

# How-to #0: Rules for transcription

Every transcription project has its own set of editorial principles, guidelines and rules. The MoEML approach is very straightforward. The initial transcription is a *diplomatic* transcription; in other words, we try to reproduce every feature of the original text as precisely as possible.

**Capitalization** should be reproduced. However, look out for instances of *small-caps*; small-caps are not really capitals, they are just a decorative typographical feature, which will be encoded in CSS later. So, for instance:

|                              |               |
|------------------------------|---------------|
| You see this:                | GOOD Cloathes |
| You transcribe it like this: | GOOD Cloathes |
| You see this:                | CHAP. 2       |
| You transcribe it like this  | Chap. 2       |

## **Spacing**

MoEML closes up extra spaces between words and punctuation marks. However, we retain the spacing in authorial initials, such as A. M. (for Anthony Munday). We add a single space after a comma when the comma has been used to separate two words.

**Unusual characters** should be reproduced. A range of archaic and unusual glyphs may appear in your texts. All of them should be available in Unicode, so you should be able to reproduce them. Examples are:

|                  |  |
|------------------|--|
| Long s:          | f  |
| Quotation marks: | “ ” (curly double), ‘ ’ (curly single), " (straight double), ' (straight single) |
| Nasal Tildes:    | õ  |

**Ligatures:** æ, œ, ff, fi, fl, ffi, ffl, ft, st

We retain the vowel digraphs using the appropriate Unicode characters (e.g., æ), but typographical ligatures (e.g., fl) should be silently expanded.

**Line breaks** should be reproduced. You will need to encode each line break later.

**Hyphenation** should be preserved, both within and at the end of lines.

**Original spelling** should be reproduced. Sometimes we don't even notice that a word is spelled in an archaic or unusual way, and transcribe it in the modern form. Watch out for this.

## **Interchangeable Characters**

We retain the interchangeable u/v and i/j and the use of vv for w.

**Typographical decorations**, including bold, italic, and underline, should *not* be reproduced in your transcription. We encode these features with CSS code later on, not in the actual transcription. For example:

You see this:                      So that you ftand not in the *Common Rancke* of *Dry-fi*~~St~~*ed*  
*Patrons*

You transcribe it like this:      So that you ftand not in the Common Rancke of Dry-fisted Patrons