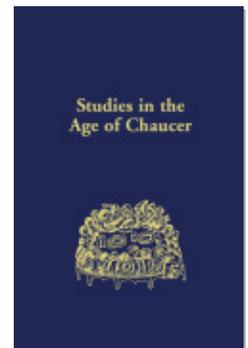*Chaucer:* The Wife of Bath' *s Prologue on CD-ROM. The* Canterbury Tales Project ed. by Peter Robinson (review)

Mark Allen

→ *For additional information about this article*
https://muse.jhu.edu/article/660876/summary

PETER ROBINSON, ed. Chaucer: *The Wife of Bath's Prologue on CD-ROM.*
The *Canterbury Tales* Project. Cambridge: Cambridge University
Press, 1996. 1 computer disk; 80 pp. booklet. $240.00.

This is the first installment of the *Canterbury Tales* Project, an under-
taking that proposes to make available all data fundamental to recon-
structing the textual history of Chaucer's poem. It is an impressive first
installment, certain to make evident the value of the larger Project and
likely to become a benchmark for similar products in other literary and
historical specialties. It demands that its users develop skills with
mouse and computer screen well beyond what I suspect most
Chaucerians have, and it demands care, even caution, in deriving its ex-
traordinary data. It rewards these skills and this caution with a view of
the future and with complete textual data for *The Wife of Bath's Prologue*,
data available in no other single source, including Manly and Rickert's
The Text of the Canterbury Tales.

The CD-ROM contains facsimile-quality digital images of all fifty-
eight fifteenth-century witnesses to the Wife's *Prologue* (some 1,200
pages of fifty-four manuscripts and four printed editions), transcrip-
tions of these witnesses in both regularized and unregularized spelling,
word-by-word collations of the transcriptions, and, as described in the
booklet that accompanies the disk, "spelling databases grouping every
occurrence of every spelling of every word in every witness by lemma
and grammatical category" (p. 13). The disk also includes Daniel W.
Mosser's codicological descriptions of each of the fifty-eight witnesses
(a preliminary version of the materials for Mosser's anticipated
*Descriptive Catalogue of the Manuscripts and Pre-1500 Printed Editions of
the* Canterbury Tales) and Stephen Partridge's transcriptions and colla-
tions of all glosses to the Wife's *Prologue*. Mosser and Partridge each pro-
vides an essay that introduces the principles underlying his contribu-
tion. Further, the disk contains two essays reprinted from *The
Canterbury Tales Project Occasional Papers Volume I* (see *SAC* 17 [1995]:
301, no. 23): one by Robinson and Elizabeth Solopova on the theory
and principles of transcription of the witnesses, and one by Norman
Blake on editing *The Canterbury Tales*.

The manuscript images on the disk can be magnified or printed.
Collations and spelling databases are available at the click of a mouse,
as are lists that enable comparison of a given line in the witnesses that
contain it. Split-screen technology makes it possible to compare
manuscript with manuscript, transcript with manuscript, or multiple

combinations of ᴍ.u.script, transcript, and apparatus. The base text for the collations is "a very lightly edited" version of the Hengwrt manuscript, but the collations can be read against any of the fifty-eight witnesses. The disk is fully searchable in both normal language and SGML code. The data can be cut-and-pasted into word processors, and users can add their own hypertext links, bookmarks, and annotations to the disk. Robinson estimates that the disk contains somewhere around 10 million items of information about individual words, their parts of speech, variant spellings, and relations with other occurrences, etc. These data are tied together in a web of some two million hypertext links. God's plenty indeed.

Such a summary description of the materials on the disk and the ways they can be used disguises the work's enormous potential and its possibilities for confusing or frustrating its users. Computer adepts will capitalize on the gold mine of information without having to pause long in confusion. Computer preliterates will probably never get past the confusion. Most of us, however, have much to learn much from the disk, not only about the manuscripts of *The Wife of Bath's Prologue* but also about computer technology. In a month or two of working recurrently with the disk, I have learned how to make my own links and annotations, how to develop sophisticated searches, how to toggle between the CD-ROM and various word processors, etc. In short, I have learned how to manipulate the technology to move toward using the vast quantities of information here, although I do not imagine that I have exhausted the potential for complex searches. The instructional booklet that accompanies the disk is helpful, but not thorough in explaining processes or abbreviations; the "help" screens on the disk itself are extensive.

Facing this first installment of the Project is something like facing Manly and Rickert's *Text of the* Canterbury Tales for the first time, despite the important differences between the two. Perhaps the most important difference is that Robinson's CD is only a beginning. Manly-Rickert includes various surmises and conclusions about the textual history of *The Canterbury Tales* as well as textual data for the entire work. Robinson's CD includes data from the *Wife's Prologue* only—the kind of data from which Robinson and Blake promise to generate their own surmises and conclusions, but raw data nevertheless. I imagine that the codes and charts and bland typeface of Manly-Rickert were as intimidating in 1940 as the electronic wizardry of the CD-ROM is today, and though the Manly-Rickert blizzard of sigils and abbreviations has become more familiar with time, Robinson's *Wife's Prologue* benefits

from more normal language in the collations. Nevertheless, the multiple-screen technology produces some baffling codes of its own. Terms like "regapp," "unregapp," "linapp," and "revwtxt" are used to label individual screens, meaning (I think) "regularized apparatus," "unregularized apparatus," "line apparatus," and "review witnesses to the text of a given line" respectively. "Fulltext" labels a number of different kinds of screens, creating possibilities for confusion when multiple screens are open simultaneously, and none of these screen-labels identifies which witness is being viewed, making possible much greater confusion. When comparing transcripts of individual manuscripts, for example, I often find it necessary to scroll to the top or bottom of a given screen to recall which manuscript I am viewing.

Moreover, scrolling itself causes some problems, especially when one manipulates the manuscript images and certain parts of the apparatus. Magnification or reduction of a given image brings one back to the top of the image, requiring recurrent back-scrolling to reach the point of interest when attempting to examine closely a particular word or flourish. However, because one can here closely examine magnifications of individual words, flourishes, even marks of punctuation, the irritation of scrolling is minor compared with the rewards.

Scrolling any collation screen produces a curious glitch. The top of the screen conveniently includes the base-text version of the line under consideration, yet once scrolled off-screen, the base-text line reappears in a scrambled version when one scrolls back to it. Such glitches and the issues of labeling above make it necessary that users be very cautious about the information they carry away from the disk.

An extended example will help clarify the kinds of information available and the caution necessary in deriving it. The CD enables users to examine any given word or line of the *Wife's Prologue* in a number of ways. Clicking on the line number to, say, line 1, produces a "revblk" screen (a bit of code I have not broken). The screen usefully informs us that in forty-seven witnesses, line 1 appears after a rubric; in seven, it appears first without a rubric; and in four the line is "out," or does not appear at all. Clicking on the icon next to forty-seven produces a "revwtxt" screen that reads "Single-click on any line to move to that line in the transcription. Number of witnesses with this reading in line 1: 47." The screen then lists each witness in which line one appears after a rubric, even though the text in the list is not fully formatted. That is, the text of the lines listed on this screen is normal-

ized in ways that it is not elsewhere on the disk, making it necessary that users are careful not to use this text instead of the more precise version of the base-text.

A similar search route produces another, almost identical, "revwtxt" screen. Clicking on the word "Experience" in the base-text results in a "regapp" screen that identifies all variants of the opening word of the *Wife's Prologue*, forty-two of which agree with the lemma, seven reading "Experiment," and four out. Clicking on the number "42" brings another "revwtxt" screen very easily confused with the one above. It reads "Single-click on any line to move to that line in the transcription. Number of witnesses with this reading in line 1:42," and it lists each witness to line 1 that has the word "Experience" in initial position, a list very similar to the one above. The information on the two "revwtxt" screens is accurate, precise, and useful, but one screen is too easily confused with another, especially when one or both is held open for future reference. There is simply no way to distinguish them except by remembering which is which or by repeating the searches. With multiple screens and frequent rearrangement of screens, memory is not sufficient. Redoing the search for certainty is easy enough, and that is the point I wish to make: users must confirm the information they derive because the system of labeling on the disk is insufficient to the point of potential confusion.

Similar caution is necessary in using the spelling databases where the complicated taxonomy of grammatical categories can lead users to make incorrect assumptions about the data they are viewing. Calling up the verb "be" in the All-witness Spelling Database produces the impressive datum that the verb occurs 12,828 times in 172 spellings. This is further broken down to tell us that the verb occurs 1,040 times as an infinitive in 12 spellings, with 902 occurrences spelled "be"; 19, "bee"; 73, "ben," etc. The five occurrences of "ben" with a terminal flourish and the three occurrences where it has a macron over the "n" are distinguished from unadorned "ben," as are other variations. It takes a bit of scrolling, however, to discover that the "be" spelling also occurs as an infinitive in initial position 59 times, in addition to the 902 listed previously, with capped (56) and uncapped (3) occurrences listed separately. Another bit of scrolling finds that the "be" spelling of the infinitive also occurs 366 more times in rhyme position, but it takes a good ten or so more screens to find listed the one occurrence of "be" as an indicative present singular, the 183 occurrences of indicative present

plural, and the 159 occurrences of the subjunctive present singular. Continued scrolling (or paging) discovers the remaining instances of subjunctive, impersonal, and participial instances of the verb in its "be" spelling. This is impressive data indeed, and with a 200 MHz Pentium computer, it took me just over a minute to do this scrolling, although on a 100 MHz model with a slower CD-ROM drive it took well over five. The minimum equipment required to run the CD is a 386 or faster PC with at least a double-speed CD-ROM drive, Windows 3.1+, and 8 MB of RAM (or Macintosh System 7 or later, double-speed CD-ROM drive, and 4 MB RAM), but I am sure that the disk would be very frustrating to use on minimal equipment. I suggest that users work on very fast machines, that they double-check their search paths, and that they make certain they derive complete information. On the other hand, for the purposes of this review, I easily cut-and-pasted above quotations from the "revwtxt" screens, and I printed the data about "be" just as easily, enabling me to check the data without having to scroll more than once on a slow machine.

Like Manly-Rickert, the CD-ROM is intended primarily for those interested in textual and editorial issues, a basic reference tool for libraries. But the disk shows the potential for the overall Project to go well beyond Manly-Rickert in some important ways. Partridge's information about the glosses is much more extensive than that in Manly-Rickert. Mosser's descriptions of the witnesses are more detailed than Manly and Rickert's, and they benefit from the many recent advances in codicology. The spelling databases of the *Canterbury Tales* Project in particular will be a resource for historical linguists unavailable in Manly-Rickert, indeed unavailable anywhere else at present with the possible exception of the archives of the *Middle English Dictionary*.

Furthermore, the CD supersedes Manly-Rickert in at least one fundamental way: the graphic capabilities of electronic publication make for very refined transcriptions and the transcripts are all here for the checking. I cannot pretend to have checked transcriptions on the CD at all exhaustively, but I can say that what I have checked has been 100 percent accurate. According to Robinson, each transcription was checked at least three times, and wherever possible, checked and double-checked against the manuscript itself as well as a reproduction. To transcribe the witnesses of the *Wife's Prologue*, Robinson and Solopova developed a unique transcription policy, detailed in their essay included on the disk. While they acknowledge that no transcription can duplicate a manuscript, Robinson and Solopova provide a system of

"graphemic" transcription that distinguishes a number of abbreviations, superscripts, marks of punctuation, and special or terminal characters that are beyond the alphabet or ASCII code. This makes for, as they put it, very "rich" transcription that is more open to error than a less complicated system. Awareness of this potential for error has apparently encouraged the transcribers to be highly cautious, and if the entire Project achieves what seems to be the standard of this disk, it will be very high indeed. When they wrote their essay describing the transcription method, Robinson and Solopova hoped that their final check of the transcripts would "find less than one correction for every four thousand characters," and presumably the final check took care of those. They hope that scholars who check their transcripts will find them "perfectly reliable." I did not check four thousand characters in each or any of the fifty-eight witnesses, but I did spot-check various collations against the manuscript images, against Manly-Rickert, and against John Fisher's unpublished Variorum collations of the *Wife's Prologue.* I did not discover any errors on the CD, and I was recurrently astonished at how much information it includes.

There can probably never be a simple, single, or final test of Robinson and Solopova's method, of the *Wife of Bath's Prologue* on *CD-ROM*, or of the *Canterbury Tales* Project more generally. Their success will depend upon how much they are used, imitated, and relied upon. So, I am at something of a loss to pronounce judgment on any of them now, so early in what will be their scholarly legacy. Nevertheless, the labor involved in producing this one disk is enormous, exacting, and justified. Peter Robinson's vision in embarking on the task is to be commended. He developed the electronic collation program that underlies the enterprise and was instrumental in developing the *Canterbury Tales* Project to make possible this first installment. The disk testifies that Robinson's vision is both practical and exciting. The Web page of the *Canterbury Tales* Project (http:www.shef.ac.uk/uni/projects/ctp/Main/occ.html) promises that a CD-ROM of *The General Prologue* will be available late in 1997, *The Nun's Priest's Tale* in 1998, and that significant progress is under way for the transcription of all of fragments 1 and 7. May they all be as careful as Robinson's first CD, and may they become less daunting with familiarity.

MARK ALLEN
The University of Texas at San Antonio

323