

KEY to Word Enhanced for TEI

Last Modified: August 2026

The purpose of this KEY is to provide further explanation and clarification regarding the markers {{ }} you will include in your MS Word transcriptions. All markers {{ }} included in each Word file will be converted to corresponding, valid TEI markup when you process the Word file with the online WET VAC system. Find online at <https://www.primarysourcecoop.org/tools/wetvac/> . Works best in Chrome.

*N.B. Use of **boldface** in this document is just to make it easier to find things here. Do NOT use boldfaced text in your Word document.*

Some basics and terms to keep in mind

Marker = anything in curly braces {{ }}.

These generally correspond to elements that will appear in the XML file, looking like this: <thing>.

All markers are either block-level or inline. “**Block-level**” means these function like (or are) paragraphs. Each should begin on a new line created by a hard return (Enter key). There should be no leading whitespace on the new line. There cannot be hard returns *within* the text associated with a block-level marker. “**Inline**” markers should be used within a paragraph or other block-level marker, between words or marking words interspersed among other text. Do not enter a hard return immediately before one of these markers.

XMLID is TEI speak for the unique identifier for an XML document. This does not have to be the filename for the Word document, but it will become the filename for your XML file.

Consequently, each Word document in your project must designate a unique ID for your document: rules for constructing those and adding them to the Word document appear below.

The name of your Word file can be whatever you wish, though it is best to avoid too many unusual characters.

----- How Do I Handle -----

[Document Body Markers—i.e., Historical Text “markup”](#)

[Document Back Markers—i.e., Annotations](#)

[Managing Autocorrect in WET Files](#)

Metadata Markers

The metadata markers are the first part of every document, preceding the historical text. There should be no paragraph breaks (hard return, ENTER key) or line breaks (soft return, SHIFT+ENTER keys) within the text that follows a marker. Preferably all markers should remain in the order shown here; irrelevant markers can be left out, but do not leave out a marker because the information is unknown. (n.b. really not sure what I meant by that last bit ...)

- {{XMLID}}** Supply the unique ID your project has assigned to this document. It must follow the LLLNNNNN format (3 uppercase letters, 5 numerals). E.g., OEL00521
- {{DATE}}** Historical date as best known for composition of document is **required**. It must be in ISO (international standard) format for the output XML to be valid. This standard requires numerals and hyphens only with YYYY-MM-DD structure, thus: 1776-07-04
- If only the year or the year and month are known, ISO will still validate with only the relevant parts, such as 1776-07 or just 1776.
- Date ranges and conjectural dates CANNOT be entered in the WET document. They can be corrected in the XML. For more info, see the encoding guide or the FAQ sheet.
- {{AUTHOR}}** Include identification for every known author after this marker, using semicolons to separate identifications if more than one. Use the HUSC (Hyphenated Unique String of Characters, which generally follows a lastname-firstname format) assigned to that individual in the Name DB. Confirm that a corresponding entry already exists in the Names DB or add one if it does not. The Word document will transform to XML without the proper HUSC, but it cannot be published without it.
- Examples: {{AUTHOR}} sedgwick-catharine
- {{AUTHOR}} richards-ellen
- {{RECIPIENT}}** If the document had a recipient, include identification for every known recipient after this marker, using semicolons to separate identifications if more than one. Use HUSC lastname-firstname format for each name, confirming that a corresponding entry already exists in the Names DB.
- Examples: {{RECIPIENT}} richards-ellen
- {{RECIPIENT}} atkinson-edward
- {{HEAD}}** The text here, composed by the editors, will be the text that is visible to users of the online edition as the heading for this document. For correspondence, the best practice is Firstname Lastname to Firstname Lastname.
- For example: {{HEAD}} Ellen Swallow Richards to Edward Atkinson

- {{EDITOR}}** Name of each member of the main editorial team for the edition. Be consistent.
- {{EDITION}}** Name of the edition, in full. Be consistent.
- {{TRANSCRIBER}}** Name of initial transcriber (if known). Be consistent.
- {{TRANSCRIPTION-DATE}}** Date of *initial* transcription—date completed if more than one day. See **{{DATE}}**, above, for acceptable date formats.
- {{MILESTONES}}** This optional marker makes it possible to set any of the four major editorial milestones (transcription, persRefs, subjects, annotation) tracked in the metadata as *completed* in the WETVAC output.
To generate a “complete” status for each that is complete at the time of XML transformation, add the name/s after the marker, using the exact terms used above. If more than one milestone is indicated, use a single space between the terms. Capitalization does not matter.
For all WET documents, the default WETVAC output *without* the Milestone marker will treat only the initial transcription as done.
- {{SUBJECT}}** Optional, can be none or many. Use each one to identify a historical topic that content in this document invokes. BE SURE to enter Subject terms exactly as spelled (including capital letters and punctuation) in the Subjects DB.

Document Body Markers—i.e., Historical Text “markup”

The following markers and formatting options should be used to indicate visible, semantic features of the **historical text** in the source document. Each will determine XML tagging in the XML file generated by WETVAC. **Note that *most* Word formatting will *not* produce any XML tagging (and some may cause WETVAC to reject the Word file).**

First, a useful tool for noting issues the team needs to follow or address later: Comments!

{{COMMENT: text of your note for edition team members here}}

Write anything you like after the colon within the curly brackets. It will be preserved as an XML comment. As long as the comment appears after the Dateline, it will appear in XML produced by WETVAC.

HISTORICAL DOCUMENT TEXT: OPENING AND MAIN BODY OF TEXT

---- *following items are all “**BLOCK-LEVEL**”: each marks a section of text that begins on a new line, like a paragraph. All markers below that are present in a document must be entered in the order in which they appear here. The absence of irrelevant pieces is fine.* ----

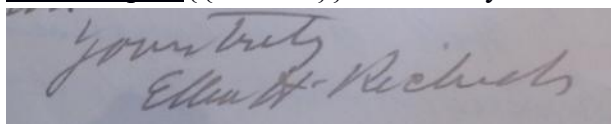
{{DATELINE}} Transcribe the dateline, if it stands apart from paragraph text, as written by the author, including any information (e.g., place, time of day, etc.) written in the same line. Indicate line breaks in manuscript dateline with / slash.
For example: **{{DATELINE}}** Sunday eve, Nov 22 1894 / Boston

{{SALUTE}} If no salutation, do not use.
For example: **{{SALUTE}}** My dear Mr Atkinson

Paragraphs ... Hey, how do we make **paragraphs** of historical text?
Just as in any Word document, hitting the Enter key will create a new paragraph.
In the XML generated by WETVAC, each hard return will become a <p> element, the standard element for a paragraph across most XML and HTML schemas.

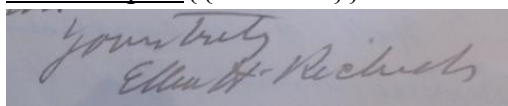
{{CLOSE}} This marker encompasses any closing sentiment expressed at the end of a letter on a line above the signature (see **{{SIGNED}}** for that) or other standard closing language that may appear in different kinds of documents.
Use / slash for line breaks within closer. If no such text exists, do not use.

For example: **{{CLOSE}}** Yours truly



{{SIGNED}} Use for the author's signature, **including** other text on the same line.
If there is no line that includes the author's signature, do not use.

For example: **{{SIGNED}}** Ellen H. Richards



{{PS}} AKA Postscript: Use for any brief addendum that follows the author's closing words and signature; it may be by another writer but should be apparently contemporaneous with the composition of the document.
Remove if not relevant.

{{INSERTION}} Use to transcribe inserted text by the author or another contemporary writer, corresponding with the inline **{{INS}}** marker, described above. If more than one, repeat. If **{{INSERTION}}** appears in a transcription, it must be the last marker before the source note, i.e. **{{SOURCE}}**

---- following items are all **"INLINE"**: used within the running text of the transcription ----

Using MS Word formatting features to markup condition of historical text

superscr^{pt} text Use superscript formatting in Word to indicate where the author has handwritten superscript characters in the document.

~~word/s deleted~~ Use strikethrough formatting in Word to indicate text that has been deliberately canceled in the manuscript but still appears as text on the

page; provide a transcription of any text that is still legible.

For text that has been canceled and is illegible, use this combination:

{{ILL}}

underlining

Use underline (U) in Word to indicate text underlined in manuscript.

italics

Use primarily in notes for book titles or for transcriptions of printed materials in which some words were italicized.

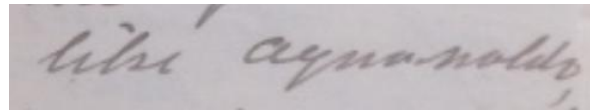
Using markers and special characters to markup condition of historical text

{{DAMAGE}}

Indicates any type of damage to the document—like a tear or where a signature has been cut—that has made text unrecoverable (damage to non-text part of document is ignored).

[conjectural]

Enter text of conjectural reading between square brackets to indicate word or words that are unclear to some degree but for which the editor is able to make a determination with relative certainty. (For illegible, see below.)



like [Aquanaldo]

[conjectural?]

As above, but the question mark indicates a much lower level of certainty.

Abbreviations

abbr==expansion

where == is two equal signs. Use to provide an expansion to an abbreviation on the page if that abbreviation is apt to be confusing to most readers. **It is best to use this markup sparingly**—only where manuscript text is very ambiguous or confusing. Does J’y stand for July or January? Spaces and most punctuation will not process correctly, and there must be **no** space before or after the equal signs. To indicate necessary spaces in either term, use underscores. E.g.: J_y==January where _ replaces apostrophe or period. Add a {{COMMENT}} in the Word file to flag that these will need adjusting in the XML after processing through WETVAC.

Regularized spelling

original++modernized

Use to provide an editor-supplied modern spelling when the author’s spelling would be confusing for readers. **It is best to use this markup sparingly**—only where manuscript text is very ambiguous or confusing. Spaces and most punctuation will not process correctly, and there must be **no** space before or after the plus signs. If spaces are needed in the expansion, use underscores. E.g.: “peas full” appears in the manuscript; mark it as peas_full++peaceful. Add a {{COMMENT}} in the Word file to flag that these will need adjusting in the XML after processing through WETVAC.

- Illegible** `{{ILL}}` Use for written text visible on the page but impossible to decipher. Refinement (extent of illegibility, reason for illegibility) can be added in XML if desired. For text that is now illegible and was canceled, use this combination: `{{ILL}}`
- Page break** `{{PB}}` Indicates a page break. Numbering will be added automatically in XML. If the page break is in between paragraphs, place the marker at the end of the paragraph before the break. Do not place on its own line.
- `{{BLANK}}` Used to represent any apparently intentional author-created space in text. This is an inline marker—that is, within a paragraph or other block-level marker—to indicate an empty space *in lieu of* a word or words. Do not use for space between block-level markers.
- ^added text^** Use for text added to the manuscript by the author or another contemporary hand. Note there must be a space before the starting caret and after the closing caret. **INLINE but can also be used for an added paragraph.**

Using markers to annotate historical text

- `{{P: text or notes}}` **AKA persRef:** Use to mark a person’s name or another word or words that refer to a person in the historical text. The last word in the reference *must* be followed with `{{ENDP}}`.
Example: I gave it to `{{P: smith-john}}`my brother`{{ENDP}}` today.
Always use the HUSC lastname-firstname format for each name. If a corresponding entry already exists in the Names DB, use that.
- `{{INS}}` Can be used to note insertions, such as marginalia, that are too large to run into the text and/or in a position that is ambiguous with regard to the rest of the text. Transcribed text of the insertion should be placed at the bottom of the document using `{{INSERTION}}`, which is block-level. `{{INS}}` itself is **-INLINE**
- `{{N}}` Note pointers: pointers within the historical text that link to an editorial annotation in the “docback” part of the XML (see below). For each `{{N}}` to process in WETVAC, there must be the same number of `{{NOTE}}` markers in the docback. Numbering will be automatic.
Do not use this to identify people referenced in the (see `{{P}}` above).

Document Back Markers—i.e., Annotations

Most docback markers are block-level, beginning on a new line created by a hard return. `{{SOURCE}}` is **required** for every transcription and must be the first of the back matter markers. Most are block-level but inline formatting and some markers can appear in notes.

{{SOURCE}}	Source note: introduces the section of the annotation that records key information about the physical document used for this transcription. NO TEXT DIRECTLY FOLLOWS THIS MARKER. Information should be divided among the following seven markers, as relevant. Use only those that are relevant to your document; all used should remain in the order presented here.
{{DOCTYPE}}	Document type: record standard identification for letters (Recipient Copy, Letterbook Copy, Draft) or indicate other type of document (e.g., Diary). Any frequently used term should be consistent language designated by the edition.
{{REPOSITORY}}	Name of repository, in full and as used by the holder; indicate private collection by name of owner or “Private collection” as determined during permissions process.
{{COLLECTION}}	Name of collection, in full and as used by the holder.
{{CONDITION}}	Any salient details regarding the physical condition of the document, as determined by editors.
{{ADDRESS}}	Transcribe recipient address from envelope or verso of the letter sheet (not unusual previous to 19th century), per your edition’s guidelines. Use / slash for line breaks within the address. If there is no address, do not use this marker.
{{ENDORSEMENT}}	This is used for any endorsement by the recipient or the recipient’s affiliate (e.g., a secretary) for filing purposes, done at or close to the time the document was received. Transcribe per the edition’s guidelines. If there is no endorsement, do not use this marker. If more than one, repeat.
{{NOTATION}}	This indicates any additional contemporary (or near contemporary) notation on the letter, such as “pr. Capt. Bigelow” or “Not sent.” When possible, indicate who wrote the notation if known. If there is no notation, do not use this marker. If more than one, repeat. NB: this is not the place to include any editorial observations about the condition of the document; that should be entered with the {{CONDITION}} marker.
{{NOTE}}	Each editorial annotation for a document must begin with a {{NOTE}} marker. These must follow the final marker affiliated with the source note. Numbering will be automatic. The number of {{NOTE}} markers must match the number of {{N}} markers in the body of the historical text.

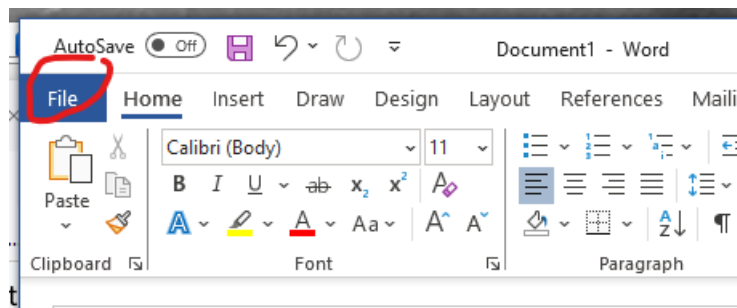
More Information on All These Things

Creating “markup” for a transcription done in MS Word is ultimately quite limited. Much more precision and nuance can be achieved in the TEI-XML encoded document. See the Encoding Guide on the Coop’s Resources page for information on how the Coop uses XML, but keep in mind that as your project develops, our schema may not contain every element you need.

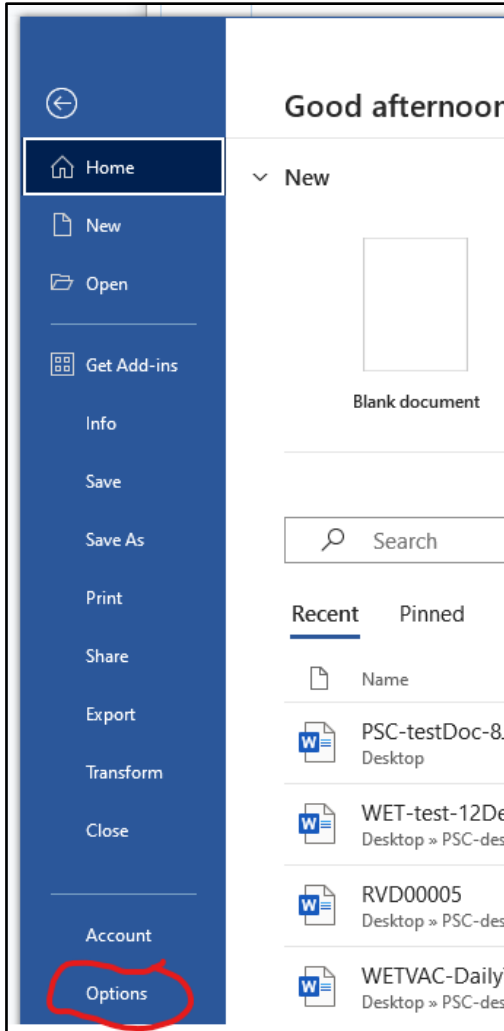
Managing Autocorrect in WET Files

Transcribing historical text in MS Word can be especially difficult when Word’s autocorrect feature is in full force. It may correct or at least question many of the irregular spellings in the historical text. It may automatically capitalize all initial letters for a new sentence.

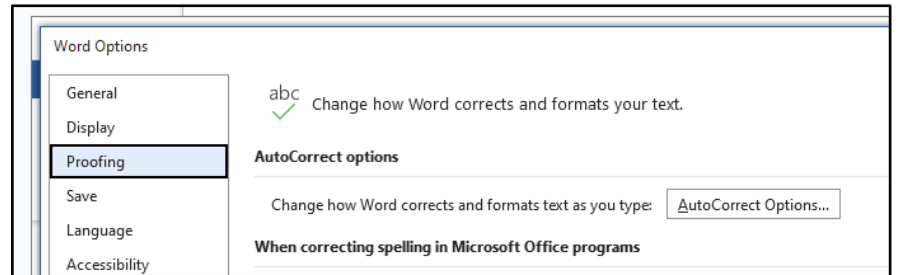
Autocorrect has especially caused problems with Coop markup by automatically capitalizing the lowercase characters of HUSCs within the `{{P}}` tag for person references: goldman-emma may become Goldman-emma. Because the XML in this circumstance is case sensitive, a HUSC value with an uppercase character will cause a validation error for the XML file.



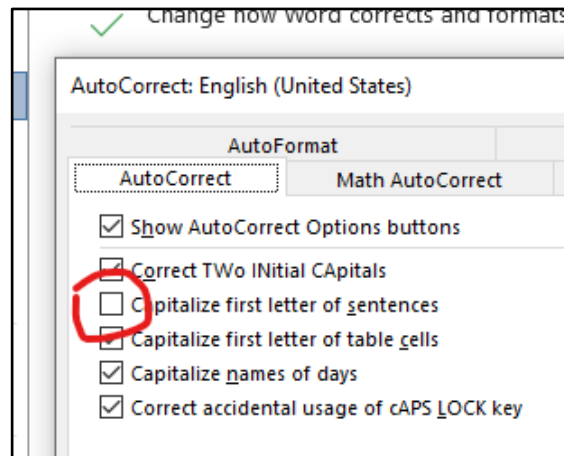
To fix this problem, the easiest solution is to change the autocorrect settings in Word. First, navigate to the **File > Options** menu.



The **Options** button will be all the way at the bottom of the screen. This will open the **Word Option** window. Navigate to the **Proofing** option* on the left side bar. To change the AutoCorrect options, click on the “**AutoCorrect Options**” button ...

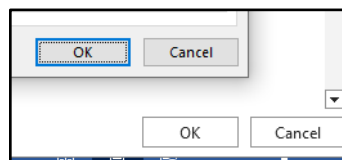


... which will open this dialog box:



To change the settings that autocorrect the capitalization in your WET documents, including the first letter of a HUSC, **unclick** the following options for first letters of sentences, names of days, and accidental usage of CAPS LOCK key.

Check all the OK buttons:



* N.B. There are all kinds of settings in the “When correcting spelling in Microsoft Office programs” area that you can un-check to remove those intrusions in your transcriptions.