EMMO conventions for three versions of transcriptions on the EMMO beta site

See EMMO tag set and downloadable XML files for more information on the actual codes

Diplomatic Transcriptions

The goal of diplomatic transcriptions in the EMMO database is to allow our users to visualize how the text is rendered on the original page, as closely as this is typographically possible. However, we do not attempt to mirror the layout and spacing on the page exactly, nor are we able to reproduce every brevigraph or symbol accurately.

The following conventions are observed for diplomatic transcriptions:

- Original spelling (including u/v and i/j and ff), punctuation, and lineation are maintained (terminal long "i" characters are transcribed as "i" not "j")
- Hyphens (including "=" characters) are preserved at the ends of lines
- Superscript letters remain superscript, interlineal insertions are tagged differently but are displayed as superscript
- Abbreviations are not expanded
- Deleted characters and words are tagged as such and then displayed with a line through, e.g., "~~whitch~~"
- Special-p characters are rendered as a symbol
- Supplied text (used very rarely) and gaps in the transcription are rendered as a symbol
- Symbols for –er/ur and -es brevigraphs are displayed in braces: {er}, {ur}, and {es}
- "y" characters that act as thorns (archaic letters) appear as a "y", i.e., not as a symbol but these thorns will be tagged as such in the XML
- Macrons and/or lines over sections of words and colons (and in some cases periods), signaling an abbreviation, appear in the display – macrons and lines signaling abbreviations are tagged; those colons and (in some cases periods) that signal abbreviations are tagged as metamarks
- Endorsements that were added later, i.e., administrative endorsements, are tagged differently than the original endorsements – these will appear in light gray text
- Characters or words obviously entered much later, e.g., bookseller marks, library call numbers, etc. are tagged and included in the XML but are not displayed

Semi-diplomatic Transcriptions

The goal of semi-diplomatic transcriptions in EMMO is to provide an intermediate version of the manuscript text, very similar to the diplomatic transcription (following the conventions above) but with a few minor alterations for readability. This version is the one taught in paleography classes at the Folger and is often used for scholarly transcriptions/editions.

The following additional conventions are observed for semi-diplomatic transcriptions:

- Superscript letters are silently lowered (though recorded in the XML with tags), <u>except</u> for the following superscript cases: characters denoting currency (e.g., ^{li}), Roman numerals forms that signal operations (e.g., iiii^{xx} meaning four-score or 80), ordinal numbers for dates and the like (e.g., 2° meaning second), and interlineal insertions (tagged distinctly in the XML)
- Abbreviations are expanded, with expanded letters displayed in italics – macrons/lines over words and colons (or in some cases periods) that denote abbreviations are not displayed, though the tags for these will be noted in the XML

- Archaic brevigraphs such as the thorn (entered as "y" character but tagged as thorn) and the -es and -er/-ur graphs are replaced with modern equivalent letters displayed in italics
- Supplied text (very rarely used) is displayed in brackets.

Regularized Transcriptions

The goal of regularized transcriptions in the EMMO database is to provide searchable and easy-to-read texts, rather than scholarly transcriptions.

The following conventions are observed for regularized transcriptions (based on the diplomatic and semi-diplomatic versions):

- u/v and i/j and ff/F are regularized
- Spelling is modernized (following standard American spelling forms) – this includes deleted words
- Expanded, brevigraph, supplied, and silently-lowered text (see semi-diplomatic section) is displayed as regular text (i.e., no italics)
- Separate words that are regularly written today as one word are modernized, e.g., "my self" would become "myself" and similarly, combined/shortened words are modernized, e.g., "thother" would become "the other"
- Words broken up over two lines with a hyphen (or "=" character) in the diplomatic and semi-diplomatic versions are converted to a modern one-word version, where applicable, and appear on one line, e.g., "possi=" at the end of one line and "ble" at the start of the next would become "possible" at the end of the first line for the regularized version
- Obscure and/or obsolete words are retained if listed in the Oxford English Dictionary (OED), e.g., "resiant", also older verb forms are preserved, e.g., "bringeth"
- The use of majuscule and minuscule forms is regularized for names – proper names of people and places and the like (e.g., days of the week) are converted to majuscule, but otherwise capitalization from the manuscript is maintained, e.g.,"George" instead of "gorge" as written
- Text of interlineal insertions (rendered in superscript for diplomatic and semi-diplomatic versions) is lowered as regular text in the display
- Proper names of people and places are converted to modern spelling versions for easier identification, e.g., "Cheadle" instead of "Chedyll" as written in the manuscript
- Roman numerals are converted to Arabic numbers, e.g., "xx" becomes "20" – superscript ordinals for dates are converted to English, e.g., "1°" becomes "1st".
- Apostrophes are added as needed for possessive case
- Latin names of months, place names, and personal names are rendered in English, e.g., "Oxon*iensis*" in semi-diplomatic would become "Oxford" in regularized