Conventions for Transcribing, Abstracting, or Extracting from Primary Documents

Although transcription is understood to be "literal", any attempt to copy an original document, short of photoreproduction, is bound to lose some of the original structure, and probably a few juicy textual ambiguities as well. So why not just publish the image itself? Because even the skilled reader can probably use some help in deciphering it. Reading ancient manuscript documents is an exercise in painstaking expertise, and even extensive experience may be an inadequate preparation for dealing with the idiosyncrasies of some particular backwoods colonial American scribe.

On the other hand, to just leave it up to the expert reader to tell us what the document is saying, seems to unduly privilege his interpretation over that of other scholars, or even amateurs who may have a deeper understanding of the referential context. After all, everyone's mind has a latent tendency to see what it expects to see (usually what it is used to seeing), and even expert readers (despite their familiarity with the hand of a particular scribe) can still misread the text, especially when it does not fall into the normal pattern s/he is biased to look for.

We know too, from modern psychological studies, that every time a text passes through the filter of an interpretive mind, opportunities for error arise. That is why one nearly always prefers a primary source to a secondary, and the original writing to the most meticulous copy. But even the expert genealogist must rely on these substitutes in large part, for reasons of time and money, and it is important, therefore, to define criteria for determining when substitutes are acceptable. Mine are: care and expertise in the transcription according to a set of explicit stated rules, and a detailed reconstruction of the records context, particularly where the material is being abstracted.

Until recently, few works of genealogical transcription have met all these criteria, and many, none of them. Modern examples of how the job should be done are: Roger D. Joslyn's *Vital Records of Charlestown...*, and Robert H. Rodgers, *Middlesex County ... Records of Probate and Administration ....* Only where the above criteria are met (care and expertise in reading; explicit transcription rules; characterization of records context) can one appropriately rely on transcriptions or abstracts of primary documents.

With the important topics of how best to preserve and explain records context, and how to abstract, I am not here concerned. No one, to my knowledge, has even begun to address these questions in any definitive, or even systematic way, although I have encountered a few collections of published abstracts that do an adequate job, not only of abstracting, but also of specifying their own rules, and of describing the records structure upon which they are operating.

Transcription "Methods"

Scholarly editors have devised sets of conventions, or "methods", of transcription - in fact they have devised a wide spectrum of these, ranging from the most literal, through an "expanded" literality, to the most modernized. Perhaps the most influential of the standard sources on these matters, the *Harvard Guide to American History* (through several editions), has defined the span of the methodological spectrum in just these terms: literal, expanded, and modernized[1].

Modernized Method (MM) is undoubtedly the most attractive to the reader, because it is the easiest to read. Literal Method (LM), on the other hand, attempts at least as faithful a transliteration of the original as modern typography will support. Expanded Method (EM) is a compromise format, which relaxes this strict literalism for the sake of readability, primarily in expanding abbreviated words "silently" (without the editorial notice of square brackets) where, in the opinion of the expert transcriber, there can be no

---

[1] Other generic guides to scholarly transcription: Mary Jo Kilne, *A Guide to Documentary Editing*, 2nd ed. (The Johns Hopkins University Press, 1998). Michael E. Stevens and Steven B. Burg, *Editing Historical Documents* (Altamira Press, 1997)

reasonable doubt as to what was intended. Of course in all three methods, no words are added to or subtracted from the original, and the usual standard editorial conventions are followed.

Which is best? In my opinion, the answer is: Modern Method, even for most scholarly purposes, but with an important proviso: *that a photocopy of the original be readily available to consult at the same time.* This might seem the Proviso the Ate New York, except that by now a large proportion of the primary source documents are available online, and low-light digital cameras and scanners have put this technology into the hands of the average serious genealogist. One of the reasons my own publications are in PDF format is so that I can provide links to key source documents that aren't worth embedding within the publication itself, and it is to be hoped that future print publishers of transcriptions and abstracts would take care to provide URLs as well as explicit citation criteria for all important primary records evidence.

The basic reason for not fussing with quasi-literal compromises like "expanded method" is that, while we may or may not have both the time and the expertise to study the original ourselves, when we consult a work of transcription, or a collection of abstracts, or even an index (all are species of the same genus), we are implicitly relying on the skill and care of the expert who has produced the work, and there is no reason not to make the fullest possible use of his labor and expertise. If we have a certain amount of time in our research project to consider a particular document, better that we spend it contemplating its meaning and import, than in struggling with its idiosyncratic and inconsistent spellings, punctuation, and chirography.

The original material on transcription methods in the *Harvard Guide* was written in the early 1950's by eminent historian, Samuel Eliot Morison, who published at the same time, the authoritative modern scholarly version of Bradford's journals—using the Modern Method of the *Guide*. And since the early 1990's eminent New England genealogist, Robert Charles Anderson, has chosen to use MM in his magisterial *Great Migration* series, even for brief extracted snippets from original records. Although it is true that my 1974 edition of the *Guide* recommends LM or EM for scholarly readers, its final word on the subject is: "Every printed document ... should be compared word for word with the original, or with a microfilm or photographic copy ...." (one hopes "every *important* document" was meant), and if this is to be done anyway, what is the use of an intermediate fussy and literalistic transcription?

Well, there is one use, I suppose: to impress the reader with one's own care and expertise as a transcriber; perhaps LM or MM are justified where this is in question, and probably LM would be more impressive than MM.

The only good reason I can think of for using LM or MM, though, is when reproductions of the original are *not* readily obtainable, or not economically so—but this covers perhaps the majority of note-taking situations, which is why I use a very literalistic method in my notes when it's not practical to make photocopies. One way or another, a very exact and literal version of the original needs to be available to back up the interpretive transcription, and when the latter is at all problematic, the literal text should be supplied in footnotes or interpolated into the transcription itself, within editorial square brackets

Although the *Guide* defines its LM, EM, and MM in some detail, thus erecting a standard that has been widely influential for the last 70 years, the fact is that every scholarly editor tends to go his own way on the details. The justification offered—that the rules must be tailored to the particularities of the subject matter—can only be judged by the particular results. The *Guide* also proposes this motto to guide scholarly editors: "Accuracy without pedantry. Consistency first, last, and always."

Unfortunately, most genealogical publications today, even the most prestigious, offer only the sketchiest of accounts of their transcription methods. This is a defect I intend to remedy in my own work. Accordingly, I have defined below my own versions of MM and EM (RMM and REM), with at least as much rigor as is to be found in the *Harvard Guide*. I have also worked out my own ultra-literal RLM, which I generally use only for notes where it is not practical or worthwhile to make a photocopy of the original document; my specifications for this appear in a separate, unpublished, document.

Robb Expanded Method (REM)

No words will be added to, or subtracted from the original.

Spelling and punctuation are transcribed literally from the original, except that a period (or dash) will be interpolated after the last word of an identifiable sentence (or sentence fragment) in lieu of extra white space, or other graphic clues in the original. Capitalization, on the other hand, will always be modernized, though where this might be problematic, the exception will be noted

Abbreviations are expanded silently, except for standard abbreviations associated with proper names, titles, and places: these are expanded within square brackets, but otherwise modified only by bringing superscript abbreviations down to the line.

Emphasis in the original will be represented thus: underlined text by italics; text underlined twice by small caps; text written large by large caps; and signatures by large and small caps.

Transcribed text appears in normal type, with italics reserved for the editorial voice. Text that can't actually be read, but which can be reasonably inferred, perhaps by taking into account other sources or comparable records, will appear in normal type within square brackets, along with any italicized editorial commentary.

Normal type within square brackets thus represents an editorial completion of what is actually readable in the original—by expansion of an abbreviation, or by interpolation of missing or illegible characters or even a whole word, which can be inferred with confidence from context, or perhaps from an associated or companion document. Where such inference is problematic, the bracketed text will be supplemented by an editorial "?", or where alternate readings are possible, they will be separated by editorial "|"(OR sign)s. Both these special characters will be in normal type, even though, strictly speaking, they should be italicized, since they represent the editorial voice; the exception is made in the interests of keeping a low editorial profile.

Editorial commentary within the transcription itself is confined to italicized text within square brackets. It is thus that we may render, e.g. the editorial [*sic*], which should, like other editorial interpolations, be used very sparingly to avoid interrupting the flow of the text itself. Italicized editorial commentary may more appropriately be used to frame the transcription or abstract, or it may be relegated to footnotes.

The only textual transformations undertaken silently are these: conversion of "ff" at the start of a word to the capital "F" that is meant; transformation of the "y"thorn to "th" (e.g. "ye" to "the"); resolution of ambiguous "u"|"v" or "i"|"j" letters to conform to modern spelling; replacement of "&" with "and:, except in the names of businesses or partnerships, and of "&c." with "etc.".

Certain special characters may be interpolated editorially into the original text, without brackets. Ellipsis dots are used in the normal way to indicate skipping of part ("..."), or the rest of ("...."), the current sentence, while omission of one or more whole paragraphs is marked by ". . ." centered, on its own line. I have supplemented these standard editorial conventions by substituting "*"s for illegible characters in the original, and "_"s where unexpected blank spaces appear in the original, one for each.

Interlineations will be silently merged into the text at the intended point.

Where marginal text is lost due to paper damage or tight binding, and must be reconstructed from context, it will appear within square brackets in normal type, prefixed or suffixed by the "‖" character to indicate where the readable margin ends, or begins, eg "[‖ing]", or "[pre‖]".

No gratuitous paragraphing or other formatting is imposed, but extended dashes or interpolated lines in the original will be interpreted as intrinsic paragraphing.

Each turn to a new page may be noticed editorially by, e.g. "[*page 123*]", "[*folio 81*]", or just "[*new page*]" where there is no pagination in the original, but this will be done only at discretion.

Indications that a signature was by mark will be rendered [*mark*].

Robb Modernized Method (RMM)

Spelling, punctuation, and capitalization will be modernized, and standard or unambiguously recognizable abbreviations are expanded silently, except for proper nouns (personal, organizational and place names) whose abbreviations are expanded within editorial square brackets. Where such expansions are ambiguous, the original will either be rendered literally, or a choice of expansions will be offered within square brackets, using "|"(OR) as a separator for the alternate text, and enclosing the text itself within double quotes to distinguish it from the editorial voice.

Whenever the reading of the original is doubtful, any illegible characters will be replace by "⁑", and possible readings may be offered with square brackets, with the text enclosed in double quotes.

The name suffixes "Junior", "Senior", "Esquire" or "Gentleman", or their recognizable abbreviations will be standardized to their most common modern forms: "Jr.", "Sr.", "Esq.", and "Gent."

Honorific or rank prefixes will likewise be abbreviated in standard modern form, e.g. "Mr.", "Col."

Evident slips of the pen, and duplicated words or phrases are corrected silently.

In rare cases where it is necessary to comment on, or juxtapose, the raw original text, such text will appear in square brackets in normal type but be double-quoted, and will be transcribed in REM.

Dates are normalized to ddMmmyyyy format, with double-dates made explicit as in these examples: "10Jan1750/1" if explicit in the original or nearly certain from context; "10Jan1750[/1]" if the inferred year is merely probable; "10Jan1750[/1?]" if this is the most likely year, but there is significant doubt, or just "10Jan1750" (leaving the ambiguity of the original unresolved) if there are insufficient grounds for making a judgement. Partial, referential, dates, like "March second next ensuing" will be rendered, e.g. "March 2nd next ensuing".

Weights and measures are standardized thus: dollars and cents as "$dd.cc"; monetary British pounds as "£", shillings as an appended "s" (e.g. 5s), and monetary pounds, shillings and pence written as a single amount as "£pp/ss/pp ". Avoirdupois pounds are just "pounds". The word "acre" is appended as a suffix "a" to the quantity, which is decimalized if necessary, e.g. "153.5a", and perches or poles (both represent 16.5 feet) are both converted to the suffix "p".

Directions of the compass are normalized to, e.g. "N", "NW", "N30W", at least within the metes and bounds of deeds, but at discretion otherwise.

A reasonable, but conservative, paragraph structure will be silently imposed where this is lacking in the original.

In addition, for convenience of eyeball scanning, certain typographical devices have been adopted.

Multiline items comprising lists are each prefaced by a bullet head, e.g. the bequest items in a will.

Names of persons have been promoted to boldfaced type (except for parts interpolated in square brackets), and excepting also names of officials.

And finally, certain key boilerplate words or phrases (especially if highlighted in the original), and all the text of pre-printed forms is rendered in small-caps (e.g. for wills "IN THE NAME OF GOD, AMEN", for deeds, "BEGINNING..."). These special conventions will generally be defined in the front matter for each published report.

Additional Rules Common to Both REM and RMM

Transcriptions of documents copied into deeds, probate, or other court books, will be supplemented by the text of any associated oaths, declarations, or other endorsements, except where the endorsement conveys no additional information, such as a date of recording that differs not from the date of a document, or of its witnessing.