



Transcription Manual

January 2013

Reviewed and Updated Annually

This guide is borrowed very, very heavily from *Transcribing Manuscripts: Rules Worked Out by the Minnesota Historical Society*, adapted in 2004 by Lydia Lucas from the original pamphlet by Grace Lee Nute (St. Paul: Minnesota Historical Society, 1935). Thanks, MHS!

INTRODUCTION

Transcription of handwritten manuscripts is usually undertaken with two goals in mind: to give a truthful representation of what the writer actually wrote, and to make the printed copy easy for the reader to comprehend. When used in a digital library, transcription is important to provide a free text searchable account of a digital image.

In transcription, the transcriber should make him/herself as inconspicuous as possible in the transcribed document, while still adding those annotations that may be needed for clarification. This manual introduces some of the standards and annotations used in transcription.

This manual is divided into two sections: XML transcription and PDF transcription.

- PDF transcriptions are used when the institution wants to make a transcript of an item that contains special characters and formatting and is human-readable. An example of this would be an oral history. PDF transcripts are not machine-readable or full text searchable. For oral histories, PDF transcriptions can be used to generate an XML transcript, making the oral history both human and machine-readable.
- XML transcription is used when the institution wants to make a document full text searchable, like a handwritten letter. XML transcriptions contain no formatting or special characters. Any intended special characters must be encoded.

XML TRANSCRIPTION

Unless you are transcribing multimedia items, such as oral or video histories, XML transcription is the most common form of transcription you will have to undertake.

All special characters are to be avoided. As such, the following rules apply:

- Convert ampersands to the typed word “and”
- Quotations, underlining, italics, parentheses, etc. are to be excluded whenever possible. If you believe your transcription needs to contain any special characters, please let LCDL know and we will work on encoding your machine-readable transcriptions before ingestion.

FORMATTING THE XML DOCUMENTS

Transcriptions should be prepared using a code capable text editor without any formatting or use of special characters.. There are several free options available:

- NotePad ++ is available for Windows based systems
- TextWrangler is available for Apple computers

Eliminating formatting and special characters helps to ensure that the text can readily be transported from one presentation environment to another, as needed or desired. The new LCDL is built upon a Fedora Commons repository. Fedora uses XML files for machine-readable transcripts. Therefore all transcript files must now be saved as .xml files. Each .xml file must have a mandatory header and footer. When saving the .xml file, please select “UTF-8” as the preferred encoding. Please name your .xml file the exact same name as the image file you are transcribing from. Therefore if you are transcribing “letter02.jpg”, your transcript file is named “letter02.xml”

MANDATORY HEADER AND FOOTER TEXT For XML TRANSCRIPTIONS

Every XML transcript file created for inclusion in the LCDL must have a specific MODS header and footer tag.

The header is always:

```
<mods xmlns="http://www.loc.gov/mods/">  
<note type="transcription">
```

The footer is always:

```
</note>  
</mods>
```

(The actual transcribed content is typed between the header and the footer.)

PDF TRANSCRIPTION

PUNCTUATION AND CAPITALIZATION

Punctuation and capitalization call for many decisions. There are two overall approaches: to use modern practice, or to adhere to the style of the original manuscript. Whichever approach is chosen should be applied consistently throughout the document and/or the entire transcription project, and this decision should be recorded as part of the project documentation. Modern practice is appropriate for reminiscences and other formal and semi-formal items. Transcriptions of letters, diaries, and the like should adhere to the original items as closely as is practical, since this helps preserve for the reader a sense of the character of the originals.

PERIODS

In transcription, dashes at the ends of sentences are regularly converted into periods, provided it is clear that the dash was intended to serve this purpose. Periods or dashes under raised letters are omitted in transcribing. In carelessly written documents, periods often resemble commas. At the end of a sentence, such an ambiguous mark is interpreted as a period. If no mark of punctuation is used at the end of a sentence, just leave several spaces.

DASHES

A dash is indicated thus in typing: -- (many word processing programs will convert this to a long single dash). When the author omits letters in proper names, you may use a triple or quadruple hyphen (so the word processor won't convert it).

QUOTATION MARKS

These are used only when an actual quotation (enclosed in quotation marks) is found within the document that is being transcribed. Use the double quote marks ("). They follow a question mark or an exclamation mark that is part of the quotation; they precede a question mark or exclamation mark that is not part of the quotation. They always follow the comma and the period whether or not these are part of the quotation. They always precede the colon and semicolon.

PARENTHESES AND BRACKETS.

In general, parentheses are used when they are part of the original document only, and ***square brackets are used for insertions by the editor or transcriber***. If the writer has omitted one curve of parentheses, it may be supplied in brackets, [)]. Usually square brackets in the original text can simply be changed to curves (. . .) without further annotation. See below for other uses of square brackets.

SMALL CAPITALS AND ITALICS.

Do not notate small caps or italics in the transcription.

UNDERLINING.

Do not transcribe any underlining.

PARAGRAPHS.

Paragraphs may either be indented, or left flush with the left-hand margin with a blank line between them. Whichever convention is chosen should remain consistent throughout the transcription. In some instances, however, retaining variances in indentation may help convey the character of the original document.

TRANSCRIPTION CONVENTIONS

When transcribing manuscripts and diaries, the copyist should adhere as closely as possible to the original format, spelling, and overall presentation of the original, enclosing supplied information in square brackets.

MISSPELLINGS

Sometimes a writer makes a mistake in spelling, and occasionally in grammar, that the reader might reasonably construe as the transcriber's error. Then [sic] is typed immediately after the word. But if the reader would not question the copyist's carefulness in transcription, mere misspellings are left without any such remark.

In instances of a grossly misspelled word, the transcriber should supply the correct word in brackets following the misspelled word: "qeshun [question]." However, if a manuscript is filled with misspellings, it is best to transcribe it just as is, reserving [sic] and corrections just for those cases in which the reader might become confused or suspect a transcriber error.

MUTILATED OR ILLEGIBLE MANUSCRIPTS

If a word or words can be guessed from the context, the expression is enclosed in square brackets.

Thus: "it is [inevitable] that this will occur"; or "he [had] fallen."

Uncertain but probable guesses are followed by a question mark,

for example: "in the [Indian?] wars."

If no guess can be made, an appropriate remark may be supplied in this manner:

[MS. burned]

[MS. blotted]ke War

se[illegible]ing ships

preserve [illegible] enlightenment

In the afternoon I went [page torn] and helped with the threshing.

UNCERTAIN READINGS

If a reading is uncertain, the entire questionable word or phrase, with its question mark, may be enclosed in square brackets:

rendersveu [rendezvous?].

Bear [beer?]

Survival of the [species?]

OMISSIONS/INSERTIONS

Whenever possible, the transcriber should supply words or punctuation, the deliberate or careless omission of which by the author causes difficulty in reading. This is most commonly done where it seems useful to preserve the flow of the text as an aid to the reader. The supplied portions are enclosed in square brackets; they are not italicized.

I started [to] go home.

In the afternoon I went [to] the village.

TRANSCRIBER'S COMMENTS

Comments that relate directly to the text are inserted at the point where the comment is needed. They are enclosed in square brackets.

[written sideways in margin]: More Prisoners [sic] 425

[“1874” is written opposite this word in the margin]

REPETITION OF WORDS

Like misspellings, inadvertent repetition of words by the author is indicated by [sic] following the extra word or words:

“care must be taken in in [sic] this matter.”

In old documents, the first word on a page is often repeated in the lower right-hand corner of the preceding page; in this case the word is not copied twice.

DATES

Dates centered in the manuscript are dropped to the beginning of the following paragraph.

RAISED LETTERS

Raised Letters are often represented in transcription by a superscript. At the LCDL, the transcriber should always choose to fill in the remainder of the word, bringing the raised letters down to the line.

For example, “Phila^r” would be “Phila [Philadelphia]” if filled in.
Wm^l would be Wm. [William] if filled in.

BLANKS IN TEXT

If a space for a date, figure, or other data is left blank in the manuscript, the form [*blank in MS.*] is used.

For Example, “Personally appeared before me this [*blank in MS.*] day of January. . .”

WORDS CROSSED OUT

In manuscripts, words are often found crossed out by the author. In such cases, the text should be followed exactly; however, our digital library software will not allow the use of strike through text. Therefore the crossed out word is omitted. If the crossed out word is an obvious potential search term, for example a proper or location name, then include that word in brackets.

For example, if the letter is thus: “I headed ~~down the road~~ to Columbia, SC” then the transcription may read “I headed to Columbia, SC” but if the letter reads like this: “I headed down the road ~~to Columbia, SC~~” then the transcription may read “I headed down the road [to Columbia, SC]”.

INITIALS

Initials should be filled out whenever the reader would be confused without the additional information.

For example, if in a particular situation “R.R.” stands for “Red River” rather than the more expected “Railroad,” the correct rendering would be “R.R. [Red River]”

SEALS

When a seal appears in a document, it should be indicated thus: (seal). When a representation of a seal is found on a copy of a document, it is indicated by: (seal mark).

POSTSCRIPTS AND ADDRESSES

When transcribing portions of a manuscript other than the text itself, the copyist may employ the conventions given below. They are typed flush with the left-hand margin, enclosed in brackets, and italicized.

[*P.S.*] is the abbreviation for postscripts. After it are copied all remarks added after the document was signed, whether they are marked as postscripts or not. If the original writer included the abbreviation P.S., then the brackets are omitted and the writer's style is followed exactly.

[*From:*] precedes, in the copy, any statement on the address sheet by the person who sent the letter.

[*Postmark:*] The postmark on a letter or envelope is copied only when it gives a necessary clue to the date or otherwise adds to the reader's information.

[*Addressed:*] introduces the address on the letter's cover or envelope.

[*Endorsed:*] Formerly the person who received a letter or document wrote on the back whatever information would tell him at a glance the author, date, and contents. Such information should be copied.

PDF TRANSCRIPTIONS - APPENDIX A

Standard Abbreviations for original documents:

Standard abbreviations for the most common types of original documents include:

MS. Manuscript

A.D. Autograph Document [basically, any handwritten manuscript that is not a letter, memorandum, or diary]

A.D.S. Autograph Document Signed

A.J. Autograph Journal

A.L. Autograph Letter [handwritten; writer is known but the letter is not signed]

A.L.S. Autograph Letter Signed

A.N. Autograph Note

Df. Draft

D.S. Document Signed [the body of the document is written or printed by someone other than the signer]

L. Letter

L.P.C. Letter Press Copy

L.S. Letter Signed

M. Memorandum

T.L.S. Typed Letter Signed

SOME EXAMPLES.

John Smith to William Crooks, August 5, 1840

[American Fur Company MSS. – L.S.]

Charles Harris to William Clark, [January 5, 1830]

[Clark MSS. – A.L.S.]

(the date has been supplied by the editor or the cataloger)

Diary of Jedediah D. Stevens, September 9, 1829-April 2, 1830

[Stevens MSS. – A.D.]

John Cole to ----, January 1, 1823

(the recipient's name is not known)

[John Cole] to James Groghan, January 1, 1823

(the author's name was supplied by the editor or the cataloger)

General store account book of John Cole, 1868-1870

[Cole Papers – A.D.]

Reminiscence of John Cole, [ca. 1875]

[transcribed in 1985 from original in Cole MSS., Wis. Hist. Soc. – A.D.S.]

(date of reminiscence supplied by the cataloger)