
Electronic Texts in English and American Literature

ANITA LOWRY

ABSTRACT

THE PAST FIVE YEARS has seen a rapid increase in the publication of electronic texts in all humanities disciplines including English and American literature. Librarians need to know more about the characteristics and use of electronic texts and text banks in order to meet the changing expectations of students and scholars. To that end, this article will examine the major electronic editions of Shakespeare's work and will survey selected electronic text publications currently available in English and American literature.

INTRODUCTION

Over twenty years ago, in a paper presented to the World Shakespeare Congress (Vancouver, British Columbia, 1973), the prominent textual critic T. H. Howard-Hill (1971) argued that, "the establishment of a generally accessible computer-readable Shakespearian text is the first necessary step towards the most useful computer study of Shakespeare" (p. 54). He proposed the creation of a database of old- and new-spelling texts of Shakespeare along with appropriate software for use by the "computationally-naive scholar" (p. 54) as a way of meeting the scholars' and students' need for "easy and fast access to Shakespearian materials on magnetic tape" (p. 54).

While Howard-Hill's vision of a comprehensive Shakespeare textual database is yet to be realized, the microcomputer revolution has made possible a proliferation of electronic editions of

Anita Lowry, Butler Reference Department, Electronic Text Service, Humanities and History Division, Columbia University Libraries, 325 Butler Library, 535 W. 114th Street, New York, NY 10027

LIBRARY TRENDS, Vol. 40, No. 4, Spring 1992, pp. 704-23

© 1992 The Board of Trustees, University of Illinois

Shakespeare's works, each promising "easy and fast access" for scholars investigating images, themes, motifs, linguistic features, and structure in his work. The many commercially available electronic versions of Shakespeare are emblematic of the explosion that has occurred since the mid-1980s in the publication of electronic primary source texts in the humanities. Today hardly a month goes by without the announcement of a new electronic text publication (or a new edition of an "old" one) in literature, philosophy, religion, or classics; many of these electronic texts represent important works in English and American literature.

This article will examine the major electronic editions of Shakespeare's work and will survey other electronic text publications currently available in English and American literature. It will also take a look at how these electronic resources are beginning to change the perceptions and expectations of librarians and library users alike.

PUBLISHED VERSUS UNPUBLISHED TEXTS

This discussion will concentrate primarily on published electronic texts. Published electronic texts are those

whose goals and processes of production and distribution parallel the goals and processes that we associate with formal publication. That is, the electronic text is intended for distribution or sale to an audience beyond its creator. The text and analytical software that may be supplied with it are subject to editorial control to ensure completeness, accuracy, and usability. The text is documented regarding source, modifications made, encoding and mark-up or tagging schemes used in the text, and so on. Instructional manuals are provided. Technical support for the text and accompanying software is available from the publisher. (Lowry, 1990, p. 16)

While publishers of electronic texts vary greatly in the degree to which they observe these standards, most meet at least some minimum standards that facilitate the acquisition and use of their texts.

In addition to published electronic texts, there are thousands of unpublished electronic texts "created by an individual or a research project without being intended for distribution beyond...[their] initial users" (Lowry, 1990, p. 16). Because unpublished electronic texts are difficult to identify, acquire, and use, this article will not deal with them, except for those preserved in, and distributed by, the Oxford Text Archive (Oxford University).

PUBLISHERS AND PUBLICATIONS

The first publishers of electronic texts and analytical software in the humanities were nonprofit scholarly editing projects or centers like the *Thesaurus Linguae Graecae* (TLG) at the University of California at Irvine; the American and French Research on the *Treasury of the French Language Project* (ARTFL) at the Centre

National de la Recherche Scientifique and the University of Chicago, the Center for Computer Analysis of Texts at the University of Pennsylvania, and the Humanities Research Center at Brigham Young University. There are now several hundred nonprofit scholarly editing projects around the world devoted to the creation of electronic texts or text banks in the humanities. Some of these distribute texts themselves while others are collaborating with a growing number of commercial publishers interested in electronic publishing.

In the world of commercial publishing, companies like InteLex Corp. (Clayton, Georgia.), Johnson & Company (American Fork, Utah), and Shakespeare-On-Disk (Clinton Corners, New York) specialize in the publication of electronic texts and software. They are now being joined by such print publishers as Oxford University Press, Niemeyer Verlag, Brepols, and Chadwyck-Healey, which are adding electronic texts to their lists. With the exception of Brepols and Niemeyer, all of these firms publish electronic texts of interest to scholars of English or American literature, including InteLex, which publishes an excellent series of works by English philosophers.

Sources of Information about Electronic Texts

The Humanities Computing Yearbook, 1989-1990 (1991) is the best, most comprehensive, source of information about electronic texts and analytical software in all disciplines of the humanities, though inevitably an annual publication is out of date as soon as it is published. Other sources include the journals *Computers and the Humanities* and *Literary and Linguistic Computing*, the newsletter *Bits & Bytes Review*, the network discussion group HUMANIST, *The Georgetown University Catalogue of Projects in Electronic Text*, and a recent article by John Price-Wilkin (1991) of the University of Michigan Library. The "computer files" file in the Research Libraries Information Network (RLIN) cataloging database contains records for many electronic texts, primarily the unpublished holdings of the Oxford Text Archive. And the new Center for Electronic Texts in the Humanities, a cooperative project of Princeton and Rutgers universities, intends to play a national role in gathering and disseminating information and resources.

HAMLET: "WORDS, WORDS, WORDS"

—*The Tragedy of Hamlet* Act II, Sc. II, Line 192

(Riverside Ed. on the *WordCruncher Disc*, Vol. 1)

There are countless megabytes of Shakespeare's words in electronic form and at least nine published electronic versions of some or all of his works. By examining some of these publications,

we can begin to understand not only the differences among them, but also the nature and characteristics of electronic texts in general. The following are some of the major characteristics that librarians must consider when evaluating electronic texts and must understand in order to assist users of these texts:

1. *Quality of the Text*: Which printed edition is being reproduced by the electronic text? Is it a standard edition? Does the electronic version omit those parts of the text not written by the author (for example, introductions, critical notes and annotations, etc.) or reproduce everything? Has the text been carefully proofread to match the source text? Have errors found in the source text been corrected in the electronic text or left so that it will exactly match the printed text? Is the editor of the electronic text a scholar or specialist who is knowledgeable about the texts and about scholarly editing standards?
2. *Software*: Is search software supplied with the text, and, if so, what are its capabilities? Can the software also be used separately from the text it comes with, and, conversely, can the text be used with different software? What kind(s) of encoding or markup does the software permit or require?
3. *Markup*: Does the electronic text contain encoding or markup of any kind? What is the nature of that markup, and what software is it compatible with?¹ Does it use the Standard Generalized Markup Language (SGML)?
4. *Medium*: Is the text distributed on CD-ROM, diskette, magnetic tape, or online via dial-up access? On what microcomputer platform does the software run?
5. *Documentation*: How well are all of the above characteristics documented in printed and online documentation?
6. *Price*: What does the electronic text and/or software cost? Are site or network licenses available?

Given these criteria for evaluation, how do the electronic Shakespeares stack up?

WORDCRUNCHER SHAKESPEARE

WordCruncher Shakespeare actually refers to two slightly different electronic publications of the Riverside edition of the *Complete Works of William Shakespeare*, one on CD-ROM and one on diskette, both of which are searched using the software program WordCruncher. Originally produced by the Electronic Text Corporation, the WordCruncher Shakespeare texts are now distributed by Johnson & Company (American Fork, Utah). The Riverside edition, edited by G. Blakemore Evans and published by Houghton-

Mifflin in 1974, is a standard and widely used edition of Shakespeare's plays and poems. The electronic text was scanned and proofread, but occasional errors do appear. For example, Kenneth Steele (1990) has identified seventy typographical errors introduced into the WordCruncher Shakespeare.

The WordCruncher Shakespeare is included, along with a number of other texts, on a CD-ROM titled *The WordCruncher Disc, Volume 1* (1990); it runs under DOS (\$249. Network licenses are available). The WordCruncher software (WCView) supplied with this CD-ROM is menu driven, interactive, and manages to be both easy to learn and use as well as sophisticated and powerful.

A user may search through all of Shakespeare's works at once, through an individual work, or through all the works in a particular category (comedies, tragedies, histories, and romances and poems). WordCruncher automatically displays an alphabetical list of all the words in the text (including titles, character lists, and stage directions) from which one can easily select words to search. A rich complement of full-text searching capabilities is at the user's fingertips—left and right truncation; searches for single words or groups of words; Boolean AND, OR, NOT combinations; proximity searches specifying the order of words and the desired context (within the same act and scene, within the same line, or within x number of characters). The displays of search results are clear, highlighting the search words and precisely identifying the location of the terms in the text; the number of lines displayed as context can be determined by the user. By moving the highlighting bar to an occurrence in the list and pressing the enter key, the user opens a full-screen window onto the text and may, from this point, read forward or backward throughout the text as he or she wishes.

The ease and quickness of this process—from formulating the search to viewing the hits in context to opening a wide window onto the text, moving easily back and forth—make possible a new way of experiencing the text, one that does not replace traditional ways of reading but that supplements them. Students find that it enables them to explore ideas or hunches about the texts that they might not otherwise have pursued and that, indeed, the process of moving so smoothly back and forth between word and text, between question and answer, between conceptualization about the text and immersion in the text itself is stimulating and thought provoking.

In addition to its excellent search, display, and output capabilities, WordCruncher has an interesting feature for showing frequency distribution and relative frequency distribution of a search word or combination of words across a group of texts or within a particular text. So, for example, not only can one find all the occurrences of

words relating to the image of masks in Shakespeare, but one can also quickly determine in which plays, acts, and scenes that imagery is concentrated—useful not only for identifying key parts of the works to examine most closely but also, perhaps, for finding patterns or structural significance in the placement of that imagery within a particular play or group of plays. Other features of the software include a program to create printed concordances or indexes from the works, the ability to go directly to a specified point in the text for reading, and the ability to save search strategies and result lists to disk for recall later. Search results can be sent to disk or printed in a variety of formats.

WordCruncher software supports up to three hierarchical levels of structural markup. In the Riverside Shakespeare electronic texts, those levels correspond to work, act/scene, and line. Thus these are the structural entities that the software can identify in the text and that can be specified as the context when searching for combinations of words within the text(s); in other words, the location of each hit in a search is identified according to the work it appears in, the act/scene numbers, and the line number, and users may specify that Boolean combinations of words appear within the same act/scene or within the same line (or within x number of characters of one another).

Documentation is the weak point of the WordCruncher Shakespeares. The pamphlet supplied with the CD-ROM does not adequately explain the more advanced features of the software, while the full manual that accompanies the stand-alone version of WordCruncher (WCView and WCIndex) is thorough but poorly organized, scattering information on a single topic among different sections of the manual. Fortunately the searching program itself (WCView) is well-organized and consistent and, with a few exceptions, the instructions for its most basic and frequently used operations are presented clearly on the menus; online help is available.

The WordCruncher software may also be purchased in a stand-alone version (\$299 for DOS machines. Network licenses available) consisting of two main programs: WCView (features are essentially the same as those of the WCView program on the CD-ROM) and WCIndex, an optional program that enables the user to index his or her own ASCII text files for searching with the WCView program. Johnson & Company also sells on diskette a number of the texts from the *WordCruncher Disc* CD-ROM, including the Riverside Shakespeare (complete works are \$300; subsets are \$100 each); these texts are encrypted so that they can only be read by WordCruncher.

With its high quality texts, excellent software, and an amazing price of \$249, the *WordCruncher Disc* (1990) CD-ROM is quite a

bargain. While a number of libraries have found out how easy to use, valuable, and popular the *WordCruncher Disc* can be, it deserves to be more widely known and available as a reference tool and as a unique resource for examining not only the works of William Shakespeare but also the other texts included on the disk (see *WordCruncher* texts discussed later).

THE OXFORD UNIVERSITY PRESS ELECTRONIC SHAKESPEARE

In 1986, Oxford University Press (OUP) published a new edition of Shakespeare's complete works under the general editorship of Stanley Wells and Gary Taylor. From this modern-spelling edition an electronic edition was prepared jointly by William Montgomery (one of the editors of the printed edition) and Lou Burnard (Oxford University Computing Service) and published by OUP in 1989 (\$300; educational site licenses are available). A meticulously produced and documented electronic text, *William Shakespeare: The Complete Works, Electronic Edition* (1989) (hereafter the OUP Electronic Shakespeare) is a model to be emulated for the scholarship, accuracy, and thoroughness of both text and documentation; the twenty-six-page manual that accompanies the text explains in ample detail the preparation of the text, the meanings of all special characters, the markup language, and how it may be used with the Micro-OCP software. However, it is a very different type of electronic text from the *WordCruncher Shakespeare*. Examining those differences will illuminate several important issues relating to how electronic texts may be used, the nature and function of markup in electronic texts, and the challenges these texts pose for librarians.

The OUP Electronic Shakespeare consists of plain ASCII text files—a separate file for each play, the sonnets, and other poems—on floppy diskettes formatted for DOS machines. Although these files can be read by word processors and other programs, they are designed to be used with a stand-alone program called Micro-OCP, also published by Oxford University Press (\$295; multiple-copy discounts and educational site licenses available). The Micro-OCP and *WordCruncher* programs are as different as night and day, and Micro-OCP betrays its roots in the mainframe software Oxford Concordance Program (a widely used mainframe program for producing concordances and analyzing texts). A rarity in the microcomputer environment, Micro-OCP is a batch program rather than an interactive one. The user of Micro-OCP must master a command language for writing short programs (called command files) that describe the text to be processed and define the actions to be performed on the text. Fortunately the program has a menu, a number

of default command values, and error trapping to assist the user in creating and “debugging” the command files. After carefully typing in the command file and pressing the F10 key to begin its execution, the user waits while the program runs and writes the results to a disk or prints them to a printer. Depending on the speed of the microcomputer, the size of the file being processed, and the complexity of the command file, this can take many minutes or even hours. When the process is finished, the user may read the results on the printout or view the results file on the disk.

The OUP Electronic Shakespeare and Micro-OCP obviously are not tools for casual users or for interactive interrogation of a text—purposes for which the WordCruncher Shakespeare is superbly suited. Who and what, then, are they good for?

One key to the unique capabilities of the OUP Electronic Shakespeare lies in the extensive markup that has been added to the electronic text. This markup not only identifies a wide range of significant features of the text but also supplies additional information about the text. The markup scheme identifies:

- author (including other names besides Shakespeare in cases of divided or uncertain authorship)
- title of each work
- relative order of composition in Shakespeare’s oeuvre
- date of composition
- kind or genre of the work (play, narrative poem, sonnet, commentary, or miscellaneous)
- line numbers (corresponding to the printed edition)
- amphibious and split verse lines
- act and scene
- prologues and epilogues
- act intervals
- whether text is verse, prose, or song
- stage directions
- name of speaker with text of his/her speech

Using these markup codes, Micro-OCP can search, sort, and statistically analyze a file of the OUP Electronic Shakespeare with great precision. For example, while WordCruncher can find all the occurrences of the word “king” in *Hamlet*, it cannot limit its search to find only those occurrences in the speech of a particular character like Hamlet himself; an abbreviation for the speaker’s name does, of course, appear in the text of the WordCruncher Shakespeare, but there are no markup codes to indicate where that speaker’s words begin and end. In contrast, Micro-OCP can be instructed to limit its search for the word “king” only to the text of the speeches of

Hamlet because the markup codes define the extent of his speeches. To give another example, the search for "king" in the WordCruncher Shakespeare will retrieve that word in stage directions, character lists, speakers' names, etc., as well as in the actual text of the play; with Micro-OCP and the OUP Electronic Shakespeare, the user may specify that "king" NOT appear in these parts of the text, or, conversely, may limit the search to occurrences in one or more of these special parts of the text exclusively. However, Micro-OCP can only search through one file—and therefore one Shakespeare play—at a time and cannot merge the results of searches in different files.

In addition to specifying exactly which parts of the text to search, the user has great discretion (including right and left truncation) in defining the words or character strings, the ranges of words, and the phrases or collocations for Micro-OCP to find. Micro-OCP is, essentially, a concordance-generating program that creates lists of words in one of the following formats: a list of words with or without frequency of occurrence; an index of words with their frequencies and location in the text but without any context; or a concordance of words with references, context, and frequency. The user can determine the order of the words (for example, alphabetically by beginning or ending of the words, or according to frequency) and nearly all aspects of the layout of the final lists.

Micro-OCP can also produce a simple statistical table appended to a word list, index, or concordance. This table shows the total number of words retrieved (tokens); the total number of unique words retrieved (types); the ratio of unique words to total words (type/token ratio—a measure of vocabulary richness); the frequency of words (how many occur once, how many twice, etc.); and the relative frequency of words (for example, words occurring once, twice, etc., make up what percentages of the total).

WORDCRUNCHER SHAKESPEARE VERSUS OUP ELECTRONIC SHAKESPEARE

While we might think of WordCruncher as opening interactive windows onto a text in order to highlight the words we are looking for and frame them in a larger context, we could think of Micro-OCP as pulling words out of the text in order to count and arrange them into a fixed order depending on the characteristics of the words themselves. Both WordCruncher and Micro-OCP will locate a word or combination of words in the works of Shakespeare and indicate where it appears in their respective editions, but only WordCruncher will find it in all the works at once, and only WordCruncher will take you there.

The WordCruncher Shakespeare meets the needs of most students and faculty who want to locate words in Shakespeare's works in order

to understand better the meaning and expression of themes, images, motifs, etc.; its major disadvantage is its inability to limit a search to the speech of a particular character or to eliminate stage directions from the search. The OUP Electronic Shakespeare and Micro-OCP meet the more specialized needs of scholars doing extended close analysis of Shakespeare's vocabulary or of the words of particular characters. The text and program are not easy to learn or use and do not constitute a reference tool, as most libraries understand that term, but they certainly constitute a significant electronic research tool. Like complex numeric data files, with which social science data librarians have dealt for some time, texts and programs like the OUP Electronic Shakespeare and Micro-OCP have a place in research collections.

FOLIOS AND QUARTOS

As T. H. Howard-Hill (1973) has pointed out:

The old-spelling and modern-spelling computer-readable texts commend themselves to investigators for different reasons. For instance, a computer-aided study of the imagery of some of Shakespeare's plays should probably use the modern-spelling text from which, so far as this can be done by an editor, errors, obscurities and inconsistencies have been removed. On the other hand, investigations into textual or orthographical questions (and some other matters of linguistic interest) should probably use the old-spelling text where peculiarities of text and language supply the scholar with some of his most significant evidence. (p. 53)

Electronic versions of fifty-five of the Shakespeare folio and early quarto texts² have been preserved in the Oxford Text Archive at Oxford University and can be purchased from the archive by scholars for approximately the cost of reproduction on magnetic tape, shipping, and handling. Like their printed counterparts, these texts are not for amateurs, and any individual or library contemplating them should read Whitney Bolton's analysis of the daunting problems relating to the editing, accuracy, and reproduction of the electronic folio and quarto texts. These problems are complicated by the lack of full documentation for the texts, variations in the markup codes used in them, and difficulties in converting them from the mainframe tapes on which they are supplied onto diskettes for use on a microcomputer (Bolton, 1990).

OTHER ELECTRONIC EDITIONS OF SHAKESPEARE

Not surprisingly, the number of published electronic versions of Shakespeare is surpassed only by electronic versions of the Bible. In preparation for this article, seven electronic Shakespeares³ were reviewed including the CD-ROM and diskette versions of the WordCruncher Shakespeare. Among them, the WordCruncher Shakespeare and the OUP Electronic Shakespeare are without rival

because of the importance of the editions that are reproduced, the quality of the texts and markup, the thoroughness of the documentation, and the sophistication of the software. The other electronic Shakespeare texts are, essentially, consumer items rather than reference or research tools, and each of these has significant, if not fatal, flaws.

SURVEY OF ELECTRONIC TEXTS IN ENGLISH AND AMERICAN LITERATURE

Shakespeare alone does not a library make. And yet, to date, relatively few other English and American literary texts have been published in electronic form. Scholars and students of English and American literature do not enjoy a substantial body of reliable electronic texts as do their colleagues in classics (who have the *Thesaurus Linguae Graecae* and the Packard Humanities Institute's [Los Altos, California] classical Latin CD-ROMs) or in French literature (who have the *ARTFL* online database whose pre-1925 texts are soon to be released on CD-ROM). However, recent announcements of forthcoming electronic publications by Chadwyck-Healey and Oxford University Press promise to increase dramatically the number of high quality English literary electronic texts.

Oxford University Press and Chadwyck-Healey

The OUP and Chadwyck-Healey publications present rather different approaches to electronic publishing in the humanities. OUP is initiating a series of separately published texts called The Oxford Electronic Text Library and has announced individual or collections of texts by the following authors for publication in 1992—Austen, Chaucer, Coleridge, Dickens, Wordsworth, Hardy, Matthew Lewis, Milton, Mary Shelley, and Mary Wollstonecraft. All electronic editions are based on printed editions by OUP (except for the *Riverside Chaucer* by Houghton-Mifflin) and, one hopes, will achieve the high standards set by the OUP Electronic Shakespeare texts and documentation. However, two noteworthy differences between these new texts and the OUP Electronic Shakespeare promise to make these even more useful and more significant as milestones in the publication of electronic texts. First of all, each text is encoded according to the Text Encoding Initiative (TEI) recommendations using the Standard Generalized Markup Language; thus the markup of these texts will be consistent among them and consistent with texts from other publishers using these standards—and they may be analyzed by any software that can recognize and exploit SGML markup. Second, OUP is publishing in fall 1992 the Oxford Text Analysis System (OTAS), an interactive SGML-based program for microcomputers for use with

these texts and with other ASCII texts containing SGML markup. Priced at around \$100 each and formatted for IBM PC compatibles or Apple Macintosh computers, the publications in The Oxford Electronic Text Library are within the reach of individuals as well as institutions (site licenses available). With the marketing force of OUP behind it, this series is likely to bring electronic texts to the attention of a broad audience of scholars, students, and librarians and make them increasingly aware of what can be done with these new tools for research and teaching.

Never one to shy away from vast projects, Chadwyck-Healey has plunged into electronic text publishing with two enormous databases—the *Patrologia Latina Database* and *The English Poetry Full-Text Database*, both announced for publication beginning in 1992. The *English Poetry Full-Text Database*—encompassing the poetical works in English of 1,350 English, Welsh, Irish, and Scottish authors from 600 to 1900—has no counterpart in print or among electronic publications in English literature. It will make possible innovative new literary, linguistic, and historical studies not only of the works of individual authors, but across authors, time periods, types of poetry, etc. Laudably, Chadwyck-Healey has also adopted TEI-compliant SGML for encoding its texts, adding another important publisher's voice to the support for these standards. Unfortunately, its impact on scholarship and on the development of the electronic library is likely to be mitigated by its extraordinary cost. At prices ranging from \$37,500 to \$54,000 for CD-ROM (with text analysis software) and \$44,250 to \$72,000 for magnetic tape (without software), the complete *English Poetry Full-Text Database* will be, by many thousands of dollars, one of the two most expensive electronic text databases in any field of the humanities, equaled only by Chadwyck-Healey's *Patrologia Latina Database* (\$60,000 for either CD-ROM or magnetic tape). As a result, few institutions will be able to make these resources available to their patrons. One wonders why the production costs of these admittedly ambitious databases might not be recouped by selling significantly more copies at a significantly lower price—a strategy that would encourage the development of institutional environments to support electronic research in the humanities to the ultimate benefit of electronic publishers as well as libraries, students, and scholars.

WordCruncher Texts

The WordCruncher Disc Vol. 1 has little competition as the best collection of texts in American literature. The many strengths noted in the review of the WordCruncher Shakespeare (discussed earlier) apply equally to the other texts on this CD-ROM, including works

by the following American authors—Cather, Emerson, Faulkner, Franklin, H. James, Jefferson, Hawthorne, London, Melville, Thoreau, Twain, Whitman (all are also available individually on diskette). These texts are produced from the Library of America editions and, for the most part, contain selected rather than complete works of the authors. The CD-ROM also includes *The Constitution Papers*, a collection of forty-two seventeenth, eighteenth, and early nineteenth century historical documents relating to U.S. Constitutional history—the *Declaration of Independence*, *Constitution*, *Federalist Papers*, colonial documents and early state constitutions, speeches, *Common Sense*, *The Virginia Statute of Religious Liberty*, etc. The Riverside Shakespeare, the King James and New International versions of the Bible, the *Constitution Papers*, the Library of America texts (and a few miscellaneous speeches and documents)—along with the excellent WordCruncher software, which makes the WordCruncher Disc a valuable and popular reference and research tool for high school, undergraduate and graduate students, and faculty.

The Library of the Future

Because of its sheer size and scope, *The Library of the Future Series Second Edition* CD-ROM (1991) straddles the fence between consumer item and reference tool. It contains 970 works (including many short stories) by over 100 authors from A (Aeschylus) to W (Wollstonecraft), all in English or English translation. Selected works by forty-one English and twenty-two American literary authors, essayists, and philosophers are included on the disk. In addition to these texts, *The Library of the Future* CD-ROM contains many other important works of world literature, philosophy, religion, and science.

With a few exceptions, the texts are produced from out-of-copyright printed editions, which are not identified (only the translator's name gives some clue to the editions of the translations). An obvious typographical error in one of the titles ("Women" is substituted for "Woman" in Mary Wollstonecraft's *A Vindication of the Rights of Woman*) makes one wary of the accuracy of the texts. Its software is menu driven and easy to use but is much more limited than WordCruncher—for example, it allows the user to find a word or combination of words in an individual text, in a selected list of texts, or in all the text(s) but does not show the number of occurrences in each or readily identify their exact location in the works. Because it does not use inverted file indexes to the texts, even simple searches can be extremely slow. Options for printing are limited to a screen at a time or to the entire work. Although the quality and documentation of its texts and the functionality of its

software leaves much to be desired, nevertheless this CD-ROM makes a large and eclectic selection of texts readily available for reference purposes.

Milton on Disk

Shakespeare On Disk has also published an edition of the complete poetic works of John Milton in English (\$97 on diskette). The old-spelling editions of 1645, 1671, and 1674 (safely out of copyright) were used to produce the electronic versions under the editorial direction of a Milton scholar, Roy Flannagan of Ohio University. The spelling of many words in these old editions is unusual and inconsistent—perhaps not a problem for a Milton specialist but certainly a hindrance to successful word searching by nonspecialists unfamiliar with the variant spellings. Like this company's Shakespeare texts, *Milton On Disk* (1990) contains no markup, but, because it represents the only published electronic edition of Milton's work, some libraries have found it worth the effort to add structural markup so that the texts may be used with programs like WordCruncher.⁴

Distek Literary Series

Distek Publishing, Inc. has published inexpensive editions of *Wuthering Heights*, *The Mayor of Casterbridge*, *Pride and Prejudice*, *Huckleberry Finn*, *Persuasion*, *Heart of Darkness*, and *Tess of the Durbervilles*; *Turn of the Screw* is in preparation. The texts are encrypted so that they may be read only with Distek software which is quite primitive.

DiscLit: American Authors

The vast majority of electronic texts in all fields of the humanities are primary source texts—works of literature, philosophy, religion, political theory, etc. However, G. K. Hall has put 143 volumes of secondary sources—the bio-critical studies in the Twayne United States Authors Series—onto the *DiscLit: American Authors* (1990) CD-ROM, along with the OCLC American Authors Catalog, a bibliographic database of over 127,000 citations to related books, serials, audiovisual materials, and manuscripts from the OCLC Online Union Catalog. The search software offers nearly all the options one could want for searching and moving around in these databases; with its pull-down menus, on-screen prompts, and context-sensitive help it is reasonably easy to use. Whether or not a comprehensive keyword index to the Twayne United States Authors Series and a subset of OCLC is worth \$995 is another question.

Other texts not examined for this article include *Sherlock Holmes On Disc* CD-ROM (\$99 from CMC ReSearch, Inc., Portland, Oregon); *Greatest Books Ever Written*, a multimedia CD-ROM (\$295), and *Monarch Notes* CD-ROM (\$99 from the Bureau of Electronic Publishing, Parsippany, New Jersey).

ELECTRONIC TEXTS IN LIBRARIES

The electronic revolution is changing scholars' expectations about how information is organized, manipulated, and delivered. Word processing has given students and scholars in the humanities unprecedented control over their texts and introduced them to simple word search operations. Library online catalogs and CD-ROMs have given them "hands on" exposure to the power of computer-based information retrieval. Words like "keyword" and "wildcard" are creeping into their vocabulary. Scholarly associations and publishers are bringing new electronic resources to their attention, and they are beginning to imagine the possibilities inherent in "dynamic texts":

Edited texts available in computer-readable text, data and image archives have become "active", "powerful", "dynamic", and "volatile" not just by virtue of the speed with which we exchange them over computer networks. The software that we use to access these texts transforms them before our senses. The text and its translations, the copy text and its variants, or the book and the images, texts, and sounds to which it is linked hypertextually by way of commentary, indebtedness and influence are superimposed on one another; we experience a kind of multiple vision. Just as remarkable are the many style- and content-analysis programs that may now co-exist with the text or data and through which they may be transformed into frequency tables, distribution graphs, and the like. (Lancashire, 1989, p. v)

These assertions can be illustrated by several concrete examples of the growing interest in "dynamic texts," drawn at random from the experience of three libraries that have made substantial commitments to providing access to electronic resources in the humanities.

In 1989, a graduate student came to the Butler Reference Department, Columbia University Libraries, to ask if the complete works of Immanuel Kant, in German, were available on CD-ROM (he had been using the Wilsondisc *Humanities Index* and imagined being able to search Kant's texts in similar ways)—the answer was "no". In 1991, when a faculty member asked if the library would acquire the new electronic edition of Kant's *Gesammelte Schriften*, the answer was: "We just received it; would you like to have a demonstration?" He brought ten graduate students and faculty members from the Philosophy Department with him to the demonstration, and the Kant texts (searched with WordCruncher) have since been used by both graduate and undergraduate students in the Electronic Text Service.⁵

In the Dana Center⁶ of the Barnard College Library (Columbia University), the *WordCruncher Disk* is one of the most popular CD-ROMs in the library. In the fall semester of 1991, two Barnard history professors gave class assignments exploring various political and philosophical concepts expressed in the *Constitution Papers*—one of the first fruits of the library's ongoing efforts to integrate electronic texts into the undergraduate curriculum.

In 1991, the 300 registered users of the UMLIBTEXT⁷ collection of electronic texts on the University of Michigan campus network logged an impressive 3,817 accesses to the system, which includes a large body of English and American literary texts.

The development of library facilities and projects like these for electronic texts is predicated upon several basic assumptions:

- that rapid growth in electronic primary source publishing in the humanities is imminent;
- that electronic texts and related computer-based research tools will become increasingly important for scholarship and teaching in the humanities; and
- that the library has a natural and central role to play in collecting, preserving, and providing service and instruction for these new resources.

The first assumption is being borne out by the many electronic text publication projects in progress around the world in all fields of the humanities including those surveyed in this article.

The second assumption is being borne out by the experiences of libraries whose patrons increasingly are asking about electronic textual resources and are making use of those that are available. In the experience of the Electronic Text Service (ETS), for example, people are using these to examine a concept, image, or motif in a single work: the work of an author or in a group of works related by time period and/or genre; investigate possible allusions or references in one work to another; locate a particular phrase or quotation; or identify, date, or reconstruct fragmentary Greek papyri. And, to date, the most frequently used text banks in the ETS are the *WordCruncher Disc*, the *TLG*, the *PHI CD-ROM #6* (Greek inscriptions and papyri), the *PHI CD-ROM #5.3* (ancient Biblical and classical Latin texts), *ARTFL*, and the *CETEDOC Library of Christian Latin Texts*; the Riverside Shakespeare and the King James Bible are the most popular individual titles.

The third assumption seems self-evident, a necessary and logical extension of the library's traditional roles—which is not to say that acquiring, preserving, organizing, providing access to and services for electronic texts and research tools will be simple. Librarians

seeking to expand their conception of the electronic library to encompass nonbibliographic electronic resources face many challenges—among them the fact that librarians have a great deal to learn about the characteristics and use of nonbibliographic electronic resources in the humanities, and that there are, as yet, relatively few sources for in-depth evaluations of them. This article has attempted to take a first step toward meeting those challenges.

NOTES

¹ *Descriptive markup* is extremely important in electronic texts because it categorizes the parts of a text so that analytical software may recognize those parts, much the same way that field labels (MARC tags, for example) in a bibliographic database define bibliographic structure. Descriptive markup is necessary because, in the words of the Text Encoding Initiative:

A text is not an undifferentiated sequence of words, much less of bytes. For different purposes, it may be divided into many different units, of different types or sizes. A prose text...might be divided into sections, chapters, paragraphs, and sentences. A verse text might be divided into cantos, stanzas, and lines. Once printed, sequences of prose and verse might be divided into volumes, gatherings, and pages. Such units are most often used to identify specific locations or reference points within a text...but they may also be used to subdivide a text into meaningful fragments for analytic purposes.... Other structural units are more clearly analytic, in that they characterize a section of a text. A dramatic text might regard each speech by a different character as a unit of one kind, and stage directions or pieces of action as units of another kind. Such an analysis is...useful...for facilitating comparisons between the words used by one character and those of another, or those used by the same character at different points of the play (*Guidelines*, 1990, pp. 11-12).

In general, the more descriptive markup in an electronic text the finer and more precise the searching and analysis that may be done using the text.

A *markup language* is "a set of markup conventions used together for encoding texts. A markup language must specify what markup is allowed, what markup is required, how markup is to be distinguished from text, and what the markup means" (*Guidelines*, 1990, pp. 9-10). There are many markup languages that have been used in electronic texts in the humanities, nearly all of them incompatible with one another; as a result, the tagging in a text encoded with one markup language cannot be read by software that is designed for another unless the codes are converted (which may not be easy to accomplish). Out of this chaos, the Standard Generalized Markup Language (SGML) has emerged as an international standard. The Text Encoding Initiative (TEI) is an international project to define text encoding guidelines and a common interchange format for literary and linguistic data using the SGML language. Increasingly, publishers of electronic texts intended for scholarly research are using TEI-compliant SGML encoding in their texts.

² These electronic texts are based on the electronic texts created by Howard-Hill during the preparation of his *Oxford Shakespeare Concordances*

published in the 1960s; some of them have markup and other changes added later.

³ In addition to the aforementioned WordCruncher Shakespeare and OUP Electronic Shakespeare they are:

- *Shakespeare On Disk*, an electronic version of the Shakespeare Head Press Edition (published early in the twentieth century); the complete plays and sonnets or selected subsets are available as plain ASCII texts without markup or software. Publisher: Shakespeare On Disk, Clinton Corners, New York.
- Distek Literary Series: Shakespeare, which offers thirteen plays, sold individually on diskette with Distek search software. Publisher: Distek Publishing Inc., Windsor, Ontario.
- Unidentified editions of the plays and poetry of Shakespeare on the *Library of the Future Second Series* CD-ROM. Publisher: World Library, Inc., Garden Grove, California.
- A version of the OUP Electronic Shakespeare that is included in the "Digital Library" that comes with NeXTstation microcomputers; Digital Librarian search software is also included. Publisher: NeXT, Inc., Palo Alto, California.

I was unable to examine the Shakespeare CD-ROM with DiscPassage search software, available for IBM PC compatibles and Macintosh computers from CMC ReSearch, Inc., Portland, Oregon.

⁴ In the Electronic Text Service I have added WordCruncher markup codes to the *Milton On Disk* texts so that they may be searched using that software. For the UMLIBTEXT project at the University of Michigan, John Price-Wilkin has added SGML encoding to these texts so that they may be searched with PAT software (Open Text Systems, Inc., Waterloo, Ontario).

⁵ The Electronic Text Service (ETS), a facility affiliated with the Butler Library Reference Department in the Humanities and History Division, was established in 1987/88 in order to enable the Columbia University Libraries to integrate electronic primary source materials into its collections and services. The ETS collection currently includes texts and hypermedia research tools in Coptic, English, French, German, Greek, Hebrew, Italian, Latin, Spanish, Tibetan, as well as translations into English from these and many other languages. Titles number in the thousands (thanks to such large text banks as *ARTFL*, the *TLG*, and others). ETS services include: a 1.5-credit course, "Research in the Humanities: A Practicum on Resources and Methods," taught in the Graduate School of Arts and Sciences; demonstrations and workshops; and individual reference and research consultations.

⁶ Established in 1990 with a grant from the Charles A. Dana Foundation, the Dana Center is a key component of the Barnard Library's program to strengthen the teaching role of the library. The Dana Center collections include electronic texts and hypertext programs in the humanities and social sciences (it was selected as a test site for the Library of Congress *American Memory Project*), and its programs focus on the integration of these resources into undergraduate research and teaching.

⁷ In 1989, John Price-Wilkin, data services librarian and English language/literature selector in the Harlan Hatcher Library at the University of Michigan, began a pioneering project to make electronic texts available over the campus network. Selecting the menu option "UMLIBTEXT" on the network now transports University of Michigan students and faculty into

an online world of Old English texts (from the Oxford Text Archive), materials from the Michigan Early Modern English dictionary, the Riverside Shakespeare and Library of America texts (from the *WordCruncher Disc*), the *Milton On Disk* texts, English philosophers (from InteLex Corp.), the King James Bible, and selected unpublished English and American literary texts (from the Oxford Text Archive and Cambridge Text Archive). In order to make texts available over the campus network, Price-Wilkin has negotiated network licenses for them from their respective publishers or owners—licenses that grant access only to University of Michigan students, faculty, and staff. The search software for UMLIBTEXT is a program called PAT, which runs under UNIX and was developed by the Centre for the New OED at the University of Waterloo. Price-Wilkin has edited all the texts, except the Old English corpus, to add SGML tagging. Thanks to the time invested by Price-Wilkin in the development of UMLIBTEXT, including a substantial amount of time spent encoding texts, the University of Michigan community has access to a significant body of electronic texts in English and American literature from inside or outside the walls of the library.

REFERENCES

- Bits & Bytes Review*. (1986-). Whitefish, MT: Bits & Bytes Computer Resources.
- Bolton, W. (1990). The Bard in bits: Electronic editions of Shakespeare and programs to analyze them. *Computers and the Humanities*, 24(3), 275-287.
- Computers and the Humanities*. (1966-).
- DiscLit: American Authors*. (1990). (computer file on CD-ROM). Boston, MA: G. K. Hall.
- The Georgetown University Catalogue of Projects in Electronic Text* (computer database compiled by the Center for Text & Technology, Academic Computer Center, Georgetown University).
- Guidelines for the encoding and interchange of machine-readable texts*. (1990). (Draft Version 1.0, 15 July 1990). [n.p.]: Text Encoding Initiative.
- Howard-Hill, T. H. (1973). A common Shakespeare text file for computer-aided research: A proposal. *Computer Studies in the Humanities and Verbal Behavior*, 4(1), 53-56.
- Humanist Discussion Group*. (Electronic listserv discussion group). Elaine Brennan & Allen Renear, Moderators. (Network address: <humanist@brownvm>).
- Lancashire, I. (1989). Forward. In *The dynamic text guide* (abstracts of papers presented at the 9th International Conference on Computers and the Humanities and 16th International Association for Literary and Linguistic Computing Conference, 5-10 June 1989) (pp. v-vii). Toronto, Ontario: Centre for Computing in the Humanities, University of Toronto.
- Lancashire, I. (Ed.). (1991). *The humanities computing yearbook, 1989-1990*. Oxford, England: Clarendon Press (the first annual volume was published in 1988).
- Literary and Linguistic Computing; Journal of the Association for Literary and Linguistic Computing*. (1986-). Oxford, England: Oxford University Press.
- Lowry, A. (1990). Machine-readable texts in the academic library: The Electronic Text Service at Columbia University. In C. C. Gould (Ed.), *Computer files and the research library* (pp. 15-23). Mountain View, CA: The Research Libraries Group, Inc.
- Milton On Disk*. (1990). (computer file on magnetic diskette). Clinton Corners, NY: Shakespeare On Disk.
- Price-Wilkin, J. (1991). Text files in libraries: Present foundations and future directions. *Library Hi Tech*, 9(3), 7-44.
- Shakespeare, W. (1989). *William Shakespeare: The complete works, electronic edition* (computer file on magnetic diskette). Oxford, England: Oxford University Press.

- Steele, K. (1990). Errors in the WordCruncher Riverside Shakespeare. *Shakespeare Electronic Conference*, 1(31), (27 August 1990). (Electronic listserv discussion group). (Network address: <shaksper@utoronto.bitnet>.)
- The WordCruncher Disc, Volume 1*. (1990) (computer file on CD-ROM). Orem, UT: Electronic Text Corporation, distributed by Johnson & Company, American Fork, Utah.