

The analytical bibliography of electronic texts

John Lavagnino

Box 1841, Brown University, Providence, RI 02912, USA

KEYWORDS: analytical bibliography, electronic texts

AFFILIATION: Women Writers Project, Brown University

E-MAIL: John_Lavagnino@Brown.edu
FAX NUMBER: none
PHONE NUMBER: (401)751-2014

Analytical bibliography is the study of “the physical embodiments of texts as evidence of the process that produced these embodiments and of the relations between them” (Williams and Abbott, 6). It is sometimes referred to as “physical bibliography,” and in discussions of this field there is always a stress on the examination of physical evidence as its defining characteristic; the analytical bibliographer looks at paper, typefaces, bindings, and so on, not at the textual content that these materials are used to convey. The goal has most often been to support the study of literature, but the approach is based on looking at aspects of transmission that readers usually disregard.

Analytical bibliography is based on the assumption that the physical form of a book contains indications of its history. It has become commonplace in recent years to assume that electronic texts do not contain any such indications. Although electronic texts require some physical medium for their storage and display, we are still inclined to think of them as non-physical, because they are stored digitally – and while it is not possible to make a perfect duplicate of a book or of any analogue representation of a text, it is quite commonplace to achieve that with digital texts. This makes it possible to copy and transmit electronic texts without any of the telltale signs that copying or alteration introduces in reproductions of books, and that analytical bibliography can reveal. For example, in 1974 William B. Todd showed that the transcripts of conversations in the Nixon White House issued by Nixon’s Administration had been extensively doctored; he did this through a study of the changing typefaces and margins in the published volume. It is generally assumed today that it would not be possible to detect such a cut-and-paste job if the volume were produced

using a word-processing program: after all, one reason for the widespread use of these programs is the ease of making such modifications in documents. Furthermore, it is commonly believed, we will cease to have the information once preserved in working drafts, and we will also lose any assurance of the authenticity of texts, because there is no physical object that stands as a guarantee of the work, only an easily-forged electronic text.

But I argue that this is not the case. Electronic texts do preserve significant traces of their transmission, and we can learn to recognize those indications and make deductions from them about the history of texts. These indications are not “physical”, but the way we need to study them is entirely in keeping with the traditions of analytical bibliography, if we understand the field in a way that doesn’t stress the physicality of the evidence.

Analytical bibliography looks like a form of science, in its concern with physical evidence rather than with literary structures; but analytical bibliographers have often said that it is misleading to think of the field as a science (see, for example, the comments of Tanselle). Carlo Ginzburg has presented an argument that helps clarify this question: he proposed that bibliography is one of a number of practices involving, not the generalization and quantification typical of the exact sciences, but instead “an attitude oriented towards the analysis of specific cases which could be reconstructed only through traces, symptoms, and clues” (104). The most familiar example of this attitude is in the method of Sherlock Holmes, who could see in tiny details of a person’s appearance indications of his or her origin, profession, place of residence, and so on. But there are specialists of other kinds who use the same sort of approach, and not just in the realm of fiction: for example, hunters tracking game, specialists in the attribution of artworks, and doctors seeking to diagnose patients (particularly in the period before our century). These are all practical skills based on experience as much as on book-learning, and on the study of “evidence that is imperceptible to most people” (98). And there are many people today skilled in interpreting the history of electronic texts in just this way – though they’re more likely to be found in your university’s computer center than in any academic department.

The evidence this interpretation exploits is not physical evidence – though the physical medium does actually convey some real information (a 9-track tape or a 5.25” disk probably weren’t created by someone using a Macintosh). It is rather in the actual contents of electronic texts, in the digital codes used to represent letters and, even more, in the codes used to indicate other aspects of the text (such as spacing and line breaks). That there is an extraordinary variety in such encoding

is a classic problem for us in humanities computing, but it is also one that most users of computers never give any thought to: therefore putting these details quite squarely in the category of “evidence that is imperceptible to most people.”

These indications can tell us a great deal about what program created a file and what stages of transmission it went through. Analytical bibliography seeks evidence about the techniques of typesetting and printing used in producing a book, because such evidence can help us resolve problems about the book’s history or textual content; similar studies can do the same for electronic texts. Because the systems used for creating electronic texts are implemented in software, rather than requiring the fabrication of new machinery, this form of analytical bibliography faces a broader array of different procedures than physical bibliography did. On the other hand, the systematic analysis of the texts is easier, because they do exist to begin with in electronic form: searches for telltale patterns are much easier to make with computer assistance than by the laborious examination of printed pages.

Here are a few of the types of historical indications that an electronic text may contain:

– *Program of origin.* Word-processor files usually have an internal format that is specific to the word processor in question, and generally even to a particular version of that word processor. This in itself provides some significant information about dating, since a WordPerfect 6.0 file can’t have been created before WordPerfect 6.0 was available. Even the names of files can provide some information of this sort, since it is conventional on many systems for the filename extension (the part after the dot) to indicate the associated program (this is the case on Unix, Windows 95, and VMS, for example). It’s true that you are perfectly free to change the names of your files; but it is generally the case that using the system is made more difficult if you don’t use the standard extensions, and the goal of most users is to get their work done, not to hide their tracks. The same applies to the internal format of files: most users are unaware that there is any variation from program to program or version to version, and even if they are aware they generally see no need to engage in measures to update the formats of their old files unless they need to use them again.

– *System of origin.* Files that are not created by word processors or other programs that use a characteristic internal format, but are in the “plain ASCII” format that is somewhat system-independent, still commonly have variations in their content that indicate their source. For example, such files when created under MS-DOS often have control-Z characters at the end, and lines are divided by a carriage return/line feed sequence; both

conventions are inappropriate on other systems (see Gaylord for further details of this sort).

– *Traces of transmission.* While it is commonly believed that electronic texts travel from system to system and place to place without alteration, this is actually not always the case. Indeed, it is easy to see errors in printed texts that can be directly attributed to various errors in electronic transmission. One of the most common is the ossified soft hyphen: the line-end hyphen that has been turned into a hard hyphen when a file has been moved from one program or system to another, because what was moved was the text with its line breaks and not the stream of words. The upshot is split words in the middle of lines, such as “mis-take”. In recent years electronic mail has become a common means of transmission for plain text, but it can introduce artifacts of its own. A common one is the conversion of the word “From”, when it appears at the start of a line, to “>From”, so as to avoid confusing some Unix mail programs (see Costales 399-400). The introduction of schemes for transmitting eight-bit characters or attachments has multiplied the possibilities for such changes in the text, changes that are easily overlooked if they only occur here and there in a document.

– *Typists and their interaction with programs.* It is often the case that a text’s author is recognizable through habits of spelling, punctuation, or spatial layout; but some of these habits can be conditioned by particular programs. The vi editor on Unix, for example, normally treats the h, j, k, and l keys as cursor keys rather than alphabetic symbols; consequently the vi user who switches to another editor or word processor is likely to salt the text with a lot of j characters.

Few word-processing programs today make any attempt to record a history of changes to a file, but typists nevertheless leave traces of their revisions in most word-processor files. Any file that undergoes much revision usually winds up with a number of superfluous font changes (a change into italic immediately followed by a change back into roman, for example, with no actual text affected); the absence of such incidentals is a sign that little change has been made in the text since it entered the computer, or that it wasn’t typed by a person at all but was converted from some other form by a program.

The chief objection that might be made to the study of these indications is that they could be faked, and that – since only a pattern of bytes needs to be faked, and not anything involving physical materials like paper and ink – it is comparatively easy to do this in an undetectable way. In practice, though, there are two substantial obstacles to forgery. The first is the problem of skill and knowled-

ge: the desire to fake the file format used by a particular word processor doesn't mean you can find out or easily mimic that format, or that you will have the skill to do it correctly.

The second obstacle is that of time and convenience. Most of the time, most of us are just trying to get our work done, and creating a deceptive electronic trail is the farthest thing from our minds. The study of electronic texts reveals a further truth implicit in the tradition of analytical bibliography: that producers of texts do so in a social matrix, constituted by the available technology, and they want to concentrate their attention on writing rather than on the technological means. The forger necessarily takes an interest in the means; but most texts are not produced by people with forgery in mind, and the incentive for the labor of forgery is rarely great. Whether the system we use for text production is mechanical or electronic, we generally want it to do the job while requiring the minimum amount of effort from us.

And even when deception is in question, it is very difficult to acquire enough knowledge to create one that is foolproof. A decade after Richard Nixon's resignation, in a White House that now used electronic mail and word processing, further criminal activities still wound up being exposed because the criminals weren't aware of the traces they were leaving. Some of the main evidence against Oliver North was in his electronic-mail messages, messages he assumed vanished when he deleted them – but which were preserved on backups (see Draper). Electronic texts aren't the free-for-all that we are often led to imagine: just like printed texts, they are created, read, and transmitted within systems we didn't create, that influence us and that we can fully break out of only through the exertion of extraordinary effort.

Works Cited

- Costales, Bryan, with Eric Allman and Neil Rikert. *Sendmail*. Sebastopol, CA: O'Reilly, 1993.
- Draper, Theodore. *A Very Thin Line: The Iran-Contra Affairs*. New York: Hill and Wang, 1991.
- Gaylord, Harry. "Character Representation." *Computers and the Humanities* 29 (1995): 51-73.
- Ginzburg, Carlo. "Clues: Roots of an Evidential Paradigm." 1979. *Clues, Myths, and the Historical Method*. Tr. John and Anne C. Tedeschi. Baltimore: John Hopkins University Press, 1989. 96-125.
- Tanselle, G. Thomas. "Bibliography and Science." *Studies in Bibliography* 27 (1974): 55-89. Rpt. in *Selected Studies in Bibliography* (Charlottesville: University of Virginia Press, 1979). 1-35.
- Todd, William B. "The White House Transcripts." *Papers of the Bibliographical Society of America* 68 (1974): 267-296.
- Williams, William Proctor, and Craig S. Abbott. *An Introduction to Bibliographical and Textual Studies*. New York: Modern Language Association, 1985.