

Text editing, scholarship, books, and the digital world

A seminar held on 29 June 2006, and convened by Marilyn Deegan and Kathryn Sutherland

Report by Marilyn Deegan, 14 July 2006

Introduction

On 25 March 2006 the AHRC ICT Methods Network sponsored an expert seminar on Text Editing in a Digital Environment which looked at a number of different projects and concepts around text editing, electronic editing, and the print world. This raised a number of important questions including:

- In an environment of print and electronic culture, how seriously do we envisage the falling away of print in respect of the electronic edition?
- What new kinds of edition are made possible through the electronic medium?
- What constitutes an edition in the electronic medium? How is this related to the notion of an electronic archive?
- Do we still see the scholarly edition as serving the needs of readers as well as users?
- What do we envisage the cultural status of the electronic edition to be?
- Is the role of the editor changing in the electronic environment?
- What new kinds of editing partnerships are emerging?

These questions were taken as a starting point on 29 June. In the morning session, we heard three short papers, and then in the afternoon all participants discussed what editing projects they were engaged in, with a structured discussion following.

Morning session

The Nineteenth Century Serials Edition (NCSE)
Jim Mussell and Suzanne Paylor

Convenient Editions
Peter Shillingsburg

Electronic Editions and Collaborative Interpretation
Paul Eggert

Marilyn Deegan and Kathryn Sutherland began by setting some of the context for the day's discussion, referring in particular to the points above.

Jim Mussell and Suzanne Paylor from the Nineteenth Century Serials Edition project based at Birkbeck College first discussed some of the particular characteristics of periodicals which makes their editing problematic in non-digital domains. Periodicals ask different questions of editors than books do, and so new models are needed. It is timely to look at some models for how periodicals are being re-published in the digital world, and to consider whether what should be provided are archives and gateways rather than actual editions in the more conventional understanding of the concept. Periodicals demand different treatment to other kinds of published products as they have multiple authors, publishers, illustrators; they have different notions of 'authorship'; different notions of beginnings and endings; different layout and composition conventions.

The Nineteenth Century Serials Edition project is working with around 100,000 printed pages from six titles, which makes it impossible to publish on paper. But there are also other aspects of this project which make paper publication difficult: the different objects which make up the periodicals have different relationships with each other, and the content is associated with certain 'moments'. The relationships within the titles are both structural and

hierarchical; both generic and thematic; and meaning is also constructed through layout and typography. There are also highly complex prosopographical relationships between people involved in producing periodicals: printers, publishers, editors, contributors, illustrators, engravers and paper merchants.

A key question that the presenters posed was, who is this material being edited for and for what purposes? Preservation is a key issue for periodicals, and digital methods are more versatile for preservation than microfilming, giving easier access to content.

The six titles chosen for editing by NCSE are:

The Publishers' Circular (1880-1890)
The Tomahawk (1867-1870)
The English Woman's Journal (1858-1864)
The Leader (1850-1859)
The Northern Star (1838-1852)
The Monthly Repository of Theology and General Literature (1806-1838)

These were chosen for a number of reasons: the titles mix genres and formats, and are very diverse. However, they do share generic elements, and they change over time in format and content. A key aim of the NCSE project is the provision of digital tools to handle the materials, including content analysis, text mining, and concept mapping.

There are many challenges for the project, some of which are posed by the choices made in selecting the journals. Some titles, for instance, are more like books or periodicals (the *English Woman's Journal*) and some are more like newspapers (the *Northern Star*). Some of the titles do not have a linear sequence of weekly numbers—often there are eight editions on one day. There is also some creative use of typography when pictures are made from words, which makes computational analysis difficult. Or pictures may relate to texts on different pages. The project is working with Olive Software to link text developed through Optical Character Recognition to images of page components.

The target users for the project are envisaged as being specialists in nineteenth-century periodicals; other academics; schools; and general users

Discussion

There was some discussion about the value of editing material such as this, rather than indexing and adding metadata. Presenting the full text of the periodicals has the advantage that the user has everything they need without interpretation: indexing and metadata description is subjective, but 'the text tells you what it is'. It would also be difficult to handle advertising, a vital area of information, and would probably be much more time-consuming and expensive.

Other examples of online periodicals such as the *Times Digital Archive* were discussed, as well as a small-scale project edition of *The Bijou*. Many projects in this area, however, do not represent fully the *form* of the periodicals, as well as their content, something the NCSE team see as vital.

Peter Shillingsburg of the Centre for Textual Scholarship at De Montfort University discussed the concept of what he called Convenient Editions: for the study of literature we need texts, and for the study of texts we need a whole range of specific texts in the forms of drafts, manuscripts, typescripts, proofs, magazine publications, books, revised editions, and scholarly editions.

To conduct this business we need a number of things, but we do not need computers: they are just a convenience. If computers are used, however, we feel we have to provide something that could not otherwise be done—something that shows that the electronic world supersedes the print world. One of the problems for Shillingsburg is that scholarly *electronic* editions are the results of attempts to be as good as the scholarly *print* editions that very few people use. Most students and literary critics don't actually use scholarly editions. They use their own cheap paperback copy or, in the online world, texts from Project Gutenberg—texts of sources unknown or misidentified, and with other problems, and the reason for that is convenience.

For Shillingsburg, the scholarly electronic edition of the future—the one that will actually be used and therefore influence literary study and criticism—will be convenient: as cheap as a paperback book, with a user-friendly interface (adaptable by the user to suit his or her condition, whether the user is a scholar, a student, or a tourist), and can be

treated as the user's own, with bookmarks, highlighting, space for marginal notes, and the ability to annotate or even change the materials that appear on the screen in what must truly feel like the user's very own private copy.

The vision of the future that Shillingsburg proposed is a Collaborative Literary Research Electronic Environment operating as shared space, not in the absolute control of the general editor or other owner but available to scholars in the field. This space would consist of interconnecting modules, and would operate as a knowledge site. Many projects would be group projects, and would create dynamic and interactive works.

Discussion

Discussion ranged around the complex commercial issues inherent in the kind of interactivity proposed by Shillingsburg, and around the problem of public funding for certain kinds of materials. Canonical materials are easier to secure funding for, even though the materials are often widely available, but these can ensure wide usage, which means that the funders receive 'value for money'. Worthwhile but less popular materials are harder to attract funding for.

Paul Eggert began by announcing that he had come not to bury the book, but to praise it, and made a very persuasive argument for what he called 'the beauty of scholarly completeness' in the production of a paper-based edition, something which is not possible with an electronic edition. In the 'finished' print edition, 'every part of the volume [is] enlightened by every other part, all of it seamlessly interdependent and unobjectionably cross-referenced, nothing said twice, all of it as near perfectly balanced as you could ever make it.' In comparison with this, there is no deadline for an electronic edition, and edits can be made at any time. Does this, he questioned, mean that the scholarly rigour brought to bear in the print world be relaxed? And what of the interactive and multi-author possibilities engendered by electronic editions? Can we engage in these new ways of doing things *without* compromising traditional standards of accuracy and rigorous reasoning?

Eggert then went on to discuss an electronic editing project that he had been engaged in for some years which uses the concept of 'just-in-time-markup (JITM)' to ensure the accuracy and authenticity of the electronic text. This system runs counter to common practice in markup, where tagsets are inserted into the text, and travel along with it when it is transmitted or transformed. JITM keeps markup and texts separate, and any corruptions or changes in the text are detected instantly using algorithmic methods (checksums) to keep track of even the slightest difference. JITM also has the advantage that various interpretative 'perspectives' in the text can be generated at will, but the underlying text is left unchanged.

Afternoon Session

Julia Briggs discussed the HyperNietzsche project, which is a kind of electronic research matrix designed to facilitate the cooperative and cumulative effort of a community of specialists and to make their work freely available on the Internet. Briggs is currently editing some of Virginia Woolf's materials for this, looking at the process and history of Woolf's writings. Briggs showed some examples of Woolf's manuscript and typescript writings which will be presented in the HyperNietzsche framework, including *A Sketch of the Past* and *The Hours*.

Linda Bree of Cambridge University Press discussed a number of editing projects that the Press is engaged in. The Ben Jonson project is producing both a print and an electronic edition, while the Jonathan Swift project is producing a print edition with an accompanying electronic active of a number of early editions of Swift.

Regular printed editions are being produced of Jane Austen and Samuel Richardson, and in the twentieth century, there are editions of Scott Fitzgerald, D.H. Lawrence, and Joseph Conrad. These editions will be also sold as XML text but will not be produced as electronic editions.

CUP are also producing a number of edited volumes of letters. These are a format which would benefit greatly from electronic representation, but there are a number of considerations which make this difficult, in particular dealing with the estates of the writers. Editions of the letters of Conrad and Fitzgerald are almost complete, and work is proceeding on four volumes of the letters of Samuel Beckett. The Beckett estate will not allow e-publication, which CUP have accepted. CUP are also working on a fifteen volume edition of the letters of Ernest Hemingway. The Hemingway estate will also not grant e-rights, but CUP are less willing to accept this as they probably need to have the rights to produce an e-version in order to make the edition financially viable. Libraries not willing to buy print versions of expensive volumes, and they might prefer to have electronic versions.

Bree made some general remarks about the publishing of editions. She remarked that a printed scholarly edition has a long shelf-life, but sales of scholarly editions have fallen away and therefore prices have gone up, which makes the readership even smaller.

Graham Law from Waseda University in Japan is working on non-canonical literary works, including periodical literature and the collected letters of Wilkie Collins. The Collins letters were originally produced as a paper edition: of the four editors, one did not want to produce an electronic edition, and nor did the publishers. Four volumes of the letters were published in paper form in 2005. During the editing process, the team used high quality digital images of 80% of the letters to transmit around the world. The second stage of the project will be the production of the letters on CD ROM. Charles Dickens' letters are now available on CD ROM. Eventually there will be a freely available version of the Collins letters on the web.

Law suggested that we need to think hard about transition formats, and he pointed out the huge gains that a move even to PDFs brings. He also pointed out the advantages in the availability of digital versions of different authors (like Collins and Dickens) even if they are not cross-searchable.

Jim McClaverty introduced the Jonathan Swift Edition being published by CUP. There will be 15 volumes of this, plus an archive of variant texts. Swift is a canonical author whose works have not been edited for 50 years. It is not the establishment of the text which is driving the edition but new annotations: the text has a subordinate role, and the editors are interested in new explanations, and hope to change the interpretation of Swift for scholars and students.

The archive will provide all the other variant texts. These will be searchable and can also be collated. There will also be essays about the texts and variants. It will then be possible to go on to produce an electronic edition at a later stage. There are three key advantages to the production of the digital archive: remote accessibility; virtual reintegration; and searchability.

Edward Vanhoutte of the Centre for Text Editing (CTB) at the Belgian Royal Academy introduced the DALF project, a text-base of encoded annotated letters of Flemish authors. There are around 4000 letters in a corpus of 63 different authors. The textbase includes facsimiles of letters, transcriptions, indexes, different views, annotations, and header information. Users can export the letters into their own corpus.

His team at the CTB have also published the WWI diary of Virginia Loveling, which sold 600 copies immediately. They also published an abridged version with a commercial publisher, and published it online as a blog, which has attracted people who would never read a scholarly edition.

His *De train der Traagheid* is an edition of 20 versions of a novel in which every text is a variant. In the edition, there is a reading text which acts a palimpsest, with all the other versions underneath with the variant areas showing. Users can view the XML, can produce a printed edition, can show versions in parallel, and can reorder the various versions. The CTB also did some experimentation for the Swift archive.

Kathryn Sutherland is currently working on a project which is producing a digital resource to reunite Jane Austen's fictional holograph manuscripts. These are working manuscripts which were revised over both short and long periods of time. This project has been funded by the AHRC and is at an early stage.

General Discussion

The discussion was broadly based and picked up on many points made throughout the day. The key points that came out of it were:

1. Supporting electronic publication on media such as CD ROMs can be very expensive as many things can go wrong; the underlying programs also need constant patches. This has engendered a change between buyer and publisher as when the CD doesn't work, they ring up and complain, even if it is a number of years since their purchase.
2. Some publishers are not selling CD ROMs but structured data on CD ROMs, with web based programs accessing the data.
3. With printed scholarly editions, lists of variants at the foot of the page or in the back of the book are difficult to interpret. Presented electronically they can be easier to see and interpret.
4. Are scholarly editions only ever going to be published in the future if they are produced with a grant? Is this a version of 'author pays'?
5. Electronic editing can be a kind of self publishing, which carries a reputation that scholars don't want.

6. When there is a hard copy component to an edition, a publisher has an interest. With only electronic, there is no robust and well-tried financial model. Publishers will often not take on an edition unless the work is going to be grant funded.
7. Do grant bodies concern themselves with outcomes too much, and not enough with research?
8. Is it a concern that it is easier to get grant funding for canonical works and authors?
9. Peer review: gaining academic credit for teams that work on electronic editions is a problem.
10. Editing is now more of a collaborative enterprise and the humanities is moving towards a more scientific paradigm. We may have to think in ways we didn't before.
11. Academic publishing is not remunerative.