Best Practices & Formatting Requirements

Updated August 2022

Table of Contents

INTRODUCTION

In the past, LCDL accepted xml and PDF transcripts for all transcribed materials regardless of format (whether, for example, they were images of handwritten letters or an audio recording). Since then, we have updated our platforms and now accept transcripts in different formats for different material types. Our new platform for audio and video recordings will offer enhanced capabilities such as syncing time-stamped transcripts to the audio file and adding indexing for more efficient searching. Therefore, it was time to refresh this manual.

This manual will cover best practices for transcribing and transcript formatting options for written materials (pages 3-9) and audio/video materials (pages 10-14). For a/v materials, transcription and indexing will both be covered.

We are always open to constructive feedback on any of our documentation, so if you find any portion of this difficult to understand or in error, please reach out so we can adjust as necessary.

BEST PRACTICES FOR TRANSCRIBING WRITTEN MATERIALS

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 also important because it provides a free text searchable account of a digital image.

In transcription, the transcriber should make themselves as inconspicuous as possible in the transcribed document, while still adding those annotations that may be needed for clarification. Below are some common best practices typically used in the transcription and annotation of written materials.

Much of this section is sourced 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).

Transcribing Common Symbols and Punctuation

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 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: -- However, many word processing programs will convert this to a long single dash. You may use a triple or quadruple hyphen, so the word processor won't convert it. Example: -- can convert automatically to –   To avoid this, simply type ---
In written documents, all dashes can be transcribed using the triple or quadruple hyphen method. In transcription, dashes are also used when the author omits letters in proper names.

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 (").

Parentheses & Brackets

In general, parentheses are used only when they are part of the original document ***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 & Italics

Do not notate small caps or italics in the transcription.

Underlining

Do not transcribe any underlining.

Paragraphs

Paragraphs may not be indented, and paragraph breaks may not be inserted.

Transcription Conventions

When transcribing manuscripts and diaries, the copyist should adhere as closely as possible to the original spelling, and overall presentation of the original. Information supplied by the transcriber is enclosed 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: "queshun [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.

Damaged Or Illegible Manuscripts

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

For example: "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.

For example:

[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 and a question mark may be enclosed in square brackets.

For example:

rendersveu [rendezvous?].
Bear [beer?]
Survival of the [species?]

Omissions/Insertions

Whenever possible, the transcriber should supply words or punctuation in cases where the author's deliberate or careless omission 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.

For example:

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.

For example:

[written sideways in margin]: More Prisaners [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:

For example:

"care must be taken in in [sic] this matter."

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^w" would be "Phila [Philadelphia]" if filled in.

Wm 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.

[*P.S.*] is the abbreviation for postscripts. After this abbreviation 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.

Formatting for Written Transcripts

This section of the manual addresses how to format transcriptions for LCDL that are created from written materials, such as handwritten letters and other types of handwritten documents. If you have print material, please contact your LCDL project coordinator as we can likely create transcripts for you and, therefore, you can potentially avoid manual transcription of typed documents.

Adding Transcripts to the Metadata Spreadsheet
Transcripts for written materials should be typed directly into the transcript column of the LCDL metadata spreadsheet. One-page objects can be created as single-image objects. All other objects, 2-pages and above, that will include transcriptions need to be formatted as compound objects (not multi-image objects). The compound object format will allow transcription to sync with each page of the written material and appear alongside each page in the LCDL catalog.

Example
A four-page letter will be
- formatted as a compound object
- leave the Transcript column for the hook record blank
- use four cells in the Transcript column to transcribe each page of the letter

	Q	R	S	T	U
	Date Digital	**Copyright Status Statement**	**Access Note**	**Transcript**	**Original Filename**
	2021-06	Public Domain.	For more information and re-use requirements contact the South Carolina Historical Society, Charleston, SC, 29401.		43-225_Lamboll_000.jpg
	2021-06	Public Domain.	For more information and re-use requirements contact the South Carolina Historical Society, Charleston, SC, 29401.	Charleston May 25th 1822 My dear Sister, I hope you not been made uneasy by the delay of this letter for its safety but it has been unavoidable if I collect rightly I mentioned in my last by the G.P. that it might be delayed a little on account of the uncertainty of collecting money and this	43-225_Lamboll_001.jpg
	2021-06	Public Domain.	For more information and re-use requirements contact the South Carolina Historical Society, Charleston, SC, 29401.	certainty about that, but whether he will take any further notice of it at this time we can't tell for he has never been to town since the winter I suppose had good or bad reasons rather for being obliged to stay out of it. They were to set off for the upper county soon or as soon as I	43-225_Lamboll_002.jpg
			For more information and re-use requirements contact the South Carolina Historical Society, Charleston, SC, 29401.	support her it would not be wonderful if she goes in a [illegible] and Wm L. J is so [illegible] telling her what she thought she might apprehend from such a [illegible] of confinement and indeed from all her symptoms I think there must be great danger of it. Poor Mrs	43-225_Lamboll_003.jpg
	2021-06	Public Domain.	For more information and re-use requirements contact the South Carolina Historical Society, Charleston, SC, 29401.	[Postmarked:] 50 Mrs. Elizabeth L. Gilchrist Philadelphia CHALSN SC MAY [Writing on the bottom of the page] I must apologize for the writing but my eyes are in a watched state and I have had to write in great haste, I wrote nearly the whole of this letter yesterday and last night	43-225_Lamboll_004.jpg

Hook cell left blank

Four page-level transcriptions synced with the page-level metadata

Reminder: What is and isn't allowed in the transcription field?

- Paragraph breaks are not permitted; therefore, you cannot use the Return or Enter button on your keyboard when creating transcripts. Use of a paragraph break will result in the loss of the transcript after the paragraph break.
- Underlining and strikethrough are not permitted.
- Special characters are allowed. This includes, for example, single and double quotation marks, symbols for currencies, mathematics symbols, the copyright symbol, etc.

Will typing transcripts directly into an Excel cell prove difficult for you?

Use a text editor to draft them! For a PC, the default text editor is Notepad. For a Mac, the default text editor is TextEdit. In TextEdit, ensure you're in plain text mode by opening TextEdit, clicking Format, and clicking Make Plain Text. You are not in plain text mode. If the only option is Make Rich Text, you are already in plain text mode.

Type your transcript in plain text, and when you are ready, cut and paste page-level transcription into the appropriate Transcript cell in the metadata spreadsheet. If you have separated your pages by paragraph breaks, when you copy and paste, be careful not to bring over any paragraph breaks into the spreadsheet.

BEST PRACTICES FOR TRANSCRIBING A/V MATERIALS

Introduction

The vast majority of A/V materials in the Lowcountry Digital Library consist of oral history recordings. Therefore, in this section of the manual, we will focus on best practices for that format. However, these best practices can be applied to any a/v recording transcribed for the Lowcountry Digital Library.

The primary record of an oral history interview is the audio or video recording. Transcripts should provide a verbatim guide to that recording, reflecting the actual words, speech, and thought patterns of the interviewee. Transcripts may be reviewed by interviewees for corrections to place names and proper names, but otherwise should remain unedited, unless for clarifications making the interview more accessible to researchers.

Much of this section was sourced from Baylor University Institute for Oral History's *Style Guide: A Quick Reference for Editing Oral History* Transcripts and Guilford College's *Oral History: Best Practices and Procedures* guide. While this manual should be adequate for most if not all of your transcription needs, we highly recommend Baylor University Institute for Oral History's style guide if you are looking for a more thorough resource.

Speech Patterns and Language

General Guidance
Change as little as possible. The narrator's word choice, including grammar and speech patterns, should be accurately represented. Verbatim renderings of slang and regional pronunciations are the prerogative of each project, but, if used, should be consistent.

Connecting Words
Most people are unaware that they use words such as "and," "but," and "so" to connect their sentences. They should be left in unless they become overwhelming.

Accents and Dialects
Do not try to reproduce accents or dialects. Use contractions only if they are used by the speaker.

Crutch Words

Words such as "you know," "you see," or "like" should be left in unless they become overwhelming.

Fillers
Leave out fillers such as "ah" and "um" since they do not really reflect a speech pattern.

False Starts
Include false starts because they are often indicative of thought and speech patterns. They may be deleted, however, if the false start is a repetition or a stumble, or if the speaker stutters. Example: "Well I–We didn't hear anything about that."

Unfinished Thoughts
Use dashes to indicate falters or incomplete thoughts, rambling speech, or unfinished sentences. Do not use ellipses.

Simultaneous Speech
Include simultaneous speech. Do not finish sentences in the transcript that were not finished during the interview. If each speaker's statement is indecipherable, use [both speaking-unclear].

Indecipherable words
Use a question mark to express uncertainty in the text. When you cannot understand a word or phrase and cannot venture a guess, use [inaudible].

Examples: "My best friend in high school was Bella Johnson [?]. If you're unsure of a phrase, put the entire phrase in brackets, followed by a question mark: "Like I said [it sounded fine to me?]

Interruptions and Off-topic Remarks

Off-topic/Extraneous Remarks by the Interviewer
Encouraging remarks by the interviewer, such as "yes," "sure," and "I see," can be left in if it is used as a direct response to a point made by the interviewee. If they occur frequently and become disruptive, evaluate them carefully. They may be left out if doing so does not affect the course of the interview.

Non-Verbal Sounds

Include and note with square brackets [ ]. Do not capitalize. If non-verbal sounds occur at the end of a sentence, place the word in brackets after the final punctuation.

Examples: [chuckles], [chuckling], [laughs], [laughter].

Interruptions

Interruptions that affect the recording (telephone ringing, clock chiming, etc.) should be explained using square brackets [ ]. If the recording is paused, indicate that in brackets.

Examples: "We were driving down to the church [phone rings] Oh, let me answer that. [recording paused]"

Style Considerations

Abbreviations

In general, avoid abbreviations.

Acronyms

Always provide the full name of an acronym if known. Use square brackets to provide the full title of the name or organization Example: "I started out with the SCLC [Southern Christian Leadership Conference].

Capitalization

Follow the proper forms of standard English in running text.

Administrative titles

Titles are capitalized only when they are combined with a name and refer to a specific person. They are not capitalized when referring to a general title. Example: "I talked to Chancellor [Dr. Lucinda] Planchett about the vice president's search committee."

Numbers

Use numerals as long as the numbers do not begin the sentence. If a year is the first word in a sentence, it must be spelled out. Example: "I moved to Greensboro in 1937 or '38."   "Sixty was the year of the sit-ins."

Times

Spell out any time that is mentioned as part of the recorded content. Do not use numerical times in the HH:MM:SS, HH:MM, or MM:SS format as that format is used for timestamping. Example: "I had to set dress sets for productions at 2 o'clock and 5 o'clock every Saturday." "She met me at the movies at four thirty." "I ate lunch with the teachers every afternoon at 1pm for over a month that year." "I remember that year I ran a mile in 6 minutes and 33 seconds."

Creating Oral History Transcripts

Manual Transcription

Using the above guidelines, you can create a serviceable oral history transcript manually by listening to the audio recording, pausing it, and transcribing it into a word processing document. This process is labor intensive and typically requires 4-6 hours of transcription time per hour of recording. There are various tools that can be used to facilitate this process, such as audio editing software (Audacity, Adobe Audition, GarageBand, etc.) and transcription software and foot pedals (ExpressScribe transcription bundles come with a foot pedal, or they can be purchased separately).

Automated Transcription

You may also choose to use transcription software or an automated transcription service. Some popular transcription software includes the Dragon line of products, Otter, and even using Office365 Word's transcribe function (online only). Some of the popular online transcription services available now include Rev, Trint, Temi, GoTranscript, and others. When using an automated service, there are typically two options: a human transcription or an AI generated transcription. Regardless of the source, you will have to proofread carefully afterwards. it can be a significant timesaver to only have to proofread a transcript as opposed to generating the transcript from scratch oneself.

Regardless of the method chosen to generate the transcript, while not required, it will be extremely beneficial to your listeners/viewers to be sure to include timestamping. We will discuss how to format your timestamps in the next section.

Formatting Oral History Transcripts

File Format

Oral history transcripts should be provided to the Lowcountry Digital Library as either a Word .doc/.docx or as a .txt file.

Headers

Headers should be included in all oral history transcripts. Header information should include:

- Interviewee - "Eloise Jane Smythe"
- Interviewer - "Jacob Hamberg"
- Date of Interview - "May 14, 1997"
- Location of Interview - "17 Church Street, Charleston, SC"
- Length of Interview - "52 minutes and 14 seconds"
- Project Name - "Lowcountry Memories: Oral Histories from Coastal Beaufort County"

Additional information can be included, such as: original format of recording (if analog and later converted to digital), editor and editing date, transcriptionist and transcription date, proofreader and proofreading dates, etc. Header information is automatically stripped out of the transcript display, but is useful for metadata, tracking, and administrative purposes.

Timestamps

Timestamps (or timecodes) are optional but highly recommended. Properly formatted timestamps included in transcripts will automatically sync to audio tracks within LCDL's a/v platform. Timestamps can be formatted in any of the following ways:

- [HH:MM:SS]
- [MM:SS]
- HH:MM:SS
- MM:SS
- HH:MM:SS:MS
- HH:MM:SS.MS

Timestamps should be the first entry on a line, followed by the speaker name, and then a hard return. The transcript text should appear below the timestamp/speaker line. Please see the example provided below.

Transcript Example

For your ease of use, we've created a [sample oral history transcript template](#). You are welcome to download a copy and use this as a template for your own oral history transcription project or you may simply use it as a reference. It is important to maintain consistency within your oral history project, so we recommend that you establish your project template and use the same format throughout your project.

Headers and Footers

Headers and footers are optional but acceptable in your oral history transcript file.

Indexing Your Oral Histories (Beta)

An index is time-coded metadata consisting of names, subjects, keywords, partial transcripts, locations, hyperlinks, etc. that is attached to an audiovisual media file. Using a timecode, an index references and provides a link to marked places in the oral history media file. This is similar to the way an index functions at the end of a book, which has similar references to the pages in the book where such metadata can be found on. Rather than page number references, though, indexing in LCDL's platform includes timecode references that a user can click on to skip to that section of the recording.

In our new platform we will be able to manually add index information to your media recordings. To take advantage of this feature, you will need to provide the appropriate metadata and the timecodes of the sections that are applicable. We strongly recommend that indexing be used sparingly, to draw attention to specific topics of interest that might otherwise be missed. The following information must be supplied for each indexing request:

- **Timestamp:** This is the exact time when the topic to be indexed is first discussed.
- **Segment Title:** This is the title of the index topic. This will be viewable by researchers so it is important to keep it informative and understandable.
- **Segment Synopsis:** This is a brief narrative description of the topic to be indexed.

Below is an example created using the Spoleto: Oral history with Carolyn Kostopoulos recording
- **Timestamp:** 00:46:48
- **Segment Title:** Ghost Stories from the Old City Jail
- **Segment Synopsis:** Carolyn Kostopoulos discusses her encounter with the supernatural while working for Spoleto in the Old City Jail.

This feature is still in beta, meaning it is still under development. Because of this, we will be limiting the use of indexing to no more than 3 index topics per resource. Please remember:
- Indexes are completely optional.
- They are not meant to be the primary descriptor of a resource.
- They should serve only as a supplement to the existing metadata.

If you are interested in providing index information for a resource, please consult with the LCDL project coordinator, and they can provide you with an index template.