

Transcription Principles

Institute for Editing Historical Documents - 2019

Presentation by: Cathy Moran Hajo

Transcription

Transcription is the backbone of the documentary edition. It makes difficult primary sources legible to the general reader, enables us to publish these sources and make them widely available.

In a digital edition it becomes even more important, because it also serves as the basis for searches and interpretive tools such as text analysis.

Methods of Transcription

- Typing into a Word processing program
 - Depending on your rules, may have to turn off autocorrect, autoformatting.
- Optical Character Recognition
 - Depending on your typed texts, be prepared to do a LOT of correcting.
 - Handwritten texts?
- Use a dictating machine
 - You will need to train it, *and yourself*, to talk to the machine effectively
- Hire typists in [insert Third World nation] and do automatic comparisons to catch errors.
 - Ethics aside...wait! Don't put your ethics aside!

Transcription

We need to rethink traditional transcription procedures in light of digital publication.

- Editors developed ~~weird~~ ↑unique↓ ways [of] publishing complex primary sources in print format to try to represent the original.
- Should we continue this when our digital edition might also include an image of the original?
- What other ways does the digital environment change the goals of transcription?

Transcription

How faithful should we be to the original text?

That depends....

- on our audience
- edition design
- search plans
- display plans.

Transcription - Print Editions

Print editions are designed for scholars. They are available pretty much only in college and specialty libraries, and they often make little concession for the average reader.

- Some editions do not translate foreign documents (based on assumptions such as “a good scholar should be able to read 18th century French!”)
- Many editions render archaic spellings and words as they appear with no explanation.

Transcription - Digital Editions

Digital editions will have broader use. Scholars, graduate and undergraduate students will still use them; but we hope that high school and grammar school students and their teachers will use them, the general public, and those overseas.

We struggle with the two pulls of making the documents accessible and making them accurate.

Transcriptions

Can we just say--right now-- that NOBODY has a transcription policy that seeks to render a facsimile of the original.

- Most editors regularize the appearance of parts of the document (standard locations for date, place, salutation, and signature; regularized paragraph formatting, addition of missing terminal punctuation; regularizing capitalization at the start of sentences. etc. etc.)

Don't think that you are selling out to aim for accessibility.

Transcriptions

Some issues to consider:

- Archaic words - modernize or retain?
- Interlineations and marginal notes - identify them in some way, or insert them silently as the author intended?
- Abbreviations - expand them or retain?
- Punctuation - supply if it is missing, or leave as is?
- How to indicate uncertainty, when you are not quite sure what the document says.

Transcription Questionnaire

Hobson Woodward, Editor of the Adams Papers developed 32 questions to start working towards a transcription policy.



Other Considerations

What is the role of the image? Are you going to include an image as well as a transcription? Does that change some of the decisions you made earlier?

- A scholar who wants to see exactly how the document was created, where notes and corrections are, can look at the image.
- General users can get a better sense of the historical item by looking at the image.
- But they are not searchable--at least not yet

Other Considerations

Transcriptions and Searching--

- Computers are stupid. If the string of text they are searching for does not match your search term exactly, it may not find a match.
- People are smart, but they sometimes can't spell or they use abbreviations in place of the real word.
- Editors use [strange punctuation] that can interfere with search results.
- Text searches don't tell the whole story.

Transcriptions

So

“if you wd. like to meet Bob and the old guy in N.O. in Jul.
come straight away”

is a nightmare for searching, though it is pretty clear when
a human reads it.

Decisions, Decisions!

- 1) You could expand the abbreviations, silently or not, so that the complete word is there.
- 2) You could correct the misspelling, silently or not, so that the correct word is there.
- 3) You could encode the expansions and corrections of words in the transcript, using XML, so that they would be used for searching, but not for display.
- 4) You could do number 3, but have either the corrections or the original appear when the user hovers over a word.
- 5) You could do number 3, but allow the reader to decide to see either the expanded or original text.

Transcriptions

What you *need* to do is take the time to think about it.

Develop a transcription policy that takes into account how you will be using the transcriptions, how you want them to appear on the screen, and how you want them to be used in searching.

There is no (gulp) right answer!

Making a Transcription Policy

Gather a good sampling of documents.

- Ones with issues: corrections, columns and charts, misspellings, abbreviations and oblique references
- Different formats: letters, diaries, postcards, typed letters on letterhead, etc.
- Transcribe them!
- Start listing the times that you need to make a decision and what you decided.

Making a Transcription Policy

What does the process show you?

- Are you more comfortable with some regularizations and less so with others?
- Do you think some things can be regularized without comment, while others should be marked in some way?
- Is the transcription you produced suitable for the broadest audience?
- Would it turn off a scholar if you intervened too much?
- Could you see producing two, or more, different transcriptions?
- Do you want to think about using XML?

Document Your Transcription Policy

What good is a policy if

- No one knows what it is
- No one uses it
- It doesn't address the issues that you face?

Documenting your policy and treating it as a living rule book is as important as developing the rules in the first place. Make your rules accessible to all.

Crowdsourcing Transcriptions

Crowdsourcing is a popular and fundable way to outsource some of your project's time-consuming tasks to online volunteers. Some major projects using crowdsourced transcriptions are:

- [Jeremy Bentham Papers - Transcribe Bentham](#)
- [Papers of the War Department - Transcription Associates](#)
- [National Archives - Citizen Archivist](#)
- [Letters of 1916](#)
- [Old Weather](#)

The National Archives, Smithsonian, New York Public Library and other museums also have crowdsourcing programs in place.

Crowdsourcing Transcriptions

Crowdsourcing transcriptions:

- probably won't save you money.
- might save you time.
- requires a good, clear transcription policy.
- benefits from having an interface that makes it hard for the transcriber to screw up.
- should be a fun experience for the transcriber.
- need to be proofread!

Proofreading

ALL transcriptions need to be proofread!

- Looking for typing mistakes
- Looking for formatting errors
- Making sure all transcription rules were applied consistently

Proofreading Methods

- Proofreading by reading aloud
 - Teams - one reading aloud, the other following the original. Stop whenever there is a discrepancy.
 - Solo - use your computer to record yourself reading the document aloud, then play it back and compare it. Some say you should read it backwards so that you do not make the same mistakes you might have when you transcribed it.
- Comparing by eye
 - Harder to do well, but if it is all you have...
- Comparing computer files
 - If you type a document twice and then compare the files you can find discrepancies and compare by hand.

Supported by:

Association for Documentary Editing

&

National Historical Publications &
Records Commission



NATIONAL
ARCHIVES

NATIONAL HISTORICAL
PUBLICATIONS
& RECORDS COMMISSION