So successfully have a feminist rhetoric and feminist critical and historical practices been appropriated and professionalized within the academic orthodoxy during the last decade that the ambitious male scholar can scarcely do better for his career than espouse and reproduce their argumentative models; yet it remains not the least ironic aspect of this 'success' that, with a few isolated exceptions, the basic materials of a literary revision, the neglected writings of women authors across a range of fictional and expository genres and through several centuries, continue to be as difficult to come by as they were twenty years ago. The rhetoric may be firmly (too firmly) in place, but (leaving pressing political and sexual considerations aside) we are only beginning to explore its textual accountability.

WOMEN WRITERS PROJECT
The Women Writers Project, sited at Brown University, Rhode Island, is a full-text database, or textbase, of women’s writings in English from c.1330 to 1830. It was begun in 1988, with funding from the National Endowment for the Humanities, by a group of scholars concerned to make visible and available the rich and diverse tradition of women’s writings which, despite the growth in feminist scholarship, were still not finding their way into the bookshops and into the classrooms. A determining event in the Project’s pre-history was the publication in 1985 of the Norton anthology of women writers, in which out of more than 2,400 pages fewer than 200 were given to women writing primarily before the Victorian period. The consequence of such an unequal representation of female talent in a major (and feminist) North American teaching anthology was bound to distort and underestimate that tradition of women’s writing it hoped to serve. Women have always been writing; and if publishers’ lists and university syllabuses fail to reflect this fact, the reasons are economic as well as cultural-political.

The Women Writers Project has from the first admitted to global ambitions. Its theoretic aim is to enter into the computer everything written in English by women in its 500-year period. Its scope includes printed books and manuscripts, the traditional literary genres of
fiction, poetry and drama, as well as sermons, prayers, household manuals, translations, letters, diaries and whatever else women can be found writing. Its list of entered texts (currently over 200 titles) includes: Anne Weamys’s *A continuation of Sir Philip Sydney’s ‘Arcadia’* (1651); the journals of two Quaker women, Katherine Evans and Sarah Cheevers, who set out to Rome in the mid-seventeenth century to convert the Pope and were imprisoned in Malta by the Grand Inquisitor; Jane Sharp’s startlingly explicit *Midwives book; or, The whole art of midwifry discovered: directing childbearing women how to behave themselves in their conception, breeding, bearing, and nursing of children* (1671); all four volumes of Eliza Heywood’s *Female spectator* (1745); numerous collections of women Romantic poets; etc., etc. The sexual and cultural politics underlying the enterprise are radical in the questions they pose concerning the traditional and gendered limits of ‘Great Literature’ and in the conjunction they suggest between previously marginalized and neglected representations of knowledge and reality and state-of-the-art computer technology.

A principal task facing the Women Writers Project has been to make available texts otherwise inaccessible, or, because so much women’s writing exists only in manuscript form or obscure typefaces, difficult to read. Meeting this challenge has resulted in pioneering work in computerized textual reconstruction and dissemination. Less constrained by the economic and aesthetic considerations of traditional book-bound publication, the computer can support multiple versions of a text. One version might be a scanned image of the manuscript or earliest printed edition; another might reproduce original typographical errors and include variant readings from other editions of the text; yet another might be an error-free version, designed specifically as a reading text.

Beyond this, a major part of the effort in creating a full-text database is the systematic analysis of the texts and the development of a representation scheme which will reflect the results of this analysis. In other words, it is important to recognize that a literary text is more than just a sequence of words; it is an organized structure of special components. A play, for instance, generally contains stage directions, character identifications, and scene and act designations, as well as lines of dialogue. Even a simple poem might include a title, dedication, stanzas, lines and marginalia. Whether a sentence occurs in the dedication, the title or a verse makes a difference as to how that sentence is understood. Like the descriptive bibliographer, those engaged in the computerization of text must anatomize and codify a range of features.

These features are also crucial for scholarly research: a scholar studying metre must distinguish the words which are lines of verse from those which compose the title or dedication. She must also be able to tell where one line of verse ends and another begins and where one stanza ends and another stanza begins. Similarly, to explore what
CHALLENGING ASSUMPTIONS

Concepts are frequently closely conjoined by an author, one must be able to tell not just whether two words are close to each other by some arbitrary measurement (such as 'within five words'), but whether they are in the same poem, same stanza, or same line. Readers and scholars ordinarily rely on background information and typographical distinctions as clues for recognizing these components and their relationships. But for the computer to assist us in studying texts or even in typesetting them, we must have an explicit and systematic identification of these components stored in the database itself, so that computer programs can recognize the different parts of a text without relying on background information or varying typographical conventions.

The question then becomes how best to represent the structure of a particular work explicitly in the electronic version of the text (Figure 1). This involves not only careful analysis of the individual work itself, but a general understanding and analysis of that work's literary form. For example, while a Shakespearean sonnet and a Petrarchan sonnet both are poems and both are composed of fourteen lines, they are also very different constructions (a Shakespearean sonnet has three quatrains rhyming abab cdcd efef and a couplet rhyming gg; a Petrarchan sonnet has an octave rhyming abbaabba and a sestet rhyming cdecde or cdcdcd). Every type of writing has its own distinct vocabulary and grammar of text components. Novels, letters, scripts, sermons and poems are all made up of different kinds of parts, although some of these parts may be more commonly discussed than others.

Many of the textual analysis and encoding practices of the electronic environment are simply applications of traditional literary classification and analysis. But in order to develop a classification scheme which will support the extended potential of computer-assisted research, it is necessary not only to standardize the terminology of literary research, but to make refinements and additions to the traditional theory. In the Brown Women Writers Project database, the components of a text are identified by markup tags using a system which conforms to SGML (the Standard Generalized Markup Language), an international standard for devising text markup systems, and to the formulations of the Text Encoding Initiative (TEI), a major international project which is designing SGML-type markup tags and protocols specifically for the electronic encoding of literary texts. The Brown Women Writers Project is affiliated with the TEI task group and is testing and reviewing its recommendations.

Figure 2 presents an overview of the Women Writers Project approach to text management. The box in the lefthand panel represents the database of fully analysed texts with their structures explicitly indicated by markup tags. The middle panel represents some of the different sorts of processing which might be performed on

1 For a comprehensive (and comprehensible) summary of SGML and TEI, see the article by Allen Renear in the present collection.
The Authors Dreame
to the Ladie Marie, the Countesse Dowager of Pembroke

Me thought I pass'd through th' Edalyan Groues,
And askt the Graces, if they could direct
Me to a Lady whom Minerua chose,
To liue with her in height of all respect.

FIGURE 1: SCREEN DISPLAY OF TWO VERSIONS OF A POEM
From Aemilia Lanyer, Salve Deus Rex Iudaorum (1611), marked up in SGML and then processed to show the original typographic features

the database. Since the database itself is independent of any particular application of software and conforms to a well-understood standard for markup, many different tools can be used to study it without encountering problems with file format compatibility or data interchange. The righthand panel represents the various information products which result from applying different special processing to the database. For instance, one common product is printed books and anthologies. Using a database management system, we can select those texts from the database to be included in a book and then systematically format them in a style appropriate for the volume being produced. A poetry collection in a series for the general reader, for example, will omit information, such as the indication of original page breaks or comparison of variant lines from different editions, which would be included in a scholarly edition of the same poems. It will also correct without comment typographical errors and misspellings or archaic spellings. Finally, it may need to be typeset according to the
FIGURE 2: FROM ONE DATABASE TO MANY INFORMATION PRODUCTS

How the Women Writers Project is using the computer to generate materials for research and instruction

general design of the publisher's series in which it is being issued. Using a database management system and exploiting our analysis of the text into its major components, we can select and arrange only those parts appropriate for a general edition. Then the required formatting rules can be systematically applied: for example, all titles can be centred, printed in larger type and italicized, without each individual title having to be changed, since all titles were previously identified with a 'title' tag.

On the other hand, if a scholarly critical or variorum edition is being created, then such things as indication of original publication, editorial variants, typographical errors and archaic spelling can be included. Here different formatting rules will be applied – rules designed to create a visually lucid text for scholarly study, rather than those which conform to the style of a publisher's series. Again, using a database management system, a different selection from the database can be made and formatted by simply specifying different formatting
rules for each kind of text component. Finally, as well as supporting tailored publication, the same database supports stylistic studies, concordances, metrical analyses and content and semantic studies in a way which is far superior to unstructured databases. We can pose questions which depend upon recognizing the major editorial features of a work and not just the incidental formatting.

PROJECT ELECTRA

Project Electra can best be described as a daughter project of the Brown Women Writers Project, since it draws upon and extends some of the textual and computational features of the earlier project. Its brief is the creation and implementation of an electronic scholarly resource as part of a major electronic and conventional paperbound edition of works for those undertaking specialized research in British Romantic Literature and Women's Studies in the period 1785–1815. Still in the first stages of development, the initial design is in terms of four modules, each with its own discrete coherence and each conceived and implemented in dynamic relation to the other three. These four comprise: an electronic textbase of female polemicists 1785–1815; an electronic textbase of women's writings for children 1785–1815; an image bank of female representations in contemporary prints 1785–1815 (fashion plates, political cartoons, etc.); and a database of historical, biographical and bibliographic information, incorporating details of authorial relations, readerships, publishing history, facsimile images of first and early edition title-pages, women's magazine contributions, etc., etc. The idea of the bio-bibliographical database is to create as dense and complex a picture as possible of the conditions for women's authorship in the period, and in doing so to provide a resource of importance to literary scholars, historians, cultural theorists, art historians, scholars of publishing history, etc.

In Greek story Electra is the mistreated daughter of Agamemnon and Clytemnestra and a type of the Cinderella figure. Raised in obscurity and neglect, she eventually overcomes her ill-treatment and shapes her own destiny with heroism. The name Electra is used to draw attention to the project's yoking of current developments in electronic storage and retrieval of information and the rehabilitation and scholarly dissemination of a marginalized but rich resource for the study of women and their representation in history. Women's writings for children and women's polemical writings represent valuable and neglected sources of literary, political and cultural evidence for the

2 Project Electra has been awarded partial funding (in the form of technical support, money for data collection and entry, etc.) for three years (from January 1992) from the Information Systems Committee (former Computer Board) of the U/C. This is as one of three academic partnership projects under the management of the C&I Centre for Textual Studies at Oxford University.
period 1785–1815. Project Electra will provide texts, images and supporting materials able to be viewed from a variety of perspectives. At the least, it is envisaged that it will make more widely available much unpublished and inaccessible material and thus provide opportunities for new scholarship.

Representative texts to be entered in the first stage of development include fictional and instructive writings for mothers and children and sexual and political polemics in magazine and chapbook versions as well as in book-formatted works. For example, Maria Edgeworth's *Parent's assistant* (1796) and *Moral tales for young people* (1801); Elizabeth Hamilton's *Letters on education* (1801); Laetitia Matilda Hawkins's *Letters on the female mind* (1793); Mary Hays's magazine writings (from the *Monthly Magazine*, 1796–97) and her *Appeal to the men of Great Britain in behalf of women* (1798); Hannah More's *Cheap Repository tracts* (1795–98); Mary Ann Radcliffe's *The female advocate; or, An attempt to recover the rights of women from male usurpation* (1799); Mary Robinson's *Thoughts on the condition of women, and on the injustice of mental subordination* (2nd edn, 1799); Sarah Trimmer's *Family Magazine* (1788–89), and *Guardian of Education* (1802–06); Priscilla Wakefield's *Reflections on the present condition of the female sex, with suggestions for its improvement* (1798); and Mary Wollstonecraft's *A vindication of the rights of woman* (1792).

The years 1785–1815 witnessed an intense period of social prominence and intellectual activity for women: in the redefinition of the domestic sphere; in shaping a proto-industrial society; and in the context of the political debates on natural and civil rights. The period saw a vital mixing of literary and expository genres. In the case of a women's tradition of polemical writings in English, with the exception of the works of Mary Wollstonecraft (only recently published in a modern collected edition), most of this material has not been reprinted since its original publication; some few works are available in later nineteenth-century reprints; some are scarcely to be found in the copyright libraries; almost all are accessible only with difficulty. In the case of Hannah More's *Cheap Repository tracts*, fragile and ephemeral pamphlets, we now use later authorially massaged reprints (themselves long out of print) which do not capture the vivid chapbook formatting of this literary assault on the poor. In the case of magazine publication, a fertile environment for writings by women which target women and children readers, we have lost the determining contexts and influential interpretive frameworks for their early reception.

Computerization can return the textual documents of this debate to critical and scholarly currency more economically than, or as an inducement towards, their republication in more traditional form; but it can do more: texts stored electronically can be represented and manipulated in ways not otherwise practical. Textual critics have recently realized the potential of computerized modes of presentation, and electronic editions of major literary works are now in preparation.
Within Project Electra particular attention is being given to assembling a complete collection of Hannah More's contributions to the Cheap Repository series of tracts, a programme of improving literature which she launched in 1795 with the assistance of supporters in the upper and middle classes and the Church of England. Many of the tracts she wrote herself, others were apportioned to her sister Sarah, and over others she exerted a proprietary influence. All were carefully modelled on the popular street literature favoured by children and the lowest ranks of society. Her original printers were John Marshall and Samuel Hazard, the prominent London printer and distributor of street literature and the owner of a circulating library at Bath. According to More’s early biographer, two million tracts were sold or distributed in the first year of issue alone. These original tracts continued to be reprinted into the 1840s by various printers, in America and Ireland as well as in Britain. From their conception, the Cheap Repository tracts were fakes – deliberate impositions upon the poor – and importantly the deception lies in their original appearance, in the visual enticements of their crude woodcuts, racy titles, and clumsy formatting, all of which (wrongly) signal their vulgar origins.

With the assistance of the Special Collections Librarian of the Bodleian Library, Oxford, Project Electra is assembling a text- and imagebase of More’s contributions to the Cheap Repository series, incorporating high quality images of the earliest issues of the chapbooks as well as fully tagged diplomatic transcriptions, as a stage towards their critical editing. Not only can we do new things with each document – with the tagged transcriptions and the scanned images – but each can interact with the other to produce a new understanding of the work. Computerized technology insists on a radical questioning of the nature of text, along with more practical issues of translation from one textual medium (print) to another (the electronic textbase). In facilitating the scanned reproduction of works in their original publication form, textbase technology can contribute to restoring that visual authenticity which is lost in their translation into the modern paper edition and which, spectacularly in the case of the Cheap Repository tracts, can be a vital component in their meaning, a material witness.

A team of scholars in Britain and North America is currently seeking funding to produce an electronic edition of *Piers Plowman* which will incorporate transcriptions and facsimiles of the major manuscripts. The Computers and Manuscripts Project at Oxford University is about to embark on a large-scale study of the manuscript tradition of the *Canterbury Tales* using computerized collation and cladistic analysis of variants.

which is essential to the interpretation of their politicized cultural relations.5

Because of the restricted availability of women's writings from this period, we have difficulty in conceptualizing a female literary and intellectual contribution; and so we minimize its contemporary significance or restrict it (in the case of sexual politics) to the falsely opposed voices of 'left' and 'right', of the radical Mary Wollstonecraft and the Tory Hannah More.6 In most cases, what is available is confined to randomly edited and reissued works of fiction – the exception, of course, is the complete novels of Austen. One purpose in reintroducing an extensive list of women polemicists working in a wide range of fictional and non-fictional modes is to show how in these writings the languages of moral economics, of sexual politics and of fiction address one another. The electronic medium offers a valuable opportunity to give collective identity (and therefore greater literary and cultural significance) to a women's tradition in writing in this period and to broaden and complicate the picture we now have of it.

For example, the tagging systems already developed allow for highly specialized searches, and techniques of stylistic analysis make it possible to retrieve concepts across a range of works and genres. As a gendered textbase, Project Electra will eventually provide a rich repository of women's written linguistic usage in a variety of readerly situations and during a vital historical period when theories of language were centrally and explicitly yoked with issues of political moment – the enforcement and questioning of class divisions in terms of restricted reading audiences, the relation of vulgar to educated discourse, the possibility of an adequate women's discourse, and the relationship between articulacy and liberty.7 The retrieval and analysis of women's linguistic usage in a range of reading contexts, from fashionable magazines and stories for children to political pamphlets and tracts for the poor, will allow us to complicate our readings of institutionalized writers like Austen, Burney and Edgeworth, still too often assessed in terms of the normative model of male practices, and to reassess their contributions to the shaping of public as well as private constructions of reality.

5 Such arguments can be seen as extending to the storage and reproductive capabilities of the electronic environment the proposals of some modern textual critics that we expand the scope of traditional bibliography to include the text as social and material construct. Under such a definition, 'bibliographical codes' (typography, layout, format, etc.) are given great prominence in the formation of textual meaning, in contrast to the 'linguistic codes' (the words of a text), on which textual criticism has tended to concentrate. See, for example, Jerome J. McGann, 'What is critical editing?' Text, 5 (1991), 15–30; and D.F. McKenzie, Bibliography and the sociology of texts (London: British Library, 1986).


Under the terms of its partial ISC funding (see note 2), Project Electra is concerned also to introduce technology into the classroom. Electronically stored text is fluid text, able to be accessed as traditional hard copy reading versions in a variety of forms – as customized single-author editions, with varying amounts of scholarly and editorial features, and as thematically or chronologically compiled anthologies. In its function as a resource for advanced undergraduate and post-graduate users, the Electra textbase will allow for the assembling and reassembling of authors, texts and images in a variety of disciplinary contexts, depending on the historical, literary, pictorial, biographical or bibliographical priorities of the user. As the text- and imagebases develop in internal sophistication and size, teaching uses will expand to include specialized searches and stylistic analyses. It is envisaged that elements from Project Electra will function as the foundations for courses at postgraduate level. For example, there are plans for designing a system for the teaching of research methods drawing on data included in the bio-bibliographical database. More speculatively, the electronic assembly of material demands its reconception: book-based as we still are, we have scarcely yet begun to understand the intellectual possibilities unlocked by the flexible electronic repository and the seemingly renewable design of electronically stored information.

Project Electra’s technical brief, like that of other electronic full-text databases in these early days of computerized information storage, is to investigate the recovery and management of materials and to test the viability of determined methodologies and authoring environments for the most flexible retrieval of data. It is providing opportunities for assessing the management of text and image resources through a hypermedia front end and for investigating the possibilities of generating dynamic linking. Like the Brown Women Writers Project, it is working to define standards for the encoding of texts and images so that key words, concepts and representations can be identified and accessed under a number of headings. Accordingly, it also uses SGML and TEI conformant tagging. Such standardization is vital to the automatic links which these resource bases are making between texts, images and ancillary materials and to their long-term renewability – their potential for expansion to include further texts and materials. The possibilities for the importation and exportation of materials between the two projects should enhance their international status as electronic and gendered archives and provide a vital test of the electronic corpus as a collaborative scholarly environment.

In the case of Project Electra, images are being captured, digitized and stored at the highest possible resolution, although current technology dictates their initial delivery at a lower level of resolution. Storage will be in a standard file format on CD-ROM. The database is designed to be hardware and software independent. The four image- and textbase modules are integrated and manipulated via a hypermedia front end with dual platform micro-computer implement-
CHALLENGING ASSUMPTIONS

platform-independent design of the individual modules, allowing their integration behind new front ends as technology advances, acknowledges that the front end of the system is the most fragile element, and ensures the long-term survival of the resource.

Currently, the four modules are being integrated and manipulated within Microcosm, a multimedia information management environment running under Windows. Microcosm is particularly suitable for the assembling of diverse materials.\(^8\) Within Microcosm it is possible to browse through large bodies of multimedia information by following links from one place to another. It is also possible for the user to add links and further information to the system. Microcosm is, in this sense, an open hypermedia system. However, it includes many significant features which place it at a higher level than most currently available hypermedia systems and which make it a particularly suitable environment for integrating the text and image data and processes currently required by Project Electra.

Microcosm consists of a number of autonomous processes which communicate with each other by a message passing system. No information about links is held in the document data files in the form of markup. All data files remain in the native format of the application which created them. Instead, all link information is held in link databases, which hold details of the source anchor (where there is one), the destination anchor, and any other attributes, such as the link description. This model has the advantage that it is possible for processes to examine the complete link database as a separate item, and also it is possible to make link anchors in documents which are held on read-only media, such as CD-ROM and videodisc.

A major strength of Microcosm is its ability to integrate other applications. In fact, Microcosm may be seen as an umbrella environment, allowing the user to make links from documents in one application package to documents in another application package. Microcosm also has the ability to generalize source anchors and to create generic links which can be followed from any document or application. Generic links are of considerable benefit to the author in that a new document may be created or imported into the application and have immediate access to all the generic links which have been defined for the system. The model can be extended to incorporate encoded material as required for Project Electra. This includes the automatic generation of links from encoded documents. For example, the bio-bibliographical database is being developed using a relational

---

\(^8\) Microcosm has been developed by Dr Wendy Hall, Department of Electronics and Computer Science, University of Southampton. A number of multimedia applications have been designed in Microcosm, including the Mountbatten Archive, which contains text, photographs, audio and video materials taken from the Mountbatten Archive which resides at Southampton University.
database package running under Windows. The database can pass information to Microcosm through the Dynamic Data Exchange mechanism which allows data transfer between Windows applications. Initial database entries are in SuperBase, but this may change as more software becomes available for Windows. The information is being structured to ensure its portability between database packages.

It is important to both Project Electra and the Women Writers Project that none of their specific applications – the production of customized hard copy, general purpose network access, and incorporation into large instructional and research hypertext systems – will alter the original databases. Each is merely a specialized processing of the existing databases which leaves them unchanged. Not only is this approach simple, powerful and flexible, it can exploit future progress in computing software and practices without sacrificing the initial investment in analysis and coding. Importantly, too, it does not require that specific scholarly needs be determined in advance. This approach – bringing specialized tools to bear on a general-purpose database – supports creative scholarship without preempting or prejudicing it. In this sense, both projects are committed to affirming the subsidiary significance of the computerization: the electronic environment is a dynamic environment and a new and exciting conceptual medium, encouraging the re-cognition of text and images; but the relationship remains weighted in favour of the scholarly material as the main subject of enquiry and the computer as the means to that end.9

ABOUT THE AUTHOR

KATHRYN SUTHERLAND is Director of Project Electra. She has written on Hannah More, Walter Scott and Adam Smith, among other topics, and is currently completing a book on women and the institutions of history in the eighteenth century.

CONTACT ADDRESSES

Post: Dr Kathryn Sutherland, Department of English, University of Manchester, Manchester, M13 9PL
E-mail: ELECTRA @ UK.AC.OX.VAX

9 The section on the Women Writers Project has been compiled from information kindly provided by the Project Manager, Ms Elaine Brennan of the Department of English, Brown University, Rhode Island, U.S.A.