Edward Jacobs

Protocols for Scholarly Proofreading

These protocols are designed to insure that transcriptions of manuscript and printed source texts into machine-readable programs like Microsoft Word correspond EXACTLY to the original source texts. Such accurate transcription is the crucial first step in any scholarly editing project, since the collation, emendation, and annotation that follow transcription can only be as accurate as the transcription of original source or "witness" texts is.

These protocols may seem excessively redundant and detailed, but such redundancy and details are crucial to remove the errors that inevitably occur when a text is transcribed into machine-readable form, whether by Optical Character Recognition (OCR) scans or by re-keyboarding of the original source text.[1]

The basic practice outlined here of proofing texts in teams where one person reads the "original" source text aloud and the other person checks that reading against a "transcription proof" has been standard practice since medieval manuscript culture, and over those centuries that way of proofreading by team has proven far more accurate than eye-proofing by an individual. The specific protocols below systematize that basic practice for maximum accuracy when proofing a machine-readable "transcription proof" of a non-machine readable "original." For convenience, I have below abbreviated the

---

[1] On OCR versus keyboarding, see Fenton, Eileen Gifford and Hoyt N. Duggan, "Effective Methods of Producing Machine-Readable Text from Manuscript and Print Sources," *Electronic Textual Editing*, ed. Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth (New York: MLA, 2006): 241-53.

person who reads the "original" aloud as the "Reader" and the person who checks that reading against the "transcription proof" as the "Proofer." Examples of text from the original are put in brackets rather than quotation marks.

IMPORTANT NOTE: The "transcription proof" from which the Proofer works can be either a printout of the pre-proofreading transcription file or a separately titled (and saved) copy of that pre-proofreading transcription file in which Word's "track changes" mode is turned on. Both of those options keep a visible record of transcription errors found during proofreading before they are confirmed and definitively entered into transcription file. By contrast, if during proofreading the Proofer directly enters (i.e., not in track changes mode) corrections found during proofreading into the base transcription file, those "corrections" become irretrievable and cannot be rechecked against the original. And they need to be rechecked against the original because Proofers may well inaccurately record needed corrections on/in the "transcription proof" during the (complicated) process of team proofreading.

Proofreading Protocols

1. Reader reads the first paragraph BACKWARD word by word, pronouncing all punctuation, all upper cases, all special fonts or characters, and all spellings of words that sound like another word or spelling, as for example [us'd] rather than "used" or [their] rather than "there" or "they're." [NOTE: BACKWARD readings is necessary because it mitigates "pattern recognition," whereby both the Reader and the Proofer will perceive the original/transcription proof to conform to the syntax and spelling to which they are culturally and personally accustomed.]

2

1.1 Special fonts should be stated AFTER reading the word so formatted. For example in reading aloud the original text [*used*, finally.], the Reader should say "PERIOD FINALLY, COMMA USED ITALICS."

1.2 Individual upper case letters should be identified by saying "CAPITAL" plus the letter AFTER saying the word. By contrast, entire words in upper case should be identified by saying "ALL CAPS" AFTER saying the word. For example, in reading the original text [It is], the Reader should say "IS IT CAPITAL I," whereas in reading the original text [IT is], the Reader should say "IS IT ALL CAPS."

1.3 Spellings of words that sound like other words or spellings should be given AFTER reading the word. For example, in reading aloud the original text [their Villanous kinde], the Reader should say "KINDE K-I-N-D-E VILLANOUS CAPITAL V-I-L-L-A-N-O-U-S THEIR T-H-E-I-R."

2. As the Reader reads the original aloud as stipulated above, the Proofer checks that reading word by word against the transcription proof and says "STOP" whenever the transcription proof differs from the Reader's reading or the Proofer has queries about the Reader's reading. After saying "STOP," the Proofer asks the Reader to confirm or clarify how the original reads and marks any needed correction on/in the transcription proof. (See IMPORTANT NOTE above and protocol five (5) below: the Proofer MUST NOT invisibly enter proofreading correction directly into the transcription proof during the proofreading process). When the Proofer has thus confirmed and marked any needed correction, the Proofer says "OK" and the Reader continues.

3. After Reader has finished reading the first paragraph BACKWARD as stipulated above, the Reader begins reading the paragraph FORWARD, pronouncing all

punctuation but NOT indicating upper case, font, or homonymic spellings as during BACKWARD reading. When and if the Proofer has queries about whether the FORWARD Reader reading makes syntactic sense, the Proofer says "STOP" and ask the Reader to confirm how the original reads as stipulated in protocol two (2). If the Reader confirms an error in transcription not identified by BACKWARD reading, the Proofer marks the needed correction on/in the transcription proof and when finished says "OK" and the Reader continues. (See IMPORTANT NOTE above and protocol five (5) below: the Proofer MUST NOT invisibly enter proofreading correction directly into the transcription proof during the proofreading process). If the Reader confirms that the original actually reads in the way that the Proofer found nonsensical, the Proofer says "OK," the Reader writes "EMEND?" in the margin of the original next to the passage, and the Reader then continues. (These "nonsensical" readings may need to be emended after collation, which is why the Reader notes them on the original. However, because the aim of proofreading transcriptions is to make them identical to the original, nonsense and all, the Proofer MUST NOT correct these "nonsensical" readings in the transcript proof). [NOTE: FORWARD reading is necessary because more than BACKWARD reading it activates the critical judgment of the team about how well the original corresponds to the syntax and spelling to which they are culturally and personally accustomed, and thereby may identify transcription errors not caught by BACKWARD reading].

4. The team proceeds to the next paragraph and follows the above protocols for it and all subsequent paragraphs.

5. When the team has finished proofreading, they enter all corrections into the machine-readable file of the transcription, effectively switching roles. Working from the

corrections marked on the transcription proof, the Proofer reads each correction to the Reader, giving the Reader the best possible of citation of where the correction occurs. When the Reader has found the correction site in the original, the Reader says "FOUND" and rereads the original according to the protocols in section one (1-1.3). The Proofer then enters the correction into the transcription proof, confirming and querying the Reader's reading of the original as stipulated in protocol two (2). The Proofer then saves the transcription proof and the team moves on to the next correction marked on the transcription proof.

Practical Notes

1. The above protocols implicitly assume that one team member will be the Reader and another will be the Proofer throughout the proofreading process. In practice, however, it is more effective for team members to switch roles every half-hour or hour, so as to avoid fatigue in one mode that may produce errors. Also, from a pedagogical point of view, it is best for all team members to act both as Reader and Proofer. If you are working in teams of more than two people, best practice is to have two or more Proofers simultaneously check different printouts of the "transcription proof" against a single Reader's speech. However, if you have difficulty coordinating so many schedules, the alternative is to work in rotating teams of two.

2. For teams that involve distance students, switching roles is complicated by the need to mark correction on a common transcription proof. Work this out as you deem best, but probably the simplest solution is for the team to use a track changes Word file as their transcription proof and email it between them each time they switch roles as

Reader/Proofer. A similar email exchange process could be used to switch roles during the final entry of corrections described in protocol five (5), although you may find it easier to each keep the same role during that process, depending on how many corrections there are to be entered.

3. It is not a good idea to proofread in sessions much longer than three or four hours. Proofreading is tedious but necessarily so, and as humans we get tired of tedium. So please call a stop when you judge that yourself or your team member(s) are getting too tired to be effective.