMod”>s</seg>Xb<seg type=“witMod”>s</seg> T<seg type=“witMod”>s</seg>” (I omit ‘rend= “superscript”’ in the seg tag because this can be taken care of in XSLT processing.) After some reflection, I arrived at a somewhat simpler form of recording the attesting manuscripts and their details. This is not yet done in the sample file, but one could define an element <attestList> to use instead of <seg

type="witnesses"> and then to express X^s by <seg type="witMod" mod="s">, which presupposes the declaration of an @mod attribute and the representation as superscript through a processing instruction. Thus we get <seg type="witMod" mod="s"> X </seg>.

There will sometimes be doubts about the certainty of a reading. Water damage and fading of ink on parchment or paper can make words nearly illegible, or if the editor relies on an image and the image is not clear enough to discern a tiny detail satisfactorily, the reading may be uncertain. For instance, on a black and white microfilm the purple ink used by some scribes for the *personarum notae* and some glosses and notes will be fainter than the text written in black or brown ink. TEI allows notation of uncertainty, and EpiDoc has clarified the usage for ancient inscriptions, but what they describe is suited to the dotted letters of epigraphic and papyrological usage. This is a problem I have not really dealt with yet in the current sample.

It is a common occurrence that Doric alphas in lyric passages of tragedy may have a supralinear eta added, often by the first hand. This may be intended to record a variant, such as might be found in another witness in which the form in the line has been trivialized to its Attic vocalism, but is more often probably a teaching aid, reminding the reader and student of the more standard form of the same word. In line 202 of *Phoenissae*, for example, ἔβαν has eta above the alpha in some Palaeologan witnesses. I am not sure yet whether I will collate all these explanatory etas, or collate only for a sample of text in each play; nor, even if I collate them, am I sure I will include them in the edition. But since the edition can be open-ended, it may be useful to consider how these items would be treated. As suggested earlier, they can be given their own subtype (etaGloss) so that they can be easily suppressed from display as a class. The eta is an

abbreviated way to suggest the full word with the one letter substituted. Abbreviations can be most fully documented with a <choice> element containing both the abbreviated and the expanded version of the word, but it will be somewhat more economical to use just the <expan> tag with the rest of the word tagged within it as <ex> for editorial expansion and the eta itself untagged. So the eta over ἔβαν gets recorded as "`<expan><ex>ἔβ</ex>η<ex>ν</ex></expan>`" and in processing one can produce some distinctive display.

The Triklinian metrical analyses require special treatment. In De Faveri's edition of the Triklinian scholia on Euripides, the scholia are edited in one body, and then in an appendix the lyrical passages are printed out as they appear in Triklinios' manuscript and accompanied by his internal numbering (he always tells how many cola are in a stanza and then gives a description of each colon as "first," "second," and so on using the Greek alphabetic numerals). This entails an inconvenient flipping back and forth from description to the text it applies to, and in addition, any user not familiar with the technical terminology may sometimes be at a loss as to how Triklinios interpreted a particular line. What is needed, I believe, is a closer integration of the text of the colon and the metrical scheme in longs and shorts with the sentence within the scholion that says "colon 7 is ionic a minore dimeter acatalectic, consisting of fourth paeon and third paeon, or alternatively of <fourth paeon and> ionic because of the final anceps." The XML edition makes this easy. For the moment, I propose to add a <note> after each such sentence descriptive of a colon. The content of the note will be a line number (does this number have to be tagged as a reference, or is that overkill?) and then the quotation of the line being analyzed, which will be enclosed in the <l> tag (for line of poetry). Using this

tag affords the opportunity to use the met attribute within it, the value of which will be the sequence of Unicode symbols specified for metrical analysis by the TLG and available in GreekKeys Unicode fonts and inputs.

Sometimes Triklinios writes a phrase like “the sixteenth colon is similar to the third.” When this happens, it will be useful to have a pointer back to the description of the third colon. For ease of reading, I am making each colon description a separate <p>, and when there is need of a pointer I am adding an xml:id attribute to the tag whose value will be “m” (for metrical) plus the two letter abbreviation for the play and the line number of the colon. So *Phoen.* 204, the third colon, has xml:id="mPh204" and then at *Phoen.* 216, colon 14, when the third colon is referred to, a pointer is added in the form <ptr target="#mPh204" />. (If it is indeed prudent to anchor each note to its associated line, as mentioned above in reference to the notes containing translations, then the xml:id attribute should be consistently specified.)

One advantage I foresee for an XML edition is the ability to tag content in some useful ways. For instance, suppose I want to identify scholia that are examples of the *problêma* genre of commentary, of which the scholion Ph. 208b is an example: it begins with a question διαπορεῖται πῶς... and offers a number of possible explanations. Or Ph. 202-260c presents a pattern of criticism (ἔδει δὲ, φασὶν,...) and defense that is fairly common. Content-analysis of this kind is provided in TEI syntax by the use of or <interp>. Use of either of these raises the question again whether every paragraph ought to have its own xml:id assigned and every sentence within each paragraph ought to be tagged as a sentence <s> and given its xml:id as well. But it is apparently appropriate to assign such ids only to the specific places to which one wants to give a content analysis.

In the sample I have used <interp> rather than because it allows one to use the same analytic categories in a number of different locations. Using <interpGrp> I have declared (at the beginning of <body>) four categories of <interp> with xml:id's of “problema” and “solution” and “criticism” and “defense”. I then surround a relevant section of text with a segment tag, which can be given an ana attribute with a value like “#problema”. So sch. Ph. 208b has a problema segment and four different solution segments. Sch. Ph. 202-260c-d (an old note and a Thoman rephrasing and expansion of it) both contain a segment of criticism and a segment of defense.

It would also be desirable to tag and index technical terms (like stasimon and parodos in sch. Ph. 202-260b) and grammatical terms (like ἀντίπρωσις) and words that have peculiar scholiastic meanings (like παρόλκει). The simplest form of tagging is to mark a word as a term with <term>. Although a digital text should be easily searchable, so that a scholar well versed in scholia could simply search for terms of interest, it is more user-friendly and educational for beginners if a traditional index is created, so that by perusing the index one can learn something about the interests and vocabulary of the commentators (Schwartz has good index for several purposes, though the arrangement of the material is idiosyncratic). Tagging for indexes is more cumbersome. The <index> element does not surround the word of the text, but is placed adjacent to it, and encloses a <term> element that contains the form of the word as it would appear in the index. An indexName attribute of <index> can be used to specify which index the term belongs to, if there is more than one index. For instance, in the same sch. Ph. 202-260b, I have marked up the word ὑπόθεσιν because this term has been much discussed (most

thoroughly in Meijering's book).²⁶ After the word I placed `<index indexName="gkwds"><term>ὕπóθησις</term></index>`, which indicates that in the index of Greek words the nominative form will appear with a reference to this passage. In the same note, after the quotation of *Orestes* 140, I entered `<index indexName="locorum"><term>Eur.</term> <index><term>Or.</term> <index><term>140</term></index></index>`: by nesting the `<index>` elements in this way, we provide an index *locorum* entry that has an entry with Eur. as the head word, then a subsection for Or., and line 140 as one of the lines (or the only line) cited from *Orestes*. For proper names deserving to be indexed, one could use the `<name>` element, but because one will need a standard form of the name (the nominative) for the index, one would need either to combine `<name>` with `<index>` (appending the `<index>` element after the name content but before the closing tag `</name>`) or to declare every name somewhere with a `<person>` element containing the standard form, and pointing to this `<person>` with the `ref` attribute of `<name>`. Or one can simply dispense with `<name>` and `<person>` and use only `<index>`, at the expense of repetitious entry of the `<term>` element when the same name appears often. The sample does not yet contain such markup, except for one instance of the last method.

In this specimen, I have not yet dealt with the second apparatus of *loci similes* and *testimonia*. It may be that the needs for reference within this second apparatus will be again call for at least some explicit tagging of individual sentences within each scholion.

Having done this much work on a specimen, under a shorter deadline than I would ideally have chosen, how does the prospect of this work compare to what I thought

²⁶ Roos Meijering, *Literary and Rhetorical Theories in the Greek Scholia*, Groningen 1987.

before I undertook it? First, I don't think the digital age makes much difference to the hard work of collation of witnesses. I think collations will still need to be recorded by hand on paper and only secondarily entered onto a computer.²⁷ Second, an XML edition involves a lot of overhead compared to previous, already complex processes I have handled in the past, like writing and editing my *Phoenissae* commentary on computer as Word files. Much of that overhead can be reduced by automatic processes, especially after the editor has overcome the initial investment of learning new software and new work habits, and there ought to be a payoff in greater convenience and more rapid production at the end-stages of the work. Third, I am less sure than I was before that I can do without a specific institutional partner, although this point will be clearer only after I make some trials of writing XSLT instructions and see how that goes with a sample XML file following the schema I have provisionally adopted.

From the presentations of others at the seminar and the discussion, at least two other important questions arise. First, is direct TEI XML markup the way to proceed, or will there be a software tool like what Mark Schiefsky demonstrated in his paper that will make the editor's task more convenient? Or would a database approach to the individual scholia be better than a text-file approach? Second, the sample is built on the idea of putting all the information in one file (or at least one file for each play's scholia). A different approach, namely, tagging in a minimal way in the main document of the scholia and providing various kinds of annotation (metrical, translation, content analysis,

²⁷ Certainly, in the quick-and-dirty collating I did straight to a computer file I found after the fact that I had made mistakes or omissions that the lack of a paper trail rendered quite troublesome. But in discussion at the seminar, it was suggested that with a tool specifically designed for collation and providing for version control there would be an adequate trail of previous work and (e.g.) accidental omission of the manuscript symbol for a reading could not occur.

apparatus criticus, etc.) in one or more separate XML documents, offers two large advantages. The text itself is simpler and more human-readable and can be relatively stable over time, and the annotations likewise will be easier to work with and can be expanded gradually and subject to further annotation by others.