AHRC ICT Methods Network Expert Seminar on Literature

TEXT EDITING IN A DIGITAL ENVIRONMENT

Centre for Computing in the Humanities, King's College, London, Friday 24 March 2006.

ABSTRACTS

Digital Editions and Text Processing

Dino Buzzetti, University of Bologna, Italy.

It was recently demonstrated (by P. Robinson) that digital editions, despite their superior functionality and sometimes their quality, have not yet gained widespread acceptance. A possible reason for this drawback may be seen in the difficulty of transferring a suitable part of the reader's 'competence' to the machine (T. Orlandi). This difficulty seems to depend, in its turn, on the lack of an adequate data model to process textual information. The standard form of text representation – consisting as it does of a marked-up string of characters – does not provide per se a satisfactory data model to analyse textual content. There have been some attempts to tackle the problem by extending XML semantics, notably the BECHAMEL project and the NARA/SDSC project for the implementation of knowledge-based archives. However an adequate understanding of the role played by markup in representing both syntactic and semantic relationships and abstractions is still lacking. It is argued here that a circumstantial examination of the ambivalent status of markup in relation to a string of characters could afford a dynamic model of the textual condition. Like any other diacritical mark, markup can be seen as part of the text, or as a metalinguistic indication of a certain textual feature. Moreover, markup can designate both an operation and its value: as the result of a restructuring operation, it provides a textual variant; as an instruction to a restructuring operation, it provides an interpretational variant. Through the introduction of an extended string as a new datatype (M. Thaller) and the integration of markup information into the object, or datatype 'string' on which an application system operates, all the injunctive and performative force of markup could be transferred to basic low-level processing and so enhance the functionality of text representation and digital editions.

Digital Editing, Text Markup, and the Construction of Textual Reality

Julia Flanders, Brown University, Providence, RI, USA.

As our experience with digital editing evolves and matures, we are coming to an increasingly detailed awareness of the interpretive inflections of text markup and its role in supporting recent editorial developments, such as explorations of immersive environments and visualization tools. Far more vividly than before, we are aware of editorial practice as the construction of textual reality - the persuasive arrangement of a field of information to permit and encourage certain kinds of reading and analytic processes.

Current text markup languages like TEI however, are still rooted in a more or less objective and structural understanding of textual research, their role focused on expressing determinate statements, and on naming and classifying structures and objects. Textual interpretation is accommodated chiefly through simple forms of annotation and commentary, which function as encrustations on the surface of the text rather than as inflections of its structure or meaning. In the new editorial world, where perspective and contingency are so powerfully foregrounded, can markup be more fully integrated into real editorial practice? How does the encoded text - with its place at the center of the digital edition - work with, or against, the grain of emerging ideas about digital editing?

Avant-textes, Intertexts, Hypertexts: Editing genetic documents
Daniel Ferrer, Institut des Textes et Manuscrits Modernes (ITEM), Paris, France.

Writers' working documents are not texts and should not (cannot) be edited as texts. They are protocols for making texts and their dynamic and axiological qualities should be preserved. Their nature is more relational than substantial, so they must be represented as networks. Genetic criticism has introduced the notion of avant-texte to distinguish the intellectual construction of this network from the passive gathering of the archive. It is important to try to preserve the open character of such a reconstruction and not to solidify it into a finished product but at the same time its critical and scientific authority must be affirmed. Traditional editions are notoriously incapable of meeting these requirements.

The density of spatial and temporal relationships within each single document, the complexity of endogenetic interrelations between documents and the diffusion of exogenetic (intertextual) connexions can only be adequately represented in electronic form, but some sections of genetic archives do lend themselves to paper editions. For practical and institutional reasons, it is sometimes desirable to publish them in book form. However such editions cannot be labelled 'traditional editions' since the use of digital tools by the editors and the availability of on-line supplements for readers confer new dimensions. The current edition of the *Finnegans Wake Notebooks* at Buffalo (Brepols) will be used as an example.

"..., they hid their books underground"[1]
Espen S. Ore, National Library of Norway, Oslo.

Books - including the manuscripts of Aristotle and Theophrastus - were buried by the inhabitants of the Egyptian town of Scepsis to prevent their being taken and absorbed into the library of Pergamum. When the books were eventually dug up they were sold to Apellicon. Strabo takes care to explain that Apellicon was a bibliophile rather than a philosopher. He was also no philologist, as according to Strabo he introduced much erroneous material to fill in the lacunae where the scrolls had suffered from mildew and other sorts of damage.

In this story we see a collection of books as something which is diminished if anything is removed, just like a dragon's treasure. The old historians also tell us how Ptolemaius II let the Athenians keep the deposit while he kept the books which he had borrowed from them to add to the Library in Alexandria. The Library in Alexandria was not a dead end for books and philologists have demonstrated that new and ideally corrected editions of the classics were created. To a certain degree we might even call them critical editions. Markup was developed which showed, for instance, which lines an editor considered suspicious or copied from other sources. There was also markup pointing from the physical copy or edition into a separate commentary. This work has not been forgotten. The traces left by the Alexandrian philologists, either as commentaries which have later been included as marginalia in medieval manuscripts or in providing canonical texts which have then been the basis for most or all of the medieval manuscript traditions, is the material which is built upon by modern editors of classical texts.

Modern edition projects, whether producing historical text critical editions or making available facsimiles of manuscripts are also producing something which can be built upon by philologists of the future. Although there has been discussion as to how far a new editorial project can build on existing digital archives, there are examples of projects which have done this, and there is no reason to believe that these are exceptions[2]. A number of projects at various stages in their proposed life-span will be used as examples and demonstrations in this discussion.

Dahlström, Mats, "How reproductive is a Scholarly Edition", *Literary and Linguistic Computing*, 19, 1, 2004, 17-33.

Ore, Espen Smith, "Monkey Business - or What is an Edition?", *Literary and Linguistic Computing*, 19, 1, 2004, 35-44

[1] From Strabo, *Geography* (English version from Jones (1924) as presented on Perseus)
[2] See for instance Dahlström (2004) and Ore (2004).

Strabo, ed. H.L. Jones, The *Geography* of Strabo, published on the Perseus website (Accessed February 2006): www.perseus.tufts.edu/cgi-bin/ptext?lookup=Strab.+13.1.1

A New Paradigm for Electronic Scholarly Editions
Peter Robinson, Institute for Textual Scholarship and Electronic Editing, University of Birmingham, UK.

One may distinguish two phases in the making of electronic scholarly editions. In the first, which has characterized the electronic scholarly editions made to this point, these editions translate the components of a print edition to the computer screen. We see in them the familiar content of the print edition: an array of texts, linked together by various kinds of apparatus, topped and tailed by various kinds of scholarly commentary. There may be many more such texts than can be present in any kind of print edition; there may be many more images of the primary textual sources; the linking between materials may be much more (or less) convenient; but we recognize the familiar lineaments of a print edition. There is an editor; the various parts are created by the editor; they are presented to the reader in forms determined by the editor (or, more likely nowadays, by whatever the current generation of software permits). Examples from existing electronic editions will be given. One might observe that most forays into electronic editions have been little more than digital facsimilies of single manuscripts or books, accompanied (if at all) by rather basic transcriptions: we are asked to compensate for the deficit of scholarship by the (occasional) ease of access.

One may discern a different model of scholarly edition: what I have called ‘fluid, co-operative and distributed editions’ (or, according to Peter Shillingsburg, ‘knowledge sites’). In these, what the reader reads may not be the single construct of a single editor or editorial group: it may be assembled by the reader himself, from various components contributed by various scholars and individuals, working together in ways unanticipated by their creators. An instance of an edition deploying these characteristics will be shown. These promise an opening up of both the editing and reading processes, and a relaxation of the sharp distinction between ‘editor’ and ‘reader’. This paradigm raises many issues, and its realization will require new tools. Especially, the starting point for these editions must be the availability of high-quality digital images of the original materials fundamental to the editions. This is a particular problem in the UK, where digitization costs are considerably higher than in other countries.

Digitizing Inscribed Texts
Charlotte Roueché, King's College, London, UK.

My aim is to present the steps by which I moved to the digital presentation of Greek and Latin inscriptions, to serve as an example of how such decisions develop, and how scholars can be supported in understanding the potential implications of digital media for their research.

The texts on which I work are also archaeological objects, inscribed stones, many of which have an function within the current excavations of the city of Aphrodisias in Caria. Each one must therefore be presented as an object, with its relevant metadata.

My initial intention, in publishing such texts digitally, was simply to offer the fullest possible account of them, supported by a rich archive of illustration. I only came gradually to appreciate the extent to which the exploitation of the texts, as texts, could be enhanced by the use of XML. Moreover, their presentation within a TEI compliant framework can be used to draw such material into wider debates over the language and textual products of the classical world; until now, standard modes of publication have tended to divorce such material from the 'literary' mainstream.
I will be showing the published site:

http://insaph.kcl.ac.uk/ala2004

and the site of the continuing project:

http://insaph.kcl.ac.uk.

I would like to raise some of the issues which have emerged from this process. One is a question of the responsibility of the originator - the person who has had unique access to a particular body of data. How extensive should my intervention be? And how can I facilitate the enrichment of the material by others (for example experts on linguisitics or onomastics)? What are the models?

Being Critical: Paper-based editing and the digital environment
Kathryn Sutherland, St Annes College, Oxford University, UK.

Classic twentieth-century theorists of text within the Anglo-American tradition (New Bibliographers) attempted in their theorizing and editorial practice to establish a resting place for text in the relentless change, transmission, and attendant corruption that represented the historical fate of the material text as they saw it. Their belief was in the substitutory power of the eclectic text, an emended and cleansed composite made up of several textual states. This they aligned with a previously unrealized or betrayed authorial intention, but which might be seen, from our digital vantage point, as a kind of compressed, synoptic, and editorially selective textual archive functioning as the authorial work. Two particular developments have helped shift the emphasis in textual criticism and in critical editing in the last three decades from ideal stasis to historical metastasis: one is the articulation of a sociological or cultural studies strain of textual criticism; the other is the rapid expansion of electronic storage and dissemination as a textual device. The interesting thing is how these two – the sociological and the electronic – have become mutually implicating: that is, how the contingent and material appear to find expression and revaluation through the digital. The very idea that the electronic archive's storage capacity might convert, through the accessing, manipulation, and analysis of multiple documentary surrogates, into a strictly materialist approach to the question of how texts exist is deeply problematic. It is indeed as problematic as the New Bibliographic faith in an eclectic text's ability to defy the historical fate that befell its component parts. Hence the importance of being critical. We are paying insufficient attention to electronic difference at almost every stage of our engagement with the architecture and functioning of the electronic archive or edition because we are too enamoured of electronic simulation. Electronic materiality, the difference that electronic instantiation makes, is currently far more of a hindrance to textual engagement than the robust, versatile medium of print. We have not yet thought hard enough about who will use electronic editions or how often or for what real purposes. This paper will argue for a critical (re)appraisal of the preparation and production of scholarly editions in paper and electronic forms, examining assumptions about stability and durability on the one hand and fluidity and flexibility on the other.

Every Reader his own Bibliographer — an Absurdity?
Edward Vanhoutte, Centre for Scholarly Editing and Document Studies, Royal Academy of Dutch Language and Literature and University of Antwerp, Belgium.

The qualifying characteristic of an edition is its scholarly basis, not its form of appearance. The difference between a scholarly edition and an non-scholarly edition is not its articulated orientation towards a scholarly or a non-scholarly audience respectively, but the scholarly quality of the research leading to the publication of a text. A reading edition, (re)presenting only one version of a text with no or almost no additional information can therefore be as scholarly or non-scholarly as a full historico-critical edition representing every single variant in every single witness of that text, depending on the scholarly integrity by which the edition has been prepared. By this statement I deliberately move away from the conventional distinction between the alleged non-scholarly editions meant to be read and the acclaimed scholarly editions meant to be worked with.

In the paper edition, established layout conventions allow the user to verify each and every editorial act and agree with or oppose the editor's practice. This fear of the undisclosed textual detail on both sides of the edition has resulted in a classification of editions on the basis of their amount of additional information (commentary, glosses, variants, annotations, and so on) and the chosen layout of the presentation of the data (inclusive text edition, synoptic text edition, parallel text edition, multiple apparatus, lemmatized apparatus, apparatus variorum, and so on). Too often the layout of the printed edition is erroneously identified with its scholarly status.

The electronic edition has failed to free textual scholarship from these layout economics invented for the printed page. Moreover, the peculiarities of reading on screen have turned the proclaimed

electronic reading edition into an unrealized anomaly. The electronic paradigm in scholarly editing has almost exclusively focused on the advantages of the size and economics of available storage capacity, the democratizing possibilities of providing access to full colour representations of documentary witnesses, and the nonsequentiality of hypertext. The digital archive as expanded text has in some cases jostled the one text away in favour of the multitude of many texts. Scholarly assessment of electronic editions is mainly based on the incorporation of the digital archive representing the textual history and the presence of collational evidence.

But the presentation of the digital archive does not discharge the editor from his/her responsibility to serve the text and support its function in society. As an academic product the electronic edition is of inestimable value, as a cultural product it is valueless. Therefore, electronic scholarly editing is in need of a reintegration of the reading edition without any compromise concerning its scholarly status. I propose to defend this proposition by using evidence from the electronic edition of the Flemish novella De trein der traagheid (The train of inertion) by Johan Daisne, which is currently underway at the Centre for Scholarly Editing and Document Studies of the Royal Academy of Dutch Language and Literature.