

"The Letter was not nice but full of charge":
Toward an Electronic Facsimile of Shakespeare

Speaking Notes for a paper delivered by

Kenneth B. Steele
Centre for Computing in the Humanities
University of Toronto

at the Combined 16th Annual Association
for Literary and Linguistic Computing Conference
and 9th International Conference on
Computers and the Humanities (ALLC/ICCH):
"The Dynamic Text."

University of Toronto, June 5-10, 1989.

[A laptop computer and overhead projection
palette were utilized for the examples.]

William Shakespeare's dramatic career began four centuries ago, and his associates John Heminge and Henrie Condell tell us that, from the beginning, the texts of his works were abused by what they term "diuerse stolne, and surreptitious copies, maimed, and deformed by the frauds and stealthes of iniurious impostors, that expos'd them." In their preface "To the great Variety of Readers" in the 1623 First Folio, they assert that finally their volume presents these pirated plays, "cur'd, and perfect of their limbes; and all the rest, absolute in their numbers, as he conceived the[m]." Of course, all the early quartos made very similar claims to editorial integrity: for example, the 1603 first quarto of Hamlet purports to present the play "As it hath beene diuerse times acted by his Highnesse seruants in the Cittie of London." This assertion sounds distinctly like the publicity for a very recent edition, which declares: "After eight years' work by the world's finest Shakespeare scholars, here are the plays as they were acted by Shakespeare's company."

The inescapable reality is that the entirety of the Shakespearean corpus comes to us through the mediation of scribes, editors, compositors, and proofreaders, and that, with every edition, from Renaissance quartos to modern diskettes, layer upon layer of interpretation has accumulated and obscured the texture of the author beneath the glossy paste-wax finish of emendation. Each edition claims to "correct" ostensible "corruptions" and "compositorial errors," in the process retreating still further from the original authorial manuscripts. Most damaging of all is the time-honoured editorial preference for a single, "definitive" text, rather than a respect for what Randall McLeod calls the "infinitive" text. This attitude results in rationalization of variants as "indifferent" and conflation of

sometimes strikingly different plays into one.

The Quarto and Folio texts of *King Lear*, for example, disagree in 400 complete lines, and thousands of variant words. Recent scholarship by Steven Urkowitz and Gary Taylor, among many others, has demonstrated the consistent and deliberate alteration of character, theme, and plot between the texts. When these distinct texts are conflated, it should hardly be surprising that many elements become contradictory, or that characters become inconsistent. Shakespeare may be responsible for the folio *King Lear*, or he may not, but certainly he never imagined the monstrous conflation which has dominated editions for several centuries.

King Lear is not a special case, although its seminal importance to the twentieth century has made it the most celebrated example of textual instability (or what we at this conference are calling a "dynamic text"). Seventeen of Shakespeare's 38 plays survive in 2 variant forms, and 2 more, *Romeo and Juliet* and *Hamlet*, survive in 3. Conflation and the editorial impulse to create a single "definitive" text is an accommodation of a naturally human but rather lazy preference for the uncomplicated rather than the uncomfortable or inconvenient. I propose that Friar Lawrence's words to Friar John in *Romeo and Juliet* have found new relevance in the twentieth century as a warning to Shakespeare scholars:

The Letter was not nice but full of charge,
Of deare import, and the neglecting it,
May do much danger[.]

(_RJ_ Q2 5.2.18)

Every letter, ligature, or punctuation mark in the original texts, although occasionally annoying or puzzling, is indeed "full of charge" -- and in more than one sense, when one is talking about an electronic Shakespeare textbase.

Twenty-five years ago this month, shortly after I was born, Trevor H. Howard-Hill performed his first test concurring of Shakespeare's *Measure for Measure* at the Oxford University Computing Laboratory.[1] The English Electric KDF-9 computer he used took paper-tape input, and possessed a core memory of what we would call 187 K. By the end of the decade, though, Howard-Hill had overseen the keypunching and quintuple proofreading of the entire First Folio, several important quartos, and a number of apocryphal plays, in the course of producing the Oxford Old-Spelling Shakespeare Concordances. To this day, those files remain the bulk of the computer-readable transcripts of Shakespeare's quarto and folio texts.

In the early 1970's, Howard-Hill claimed it was still "too difficult, time-consuming, and expensive for all but the most determined scholar to work with the computer." [2] Downloaded to a modern microcomputer, though, the 55 original texts of the 38 plays now occupy 10.2 megabytes of disk space, and

even with their WordCruncher indexes, can fit on a 20-megabyte hard disk. For those who steadfastly prefer an edited text, the Riverside Shakespeare and the Oxford Complete Works are now available commercially on diskette for a few hundred dollars, and Oxford is considering the release of its recent old-spelling edition as well. In many ways, it is now too difficult, time-consuming, and expensive for all but the most determined scholar to work without the computer.

Early last year, the Centre for Computing in the Humanities began the Shakespeare Text Archive project, to make Howard-Hill's mainframe files part of an accessible and convenient textual tool for students and faculty at the University of Toronto. Although still incomplete, the textbase was formatted for use with ETC's WordCruncher software, and made available through the CCH's network. The process of verification, encoding, and keypunching continues even now, but the archive currently contains the First Folio, all of the so-called "good" quartos, and most of the "bad" quartos. Hopefully, once the Shakespearean texts are complete, the means will be found to add a collection of Shakespearean sources and perhaps even a modern edition or variorum annotation for cross-reference. Ultimately, the archive could be converted into a form of hypertext, perhaps linked with optical facsimiles of the texts in question. This form of "electronic variorum" might finally approximate a genuinely "infinite text," and distributed on CD-ROM could be easily accessible to all Shakespeareans.

Using text retrieval software such as WordCruncher or even TACT, the textbase is currently not so much useful for computerized stylistic, authorship, or collational analysis, as it is a remarkable tool to accelerate and refine more traditional approaches, such as close reading, imagery analysis, or source study. First of all, the text becomes what I liken to a "doodle pad": through casual browsing, random experiment, and plain dumb luck, the user often stumbles across something which sparks his imagination. Reading Shakespeare via WordCruncher is another form of stimulus to the critical imagination: a diachronic corpus suddenly becomes synchronic, multiple plays interpenetrate on a single word, and one is reading vertically, by cross-section, rather than horizontally, from beginning to end.

Working with the original texts can be complex and confusing, or "inconvenient" as some would say: the archive contains more information than many users may think they want. For literary rather than textual purposes, the user must account for a bewildering array of duplicate texts, variant spellings, typographical eccentricities, textual cruxes, and apparent compositorial "errors" familiar to anyone who has tried to read Shakespeare in facsimile. Nonetheless, with a little practice, the archive can yield better results than an electronic edition and with considerably greater textual fidelity.

Another purpose of research with the Shakespeare Text Archive is lexical: the Oxford English Dictionary reports the Renaissance definition of any given word, but does not easily reveal Shakespeare's personal sense of its connotations and denotations. Examining each occurrence of the word in its

Shakespearean context can reveal Shakespeare's intellectual and emotional attitude toward it: for example, Shakespeare seems to consider "dogs" fawning, subservient, low-life creatures, but "hounds" loyal, noble hunting animals.

It is also possible to search for words in conjunction with the codes for title pages, stage directions, speech prefixes, prologues, etc. (There are 60,769 Cocoa-type codes in the texts). One quickly finds that, of 6276 stage directions, and 554 beds, only 6 beds occur within the original stage directions, in *2 Henry 6*, *Romeo and Juliet* (Q1 only), *Othello*, and *Cymbeline*.

[Example: eyen, eyne, eine]

An examination of the fifteen occurrences of the obsolescent plural "eyen" (in its various spellings) reveals that in all cases but one (*Pericles*), the word is used primarily for purposes of rhyme (and eight times to rhyme with "mine"). Furthermore, it seems clear from Shakespeare's use of the word in *A Midsummer Night's Dream* and *As You Like It* that he considered it preposterous: appropriate only in circumstances such as Bottom's performance as Pyramus, or Phebe's love poem to Rosalind.

[Example: Ears.by, Ears2.by, Ears3.by]

Another example may demonstrate the usual investigative sequence more clearly. The ear is referred to 603 times in Shakespeare's original texts. As one reads through these occurrences, 25 references seem strangely sexual -- specific evidence which is more convincing than vague Freudian generalizations. Furthermore, when one searches for co-occurrences of the ear and poison or infection, one finds seven places in which pouring poison in the ear is closely associated with some form of male sexual jealousy -- which I would argue helps to explain the symbolism of *Hamlet* as it may have formed in Shakespeare's mind.

[Example: Theatre.by]

The user can create a list of theatrical terms [act, appear, comedy, costume, globe, illusion, mask, pageant, part, perform, platform, play, revel, scaffold, scene, show, stage, theater, and tragedy] to conduct a preliminary study of metatheatricity across the corpus. My word list matches 3384 occurrences, largely in the tragedies and good quartos, and concentrated in the Elizabethan plays. The histories and the First Folio seem comparatively lean in metatheatrical references. *Hamlet* and *A Midsummer Night's Dream*, understandably, have by far the greatest concentration of theatrical language, owing to the internal performances, and their overall thematic concerns.

The preceding illustrations have been primarily literary in nature, focusing on lexicon and content rather than textual minutiae, which is the forte of the Shakespeare Text Archive. Unlike electronic editions of Shakespeare, the archive aims to be an electronic facsimile of the original quarto and folio texts. Obviously, it cannot yet replace more conventional photographic facsimiles, but the coding of signatures allows easy cross-reference to the Norton facsimile or the Allen & Muir quarto facsimiles. Using text retrieval software, an electronic facsimile can provide virtually instant answers to questions previously unthinkable. An absolutely precise electronic facsimile is still years away: only some form of flawless optical process will satisfy bibliographers. Yet despite its minor imperfections, the Archive as it stands can eliminate much of the drudgery of mechanical and repetitive tasks, liberating the scholarly imagination to dream up new questions, rather than laboriously seek to answer old ones.

[Example: I2.byr, I.byr]

For example, the word "aye / ay / aie" [I2.byr] occurs a total of 76 times in the texts. In 41 cases, it occurs in the expression of self-pity, "ay me / mee," in 27 cases it means "ever" (17 times it is "for ay / aye"), 7 times it means "yes," and in 3 other cases it is the French first person singular verb. The spelling "i" (the single letter), occurs considerably more often: 31,540 times in the texts: in 30,743 cases, it is the pronoun (in 202 cases in contracted form), and 999 times it signifies "yes" (remember, in only 7 cases was it spelled "ay / aye"). [I.byr] "I" as a single letter is never the desperate "ay me," nor is it used to mean "ever."

[Example: Italics]

Italic typeface is used 74,655 times in the corpus, and *Troilus and Cressida* and *Antony and Cleopatra* seem to have double the average at about 2000 occurrences each. This may be partially explained by the lengthy italic advertisement prefixed to the quarto *Troilus and Cressida*, and the italic Prologue which begins the folio version. WordCruncher is not, however, counting each italic letter, but only the font codes at the beginning of each italic section. The bulk of the italic occurrences seem to be caused by italic speech prefixes, intensified by lengthy stichomythia, and the compositors' scrupulous italicization of proper names in the text (and both Roman plays seem to delight in allusions to historical figures).

[Example: Hyphenation]

Hyphenation occurs in only 8561 words, but because these are largely single occurrences, it took my 12 MHz machine two and a half hours to come to that conclusion

(fortunately I saved my results for the purposes of this fifteen-minute talk). The hyphenation, for the record, seems to be heavier in the folio than the quartos, and occurs with especial frequency in *The Merry Wives of Windsor* (532 times).

[Example: Punctuation (, . : ; ? !)]

Punctuation study, previously almost unthinkable, is extremely quick and painless on an electronic text. In a conference held here at Toronto last year, a presenter received an ovation for having diligently counted all the commas in a Renaissance text. Within seconds, however, an electronic facsimile can determine that there are 138,198 commas in the original texts (which would take over 23,000 WordCruncher screens to view), and can offer distributions across the works. Distributions of the 104,928 periods, 26,974 colons, 5820 semi-colons, 15,785 question marks, and 1162 exclamation marks can also be examined, to determine, for example, that the Folio seems light in periods, but heavy in semi-colons and colons.

WordCruncher is not a collation program, and other programs I've tried thus far seem too particular to handle the variant spellings and significant alterations in the Shakespeare texts. It is possible, however, to index together versions of a play with WordCruncher, and to output a comparison word frequency file.

[Example: "type hamlet.cmp"]

This is rather less interactive than WordCruncher's other functions, but it does produce some results. Of course, these results cannot finally be interpreted without division along compositor stints, which are coded in the text itself but which the current software cannot analyze.

Comparing the word frequencies of Q2 and F1 of *Hamlet*, a number of patterns become immediately evident. The folio seems to double the number of prose lines (perhaps because of shorter column width). Question marks and semi-colons double, colons quintuple, and exclamation points all but disappear. The compositors of the folio *Hamlet* overall prefer to drop silent e's, except in *finde*, *minde*, *kinde*, *winde*, and eleven other words which generally have doubled vowels or consonants. Quarto 2 demonstrates a marked preference for the spelling "o", while folio markedly prefers "o-h." The list could be endless, and is (as you can see.)

Not only is the Shakespeare Text Archive textually

incomplete, but the software engine which drives it, currently WordCruncher, is too limited to perform all the research which the text could ultimately support. The ideal workstation, or software package, for the Shakespeare Text Archive would have to be capable of performing more complex multitasking functions, such as on-screen collation of multiple texts or comparison of alternate pairs (for example, the distributions of "go" versus "goe" (with an "e") through compositors). The text retrieval software, or a module compatible with it, would have to be able to search the textbase for rhetorical repetitions of any kind: anaphora (the repetition of a word at the beginning of syntactic units), epistrophe (end), epanalepsis (beginning and end of same unit), anadiplosis (end of one unit and beginning of next), and perhaps even alliteration and rhyme, if adequate phonemic encoding could be included. This ultimate workstation would be able to facilitate the detection of patterns in the significant differences between variant texts or compositor stints. The software would be flexible enough to reorder ranges based on the chronological order of composition, performance, typesetting, or publication. It would allow operations to be conducted on the entire corpus or on a single play or plays.

Ultimately, the textual study of Shakespeare will lead to a combination of optical document and machine-readable text: a form of hypertext which would also allow computerized comparison of damaged type, press variants, etc. If Herner Schnelling's calculations in his 1987 Literary and Linguistic Computing article are correct, however, a single Folio page would require 3,952 million bytes of storage to meet the CCITT [3] facsimile standard. To the average Shakespearean, optical storage is currently something "too difficult, time-consuming, and expensive" -- but as we all know, technological limitations are transitory. For today, a readily accessible ASCII text archive is a step toward the day when we, along with Prospero, can declare, "I'll drown my book."

N O T E S

1. Howard-Hill describes the process in detail in his article, "The Oxford Old Spelling Concordances," *Studies in Bibliography* 22 (1969): 143-64. The later history of the texts is described in an anonymous note, "Shakespeare and the Computer," *ALLC Bulletin* 8:1 (1980): 72.
2. "Common Shakespeare Text" 54.
3. (Comite' Cosultatuf International Te'le'phonique et Te'le'-graphique).

The contents of this electronic file are copyright (c)1990
Kenneth B. Steele, University of Toronto. Quotation for scholarly
(non-commercial) purposes is permitted, but please contact the
author (or)
to verify the material in question and advise him of your intention.
Please do NOT distribute.