Electrifying Wordsworth:
Using SGML and HTML for dynamic collation

Ronald Tetreault, Department of English, Dalhousie University, Halifax, Nova Scotia, Canada

It has become commonplace to draw parallels between the rapid growth of digital communications systems and the introduction of the printing press over five hundred years ago. George Landow's claim (or is it a warning?) that "electronic text processing marks the next major shift in information technology after the development of the printed book" takes its cue from Marshall McLuhan's provocative comments a generation ago.(1) As that Canadian communications theorist observed, media are not transparent windows-- by their very function of mediation, they have power to reshape reality. The view of the world that the print medium conveys is likely to be quite distinct from the view made possible by electronic media, not just in content but also in structure.

At the same time we must bear in mind that the development of print technology did not occur overnight. It was a slow and gradual process that unfolded in several phases, in which the revolutionary introduction of movable types attributed to Gutenberg was a crucial but by no means ultimate innovation. The earliest printed books adopted the conventions of the medieval manuscript codex which preceded them as the chief instrument of cultural dissemination.(2) Such incunabula (that is, books printed before about 1500) often tried to imitate the written script of manuscript volumes, employing typefaces that were not easily legible and retaining the abbreviations and ligatures common in their predecessors. Only gradually were tables of contents introduced, and title-pages were a late development growing out of the practice of placing a colophon at the end of a volume. Even the practice of numbering pages evolved slowly, beginning with the numbering of leaves and only then moving to the separate numbering of recto and verso.

Though Gutenberg and his associates were active in Mainz by the mid-1450s, the conventions of the modern printed book actually evolved in Venice only later in the century, and gained impetus shortly after 1500. Here the Frenchman Nicholas Jenson began to design type-faces of great clarity and beauty, and the Italian scholar-publisher Aldus Manutius introduced many of the features we now take for granted. Aldus is justly famed for the clarity of his Greek fonts, but it was his printings of Latin and vernacular authors that set the standards for the modern book.(3)

If parallels are to be drawn between the print revolution of the early Renaissance and contemporary electronic publishing, it may more properly be said that we have passed beyond the Gutenberg stage and entered the Venetian phase of our electronic medium. No longer is hypertext a novelty nor is it peculiar to find e-texts in the form of highly-structured complex documents. Encoding methods using HTML and SGML are now widely understood, but less certain are we of Aldine values like visual clarity, scholarly accuracy, and ease of use. These are the focus of my work on the Wordsworth

digital archive I am preparing with my colleague Bruce Graver, of Providence College in Providence, Rhode Island.

Using a variety of software and an appropriate mix of SGML and HTML mark-up, Graver and I are currently engaged in creating an electronic edition of William Wordsworth's and Samuel Taylor Coleridge's Lyrical Ballads, a pioneering collection of English poetry first published in 1798. Wordsworth's poems dominate the book, at least in sheer volume, and in subsequent editions he added so many more of his own poems that he felt entitled to put his name alone on the title page. Though revisions on the part of Coleridge will be given attention in the four lifetime editions of Lyrical Ballads (1798, 1800, 1802, and 1805) that we will reproduce in full, the textual alterations made by Wordsworth to his poems are so extensive and complex that they have posed problems for editors for generations. After the Lyrical Ballads poems were subsumed into his collected works, Wordsworth continued to revise them repeatedly, until the final versions left upon his death in 1850 were often markedly different poems from those he wrote half a century before. We therefore chose Wordsworth for treatment in the electronic medium not because of his prominence in the literary canon nor because of any perceived editorial defects in previous printed editions of his works, but precisely because his restless creativity has repeatedly challenged editors to represent it adequately. Printed texts have always had to choose one version or another, except where the expedient of using facing pages to show two versions side by side has been adopted. But scholarly editors have always known that inevitably there are more than one or even two versions of any Wordsworth poem, and have resorted to the complex system of footnotes common in printed variorum editions. This critical apparatus places great demands on the reader, who must sort through, interpret, and plug in variant readings. Wordsworth is thus an ideal candidate for this project, because his habits of revision stretch the print medium to its limits and call forth the potential of the new electronic medium to represent change.

Wordsworth was fond of claiming that "poetry is the spontaneous overflow of powerful feelings", but his practice of inveterate tinkering was an equally important part of his creative process. The question for his editors has always been whether to prefer the early versions of poems closest to their inspirational source or to rely on the final authorized texts. Spontaneity was highly regarded by the poets of the Romantic movement, though it might as easily be argued that perfection is the work of time. But there are also the versions in between, where a poem finely polished may be marred by subsequent over-refinement. The choice of copy-text has always been the editor's dilemma: establishing a text in print has always meant that we must privilege one version over others, and settle for a static representation of what might be better understood as a dynamic process.

Ernest de Selincourt's landmark edition of 1940-49 confirmed the authority of Wordsworth's final texts. "No poet ever paid more meticulous or prolonged attention to his text than Wordsworth," wrote de Selincourt (4) in justification of choosing the last lifetime edition of 1849-50 as his copy-text, though he was careful to indicate evidence of development by citing early manuscripts and printed volumes in an apparatus criticus at the bottom of the page. Unease with this portrait of the poet in old age has become more

acute among his recent editors, who have attempted a snapshot of a Wordsworth closer to the moment of inspiration by choosing to base their text on the earliest completed version of the poems. This movement is exemplified by the Cornell Wordsworth, a series of scholarly editions begun in 1975 and ongoing. This edition aims "to bring the early Wordsworth into view", according to General Editor Stephen Parrish, by printing "clean, continuous 'reading texts' from which all layers of later revision have been stripped away."(5) An edition sharing the same editorial inclination, but meant for students, has appeared in the Oxford Authors series, which the publisher says aims to restore the "original identity" of the poems.(6)

But the idea of a poem's "identity", like that of the poet's, is destabilized when we begin to regard it as a work in progress. In his recent book, <u>Revision and Romantic Authorship</u> (Oxford University Press, 1996), Zachary Leader argues that the protean Romantic self cannot be captured in the earliest versions of poems. All writers revise their texts, and perhaps none more so than Wordsworth. In a review of this book, Frank Kermode writes that Leader

wants to know what notions of identity underlie the assumption that a poet in his twenties could be identical with the poet who, in his seventies, was still tinkering with his early writings, as if they were essential to the expression of the singleness of a life or a life-work, rather than leaving them alone as virtually the work of a different person, or at any rate of a person in no need of being assimilated to a later one.(7)

Revision for Leader is a vital part of the creative process, and reveals a complexity in acts of composition that are at odds with comfortable notions of a unitary self. It is dubious not only to think that age is best which is the first, another instance of that bias Jerome McGann warned against, stemming from our "uncritical absorption in Romanticism's own self-representations".(8) Besides the merits of mature reflection, more than one hand might be involved in the making of changes (as is clearly evident in many Wordsworth manuscripts), further undermining the notion of the Romantic poet as a solitary genius who sings as artlessly as a bird. But the print medium is not well-equipped to show change over time except by the cumbersome apparatus of footnotes, cannot possibly for reasons of space present all versions as they evolved, and can only with difficulty make evident the extent to which composition and revision can be a social activity. Wryly, Kermode observes that these early, late, and intermediate versions "may one day be represented by hypertext, ...[although] these plural texts are not likely to be of much use to people who simply want to read Wordsworth and leave it to the experts to give them a text."

True, an electronic Wordsworth may not be eagerly sought out by the general reader; I've met very few lovers of reading who prefer a screen-image over an affordable book you don't need elaborate equipment to use. Indeed, such a Wordsworth edition may not in fact be meant to be "read" at all, at least not in the sense in which we still use the term today.(9) Instead, a hypertext Wordsworth would be meant to be explored, studied closely in a fashion that the linear structure of print makes difficult, and used at a distance

by scholars who don't enjoy the privilege of proximity to a major research library. Even more important, Wordsworth in hypertext may be the most effective way yet to represent Wordsworth in development. To know Wordsworth is to know not one but many selves, expressed in a succession of texts that mark the different stages of his personal development. New media functionalities of electronic text, multiple windows, digital images, and interlinked hypertext may offer an avenue into the reality of this poet's diversity. Furthermore, this new medium reinscribes textual stability as a series of moments in a lengthy creative process; it is hoped that by adding motion to comparative views, hypertext will enable us better to represent Wordsworth as an evolving self. Finally, if electronic texts are to justify themselves, they must make a space for themselves beyond the book. It is not enough to produce texts that are indexed and searchable; like the Aldine editions of old, electronic texts must establish their own unique conventions and explore easily-intelligible forms of presentation that distinguish them from what went before. The new medium should neither replace nor reproduce the book, but must strive to do things books could never accomplish.

There have been several other pioneers on this particular frontier of cyberspace, but none so far has conceived their project as we have. Chadwyck-Healey's English Poetry Full-Text Database is perhaps the best-known of the CD-based products, but it by no means focuses on Wordsworth nor does it print more than one version of any poem, and that usually from an out-of-copyright late edition. The same might be said of their Literature Online project, though I have had little opportunity yet to explore it. Their student version, English Poetry Plus on CD- ROM, contains just 39 Wordsworth titles, and never acknowledges the source of any of these texts. Using software designed by Electronic Book Technologies of Providence, Rhode Island, it does have a pleasing user interface, an effective search function, and links to brief biographical sketches. A more scholarly work is promised by David Miall of the University of Alberta; his Romanticism: CD-ROM, forthcoming from Blackwell's, is an electronic anthology offering a "generous selection" of poems by Wordsworth, but how these are to be chosen and what texts they are to be based on is not specified.

A networked environment promises much wider dissemination than a single-user CD, so it is gratifying to see Wordsworth already prominent on the World Wide Web. The best guide to Wordsworth on-line is to be found in Alan Liu's gloriously encyclopedic "Voice of the Shuttle" Web Page for Humanities Research , a resource to which we are all indebted. At least a couple of what can best be described as "fan pages" have been posted: Thomas C. Gannon's Wordsworth Page (which features a thoughtful quote for the day) and Richard Darsie's Selected Poetry of William Wordsworth, an eclectic gathering of a dozen poems. More systematic collections have been made available by academic institutions. Columbia University's Project Bartleby reproduces all the poems in the 1888 edition of Wordsworth's Complete Poetical Works, indexing them chronologically and by first line. The Representative Poetry Archive at the University of Toronto offers a selection of 44 poems from a wide range of textual sources, all documented. Individual scholars have also begun to experiment with posting electronic texts for study purposes. Michael Gamer's collection of the prose associated with Lyrical Ballads is a good case in point, while Richard Bear (a Ph.D. candidate in electronic book

design at the University of Oregon) has attempted to recreate the text of the rare 1798 Bristol first edition, first issue of Lyrical Ballads, probably based on a copy of the 1926 Noel Douglas replica.

Our project differs from the foregoing in several significant respects. First, the poems we will present have not been arbitrarily chosen, but are a reasonably coherent group stemming from a single collection. Second, our "copy-text" will not be chosen or reconstructed; instead we are carefully transcribing in digital form the texts of all the original editions themselves held in libraries around the world, though of course our procedures will be informed by the findings of previous scholars, especially the editors of the Cornell Wordsworth series. Hence, we will offer not just one version of each poem but all the versions of poems by Wordsworth and Coleridge appearing in the four lifetime editions of Lyrical Ballads, together with later important versions of poems revised by Wordsworth in his lifetime, and these will be linked hypertextually. Third, our e-texts will be "marked-up" or tagged using SGML (Standard Generalized Markup Language) in conformity with the principles of the Text Encoding Initiative (TEI). These files will then be collated using special software developed by Peter Robinson at Oxford University that will enable us to discover variants and generate an apparatus criticus. Fourth, we plan to link our transcribed e-texts to scanned images of the original printed editions in order to give the reader some sense of the look of the poems upon the page, and to link the differing versions of each poem to one another in their own hypertext web, showing all versions together for the sake of immediate visual comparison. Finally, this scholarly hypertext edition will be issued on CD-ROM in the first instance, with the intention of proceeding to network distribution as soon as it becomes practical. With luck, it will be issued by Cambridge University Press in 1998 to coincide with the bicentenary of the first publication of Lyrical Ballads.

Cambridge's CD-ROM series currently consists of four offerings: Peter Robinson's edition of The Wife of Bath's Prologue, a World Shakespeare Bibliography 1990-1993, Samuel Johnson's Dictionary, and the Works of John Ruskin. The first of these (the one with which I am most familiar) assembles all 58 of the pre-1500 manuscripts and printed editions of the prologue, using Robinson's Collate program to compare the different versions and to generate an apparatus of variants. The 58 versions, together with the apparatus and digitized images of 1200 manuscript pages, are linked together using Electronic Book Technologies' DynaText software to create a hypertext "book". Typically, a DynaText book is presented on the screen in two adjacent frames, one a narrow column on the left containing a table of contents and the other a wider data-frame on the right containing the text or texts compiled into a master file which give the effect of a single large file document that is scrollable and therefore searchable from top to bottom. Rather than accessing the contents of the DynaText book linearly, though, a simple mouse-click on one element of the table of contents takes the user directly to the designated section of the text. Each such section may be opened as an independent window, and the windows so generated can be scaled and moved about the screen to permit the comparison of texts. Up to this point, I have only been able to test our transcriptions and their SGML markup in Panorama Pro, an SGML browser from

SoftQuad of Toronto, which yields effects similar to DynaText, though lacking the capacity to generate multiple windows.

Although DynaText is powerful SGML software that can handle a huge mass of complex material, I find myself as the member of the editorial team responsible for hypertext design chafing at some of its limitations. Though it is very adaptable, it was not designed with the literary scholar uppermost in mind. DynaText was initially conceived as a commercial product designed to provide rapid indexed access to large prose documents such as catalogues and manuals; its table of contents function gives effective access to works with section titles and sub-headings, but is a blunt instrument when it comes to the presentation of poetry, where the user quite rightly demands access line by line. Short of embedding every line of each poem into the table of contents, there seems to be no easy way of navigating through the variant texts. The Collate-generated apparatus helps if opened into an independent window, but then other windows have to be opened and manipulated on the screen in order to compare lines once interesting variants are found. Relying as heavily as it does on the traditional critical apparatus of variants, this display is not much of an advance on the format of the printed book, and considerably more unwieldy. The cluster of multiple windows we need to generate can be awkward to scale and arrange on the screen, nor is it easy to keep track of which window holds which portion of the immense compiled document. With a multiplicity of texts such as the Lyrical Ballads offers, it is all too easy for the reader to get lost in cyberspace. The navigational tool so far proposed, the inclusion of a "base text for collation" at the top of the long document in the data-frame, is as much a relic of print culture as the model of the single-document "book" divided into subsections. In addition to the multiple versions of each poem, the editor is obliged either to construct a further ideal version of the poem to serve as the "base text" or, worse (because it violates the principle that this new medium should not merely reproduce the book but transcend it), to choose one of the existing variant texts, say from the first edition or the last lifetime edition, as the "base text". I had thought to address this difficulty by including texts of all the poems from all four lifetime editions of Lyrical Ballads as potential base texts, and then asking the reader at the beginning of each session to choose which of the editions he or she wished to use as the base text against which all the others would be collated. But this was not interactivity so much as an offloading of editorial responsibility onto the reader, and furthermore resulted in an unnecessarily vast proliferation of alternative texts within the DynaText file.

Instead, to address these navigational problems, I have begun to experiment with functionalities associated with Internet delivery of documents over the World Wide Web. An array of windows is essential to the display of the Wordsworth project, for as Sherry Turkle observes "windows have become a powerful metaphor for thinking about the self as a multiple, distributed system."(10) What seems most attractive about HTML standards at the present stage of development is that they give control over the use of frames, a system of layout in which windows are generated on the screen according to a pre-determined pattern. Though this pattern can be easily altered by the user, its grid provides an intelligible starting point for the display of multiple texts of the same poem on the screen at the same time. The prototype pages I have developed so far take "We are

Seven" as a test-case, and consist of two successive contents pages which lead to a galaxy of pages each composed of five frames: one narrow vertical column on the left plays the customary table of contents role, while a grid of four squares on the right allows the reader to compare four different versions of the poem on the screen at once. The text in each frame may be scrolled through manually, but by the use of HTML's internal anchors a simple click on a live hyperlink in the left-hand frame causes all four texts in the squares to scroll simultaneously to the same line. There are two advantages to this scheme of display: first, a degree of animation is introduced into the text that cannot be duplicated in print, and, second, a balance is struck between exploration and direction in the reader's examination of textual complexity.

There is a second unusual feature of this scheme that is meant to help readers find their bearings. The left-hand column in this HTML treatment contains no mere list of contents but what I call a "variant map" of the poem being studied. The variant map is a guide to revisions that were made at various stages in the poem's development. It is based on the poem to the extent that it reproduces the text of the poem wherever changes were not made, but whenever a change in any of the versions under consideration is encountered it substitutes a descriptive hyperlink for the variants themselves. The reader is thus alerted by a sort of palimpsest that an alteration has been made, and by clicking on the "hotspot" can summon up the parallel passages. Replacing the base or reading text with a variant map turns the annotation process inside out in a way that seems appropriate to the dynamism of this medium, for the hiatus in reading caused by running across a link cues the reader to click on the spot and look to the right (in the direction of the normal flow of reading) to learn the word or character elided and to compare texts. Rather than footnotes which distract attention from a definitive text, the variant map is an abstraction of the poem which does not privilege one version of the text over another and that piques the reader's curiosity by means of gaps in the text to pursue the significance of revisions made in successive versions. Together with its links to the four display windows, the variant map makes possible a dynamic collation of variant texts that surpasses print.

Without doubt, there are numerous disadvantages to this proposed display scheme. Currently available bandwidth is simply too narrow and therefore slow to deliver over the Internet the amount and complexity of digitized texts and images we envision. Besides, not everyone is yet using a frames-capable browser, such as Netscape 2.0 or better (and some users are still restricted to the text-only Lynx system). We have therefore thought it expedient to stick with CD-ROM distribution for the present. Furthermore, whether reading out from the 'net or from CD, the 14-inch monitors commonly in use today do not have enough screen real-estate to display more than four versions of the text at a time, and even this minimum array for our purposes appears somewhat cramped. While these drawbacks may soon disappear given the rapid advances in computer technology we are experiencing, a more lasting objection is raised when we consider that HTML is a much simplified subset of SGML and therefore inadequate to represent the complexity of the texts in question. However, in the case of the variant map element, it should be recognized that it can function effectively in the context of either markup scheme. The way its links interact with the frames environment is another matter, though, since web-browsers like Netscape process internal anchors much more readily

than an SGML viewer like Panorama Pro can. Indeed, I have been unsuccessful in getting Panorama Pro to run my HTML-based dynamic collation files properly, even when they are linked to the appropriate HTML-DTD. Hence I have been driven to the expedient of employing two software applications where one should do. Peter Robinson assures me that DynaText will successfully process both SGML and HTML files with internal anchors, so I am hopeful that the final product will not require the reader to constantly be switching software. At present, though, I see no alternative to employing both SGML and HTML, since I am not convinced that the TEI-DTD has an adequate hyperlink capability. I have experimented with <ref> and <xref> tags, but have not been able to reproduce the simultaneous scrolling function I can achieve with HTML. At the moment, it remains unclear to me whether the fault lies with the TEI subset of SGML, the software I am using, or my own ignorance. I have thought of experimenting with HyTime links in SGML, which perhaps will offer a solution. Meanwhile, I shall make pragmatic use of HTML along with TEI-conformant SGML for the project, since I see the two designed for different purposes. SGML, especially in the form recommended by the Text Encoding Initiative, is the best way to represent the logic of complexly-structured literary documents, by this means giving us rapid indexed access to large documents and enabling them to be searchable and self-concordancing. It was never meant to perform complex hypertext functions beyond those. In my experience, I have found that the HTML subset has a much more powerful and obvious linking functionality that makes dynamic collation as well as the coordination of text and images possible. Perhaps it is just as well to regard them as complementary, and allow DynaText to process documents with either DTD seamlessly. But I welcome your advice on these matters.

In whatever ways this project may develop, it takes its cue from a wish to attempt to do things through this medium that print cannot, and in the process to discover what unique capabilities electronic publication has, if any. The fixity of print has been able to give us one Wordsworth or the other, but has not proven capable of fully representing his evolving consciousness. The poet was well aware that each of us passes through a succession of selves, not completely fragmentary and unrelated to one another but sufficiently discontinuous for him to wish for our "days to be/Bound each to each" by some means or other. Even if no more successful than natural piety, digital media can at least enable us to give attention to the various stages of composition and revision and to give each of them the respect they deserve. In the end, we may find that we have done no more than translate the contents of one medium into another, but that will not be without value, as textual critic James Thorpe reminds us:

> All editions carry the taint of time. They are for the here and now, whether that is
> a   decade, a generation, or a century. Ultimately they must all be replaced for most
> purposes.(11)

To the extent that computers can help us to embody postmodern concepts like the decentered self and show them at work in specific instances, they are the tools of the time and will certainly have an impact on the study of literature. At the least, digital hypertext editions can spark a renewal of interest in editing and textual scholarship, analytical and

descriptive bibliography, and even assist in the realignment of literary studies from the theory and practice of critical interpretation to the new cultural history's focus on the literary marketplace and the material production and representation of "literature".(12) What consequences they will have we cannot know until we try.

Notes

1. Hypertext: the Convergence of Contemporary Critical Theory and Technology (Baltimore: Johns Hopkins University Press, 1992), p. 19. We might also compare Jean Jacques Servan-Schreiber, The Knowledge Revolution (Pittsburgh: Carnegie-Mellon University Press, 1986), p. 1: "What the invention of the printing press did for mankind five centuries ago, the computer is doing today." A characteristic McLuhan observation cautions us that "with the Gutenberg technology we move into the age of the take-off of the machine"; see The Gutenberg Galaxy (Toronto: University of Toronto Press, 1962), p. 155.

2. See Pierce Butler, The Origin of Printing in Europe (Chicago: University of Chicago Press, 1940), pp. 140-143; Lucien Febvre and Henri-Jean Martin, The Coming of the Book: the Impact of Printing 1450-1800, trans. David Gerard (London: NLB, 1976), pp. 82-88; Karl Schottenloker, Books and the Western World, trans. William D. Boyd and Irmgard H. Wolfe (Jefferson N.C.: McFarland and Co., 1989), pp. 78-79.

3. See Martin Lowry, The World of Aldus Manutius (Ithaca: Cornell University Press, 1979), pp. 110-147.

4. "Preface" to The Poetical Works of William Wordsworth, ed. E. de Selincourt (5 vols.; Oxford: Clarendon Press, 1940), I, v.

5. Lyrical Ballads, and Other Poems, 1797-1800, ed. James Butler and Karen Green (Ithaca: Cornell University Press, 1992), p. v.

6. William Wordsworth, ed. Stephen Gill (Oxford: Oxford University Press, 1984), back cover.

7. "Floating Hair v. Blue Pencil", London Review of Books, 6 June 1996, pp. 15-16.

8. The Romantic Ideology: A Critical Investigation (Chicago: University of Chicago Press, 1983), p. 1

9. Sven Birkerts points to this distinction quite clearly in The Gutenberg Elegies: The Fate of Reading in an Electronic Age (Boston: Faber and Faber, 1994), p. 15: "It is precisely where reading leaves off, where it is supplanted by other modes of processing and transmitting experience, that the new dispensation can be said to begin."

10. Life on the Screen: Identity in the Age of the Internet (New York: Simon & Schuster, 1995), p. 14. Turkle follows George Landow's Hypertext and Richard Lanham, The

Electronic Word (Chicago : University of Chicago Press, 1993) in writing of the way identity is understood as fragmented and discontinuous in the digital age.

11. Principles of Textual Criticism (San Marino: The Huntington Library, 1972), p. 179.

12. A sign of this latter trend is Alan D. Boehm's recent article on "The 1798 Lyrical Ballads and the Poetics of Late Eighteenth-Century Book Production", in ELH, 63(1996), 453-87, which alerts us to "a set of rhetorical as well as technical and commercial practices" involved in the creation of Lyrical Ballads as an artifact of print culture.