

The World Wide Web as we know it today is a marvelous tool for making texts, pictures, music, and other documents available, and for encouraging collaboration among scholars in various places and various disciplines. But the Web as we know it will not last forever. Projects tied to the most primitive technologies of the web will have to convert, adapt, or evolve their documents to keep up with technological change. Just as many texts were lost in the period of conversion from codex to volumen in the 4th and 5th centuries AD, so now we risk losing electronic texts in periods of even more radical format changes. In this session, we will discuss technologies built for adaptation, and some of the benefits classicists can gain from them.

The organizer will introduce the session and the speakers, and will distribute a bibliography on text encoding and electronic publishing. Questions and comments will be solicited after each paper, and the session will close with general discussion.

The native language of the Web is HTML, but HTML encodes the appearance of a document, not its underlying structure. Elli Mylonas, Speaker #1, will explain structured markup with XML. XML can encode the semantics of a document, and a text in XML can easily be published in any other desired form. At first, XML may appear more difficult than HTML, but in fact the ability to say exactly what you mean, and the separation of content from presentation, are both liberating for writers.

One well-known problem with the Web as it exists is the difficulty of presenting Greek. Classicists have devised many rough-and-ready solutions, none of them ideal. Debbie Anderson, as Speaker #2, will present Unicode, a technology that will allow not only Greek, but Sanskrit, Arabic, and other languages to be displayed gracefully. This speaker will also discuss how the classics community is working with the Unicode Consortium to ensure that the characters and symbols we need will not be left out.

Once you have well-structured texts, you can do far more with them than simply print them out or display them as ordinary web pages. Speaker #3, Jeff Rydberg-Cox, will explain the use of structured texts in the construction of a new intermediate lexicon of ancient Greek, one of the most complex and intellectually demanding tasks in philology.

With each new technology and each new communications medium come new texts and editions that take full advantage of the new medium's capabilities. These texts seem so tied to their physical form that it is hard to conceive of adapting them to a new form. Ken Haynes, Speaker #4, will consider some early Bible editions that exploited the possibilities of the then-new printing press. Electronic versions of these editions must not lose the information encoded in the arrangement of the text on the printed page.

In the final paper, Speaker #5, John Thomas, will discuss peer review in the context of the MERLOT system.

Classics as a discipline has had electronic versions of many of its primary texts for almost 30 years. Classical texts have moved from ASCII representations to HTML files, and as digital documents, have been concorded, analyzed, typeset, and disseminated on the WWW. Unlike scholars in other disciplines, classicists today assume that they have access to electronic corpora and multimedia data collections for their research and teaching.

There are also ever more electronic resources being created: corpora of papyri, inscriptions, prosopographies and collaborative translation projects are all underway. Much of this material is either plain text with minimal formatting information in it (the older materials) and varying types of HTML. Less is in SGML or XML, and much of what is doesn't expose encoding information to its users at all.

Digital text, however lacking in markup, is an empowering tool in and of itself. Scholars and students can search it successfully for strings or words, make concordance lists, and process the documents in various ways. Most importantly, the existence of large, manipulable corpora of classical material allows scholars to easily treat the texts they study as part of a collection, and not as single instances.

The widespread adoption of HTML markup and the HTTP protocol, that is, the existence of the WWW, have greatly improved many aspects of electronic primary text in classics and in other disciplines. Digital documents can be presented in a more attractive and navigable form, and can be enhanced with links and images easily, and by individuals. The example of today's WWW is only the beginning, however. With more sophisticated markup and the technologies for processing it, it is possible not only to publish documents better, faster and for more purposes, but also to move into the promised areas of collaborative work and customized environments.

New WWW standards like XML and its associated linking specifications, XLink and XPointer, the Open eBook specification for marking up documents that will be compiled into eBooks, and the Text Encoding Initiative DTD are more complicated than HTML, and lack the user friendly tools that make HTML so easy to produce. The wider adoption and understanding of these standards and guidelines, however, will make online materials easier to maintain and more versatile.

Sophisticated markup is not easy for an individual to implement on their own. The difficulty, however, lies as much as in the intellectual sophistication the task requires as in the technological hurdles that must be overcome to configure software. Where an editor like MS Word or GoLive encourages its users to make a document *look* good, and the XML encoding process insists that its users understand the structure of the document and its *editorial* questions. Formatting features express meaning. A WYSIWIG system lets the user make assumptions about the meaning. An XML system asks the user to disambiguate it.

A text with even minimal XML markup, if it uses a common DTD, can be processed by more than one application. It can be published on the web, searched as part of a corpus, and anthologized, or reformatted appropriately so it can be read on a handheld device. It can be studied using both sophisticated word searched and analyses based on the structure of the document. Finally, it can be moved to newer platforms as they appear.

A community that has become familiar with the power and shortcoming of HTML can naturally progress to using XML in order to get more benefit from their documents.

Getting Ancient Greek to Appear Correctly in Electronic Text Documents: The Unicode Solution

Debbie ANDERSON

Typing, displaying, and printing ancient Greek (or any ancient or modern script) is often viewed as just a matter of finding and using the correct font. Now that the Internet is becoming a primary means of communication, getting Greek characters to appear correctly — whether in email messages, within databases, on maps, or when publishing a text on the Web — has become increasingly complex. Differences in computer platforms, software, even varying national standards, all contribute to make handling Greek a challenge to the student and scholar in Classics.

Various methods have been used to overcome this problem, and three are widespread on the Web. One option is to use the Beta Code transliteration system for Greek, which uses the “standard” characters on the keyboard (ASCII). Another alternative is to use images of the needed letters. Lastly, one can rely on proprietary software, such as Adobe’s PDF, to embed the needed characters in a document. While these methods appear to answer the problem, they don’t, since they cannot fulfill two critical requirements for electronic texts: the need to have the characters appear as Greek, with all diacritic marks in place (hence, the Beta Code fails), and the need to search for Greek characters (not possible for an image such as a GIF, and not possible for some characters in a PDF).

However, a different solution to this problem has appeared: the international character code standard, Unicode. Unicode assigns a unique number to each character, which remains the same on any type of computer, operating system, software, or font. Because it is an international standard, major computer companies worldwide are producing Unicode-compliant products, which allow one to type, print, search, and display Unicode-supported characters. Hence, with Unicode, no transliteration scheme is needed since the original character can be used and the characters are searchable.

One significant benefit of Unicode is that it, along with other standards (such as XML), will help to open up the publication of Greek materials beyond the standard publishing houses, hopefully lowering prices significantly, speeding up publication, and making materials more widely available. Small departments will be able to publish Greek materials and make them available to anyone with an Internet connection.

As a means of testing out Unicode, a project was devised at UC Berkeley (in conjunction with the UC Library, with seed-funding from Bryn Mawr reviews) to create an online version of a small publication, the UCLA Indo-European Studies Bulletin, which contains Greek, as well as other ancient Indo-European languages. Results from this project reveal that certain specialized characters were missing from the Unicode standard. The project further demonstrated that Unicode-enabled products need to be used (operating system, browser, fonts, etc.), but not all available products offer full support yet. Currently work is progressing on Unicode proposals for the missing characters.

This paper concludes with several recommendations: (a) Unicode-enabled products should be used by Classicists, since the adoption of Unicode offers the best — and only good — solution for getting Greek and other scripts to appear on the Web; (b) interested parties should provide feedback to Unicode if missing characters are found; (c) scholars, probably under the aegis of scholarly societies such as the APA, should provide feedback to companies when inadequacies are found (in fonts, operating systems, etc.), so the needs of scholars are met. In this way, ancient Greek, as well as other historic languages, will be able to take full advantage of the Internet in the twenty-first century.

Lexicography is an intellectually demanding and detail oriented task that requires careful consideration of all of the contexts in which a word appears. Before lexicographers can begin their study of a word, they must first gather their raw materials — the texts where these words appear. Even for a finite corpus such as the New Testament or a single author such as Lysias, finding all of the examples of a word can be a daunting task. Certainly the task has been made easier by the creation of electronic text corpora such as the Perseus Project or the Thesaurus Linguae Graecae, but the lexicographer still faces the laborious task of executing searches and compiling the results in a useful fashion. If these basic tasks can be automated, lexicographers could spend more of their time doing the intellectual work necessary to write the definitions. In this paper, I will describe some of the work that is currently underway to leverage computational techniques to help in the writing of a new intermediate Greek lexicon.

Computational lexicography is an active area of research and a well developed field. Since 1984, at least two English language dictionaries have appeared that were constructed almost entirely with computational techniques, the Cobuild English Dictionary and the WordNet Lexical Database. The impact of computational techniques on the task of lexicography is only increased by the practical reality that computers have doubled in speed and halved in price almost every eighteen months. Thus, it is now possible to think of performing tasks that even five years ago required access to a supercomputer. Given the easy availability of processing power, the question becomes how exactly computational techniques can aid in the construction of a lexicon of Ancient Greek.

At its simplest level, the computer can automate the tasks of identifying the words in a corpus, building a concordance, and presenting them to lexicographers to write definitions. Using the Perseus morphological analyzer, we can scan every word in a selected corpus, determine the possible headwords from which each word might be derived, note the passage in which it appears, tally of word frequencies, and then present a list with headwords, frequency data, citations of every lexical form, and extracts from both the Greek and the English texts on a single page. All of these tasks are possible by hand given enough people and enough time. With computational techniques, however, the time savings are enormous. For example, we would be able to create this sort of database for all of the words on the TLG CD ROM in a matter of hours, gathering lexicographic slips in less than a day for a corpus that is at least as large as the one that took Murray and his volunteers many years to complete for the *Oxford English Dictionary*.

Computational techniques are not, however, limited to speeding up tasks that are possible by hand with an unlimited amount of time and people. It is also possible to gather information that simply could not be gathered on a large scale without computational techniques such as the determination of word collocation or co-occurrence patterns, words with similar definitions, words in the same semantic range, words with the same stem, and perhaps the initial division of the raw materials in the slips into initial sense classifications. Computational linguists have demonstrated that each of these tasks is possible for English language texts. The challenge we face is adapting these algorithms for texts written in Ancient Greek so that they can be used for philological research and serve as the raw materials for a new Greek lexicon.

The relations between medium and message have been more finely studied since the days of McLuhan. Anthropologists have pointed to literacy as a precondition for incremental commentary and therefore for the processes of social and cultural rationalization; that is, a written text makes it possible to identify and resolve contradictions via higher-order explanations. Scholia, mishnah, exegesis in general, are possible only when texts are relatively fixed. The mechanical printing of texts had similar radical implications. By increasing exponentially the density of information that was accessible, it has been argued, it played a crucial role in phenomena as diverse as the Protestant Reformation and the rise of Western science. In this paper, I wish to focus on two sixteenth-century printed texts, by focusing first on the implications of their printed condition and then by raising questions of the translation of these texts from print to electronic form.

I would like to direct our attention to the ‘Geneva’ Bible of 1560 and the ‘Rheims’ New Testament of 1582. Both were products of English exiles: Protestants in Geneva and Catholics in Rheims. Both groups regarded themselves as persecuted and in their translations were offering encouragement to the faithful and the backsliders. Both groups appealed to the same key words, ‘zeal,’ ‘diligence,’ and ‘patience,’ with roughly the same meaning, despite the fact that in these words would subsequently be satirized as Calvinist. Both assented deeply to some of the same theological doctrines (especially the doctrine of original sin).

The profound differences between the translations are located in their annotations and in their visual presentation. The ‘Geneva’ Bible might best be called hortatory and the ‘Rheims’ New Testament disciplinary. It was a central concern of the Protestants that the Biblical text be presented as self-evidently true. Therefore, their annotations favored a kind of self-validating grammar (“thus [Daniel] spake, being moued by the Spirit of God”) and the presentation of the text emphasized a clear visual hierarchy (Biblical text first, then annotation, then running heads) whose purpose was further to indicate the clarity and self-evidence of God’s purpose; the running heads played a particularly important role in this presentation. The Catholic New Testament, in contrast, constantly indicated to the reader that he was in danger of erring and that safety lay in appealing to the best authorities. Visually, the triple set of annotations enmeshed the reader in a web of linguistic, historical, and theological disputation; the language of the Bible favored a distantiating Latin (“Giue us this day our supersubstantial bread”); and the language of the annotations emphasized pastoral discipline (“Here the Apostle staieth the rashness and the presumption of such poore wormes”).

The degree to which these texts are embedded in their physical form requires that we attend to them carefully in producing subsequent editions, either print or electronic. It is especially with electronic texts that we are obliged to analyze structural features of printed texts and therefore must distinguish between the hierarchy or web of the original.

Introducing MERLOT: Peer Review and Collaboration for Online Teaching and Learning

John THOMAS

This paper will introduce the online organization called “M.E.R.L.O.T.” (Multimedia Education Resource for Learning and Online Teaching). Merlot is an organization of college and university professors in the U.S. and Canada. Sponsoring universities span the country, from the University of Hawaii to SUNY, from the University of Wisconsin System to the Louisiana Board of Regents. Merlot, which itself was featured in the October 2000 issue of *Syllabus*, is dedicated to providing peer review for online educational resources in a wide variety of disciplines, including those in Classical Studies. This organization is attempting to remedy the two most vexing problems with online educational materials—(1) lack of quality or gatekeepers: Merlot provides a check on quality, produces feedback to authors, and a centralized website to which faculty can send students for materials as at least a starting point; (2) Merlot provides faculty with tangible feedback for the online materials they produce, feedback needed in the promotion and tenure process. The lack of such feedback to date has often kept untenured faculty out of the process or has penalized them for spending their time in the production of online materials. These are the very faculty who are often the most innovative and need to be rewarded for quality efforts. The target audience of this paper includes those instructors who either just beginning to creating their own online materials and seek guidance from others in the field, or those whose portfolio of such materials is in need of peer review and recognition.

A brief overview of the young history of the Merlot organization will be given, beginning with its genesis at California State University in 1997. The home of Classical Studies is within the World Languages editorial team, one of 12 such discipline teams within Merlot.

Faculty in the various discipline-teams produce peer reviews and give feedback to authors for selected online resources. All reviews are published online at the website www.merlot.org. The website also provides links to a wider array of resources in each area which have not yet been reviewed, but are still under the oversight of the editorial team. Anyone may access the resources on the site anonymously; Individuals who sign-in, identify themselves and join the free organization, may submit materials they have created or those of others they consider useful. Such individuals may also submit “user comments” on any site, which can include suggested teaching exercises of their own based on a particular website. These user comments are categorized separately from the formal peer reviews, but are a tool for faculty collaboration. Merlot can thus be used as a forum for interaction of faculty who are active in the production and use of online materials.

The methodology of evaluating teaching and learning materials on the web will be explained. If the gods of the Internet are kind, a live online sampling of the currently peer-reviewed sites in Latin and Greek will be presented (if the omens are not favorable, this the presentation will be ‘canned’). Completed reviews include the VROMA site, The Silver Muse site, and Nuntii Latini. Pending reviews include the PERSEUS site and the Latin Library site. Examples of what works for teaching and learning Latin and Greek on the web, for elementary and advanced courses will be discussed. Issues will include effective strategies for web-based pedagogy particular to ancient languages, use of multimedia content such as streaming audio, viewing characters and non-Roman scripts on the web, online dictionaries, text repositories, and the current vogue, the production of online annotated texts.