

How to type character variants

1. Character variants in Unicode

If a character variant is included in Unicode 5.1, type it.

The remaining sections apply only to character variants which are not included in Unicode 5.1.

2. Character variants not in Unicode

Some character variants should be silently normalized, some should be marked, and some should be treated as unknown characters. The rules in section 3 specify which case applies to a given character variant.

Begin a new Marked Character Variants List and a new Unknown Characters List for each text. (You can also add all marked character variants and all unknown characters to a single list.)

If a character variant should be marked, add it to the Marked Character Variants List, i.e. add the image and how you have marked it. In addition, mark the first occurrence in the text. After its first occurrence, silently normalize it.

Unknown characters should be typed with their unknown character code (e.g. <001>) whenever they occur in the text.

3. When to normalize, mark, or treat as unknown character

Rule 1 (normalizing simple changes): If a character variant differs from the standard character only in the following points, type it as the standard character and do not mark the difference.

a) Differences in rotated strokes/dots

Examples: 半·半, 勺·勺, 羽·羽·羽, 酋·酋

b) Differences in overshoot at the stroke initiation and/or termination

Examples: 身·身, 雪·雪, 拐·拐, 不·不

c) Differences in contact of strokes

Examples: 奧·奧, 酉·酉, 兕·兕, 查·查

d) Differences in protrusion at the folded corner of strokes

Example: 巨·巨

e) Differences in bent strokes

Example: 西·西

f) Differences in folding back at the stroke termination

Example: 朱·朱

g) Differences in accent at the stroke initiation

Examples: 父·父, 丈·丈, 夂·夂

h) Differences in "rooftop" modification

Examples: 八·八, 宀·宀

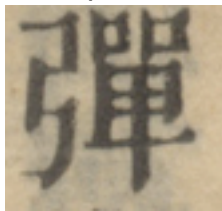
j) Straightforward combinations of the above differences

Example: 刃·刃·刃

Rule 2 (marking old radicals): If a character such as 絨 (U+7D68) exists in Unicode 5.1 only with the new shape 糸 (U+7CF9) of the radical and not with the old shape 糸 (U+7CF8), mark the character variant with the old radical with < R>, for example <絨R>. Apply this rule only if the difference between the old and the new radical is not covered by rule 1.

Rule 3 (marking other variations): If you can identify the standard character of a character variant and the difference is not covered by rule 1 or rule 2, mark the character variant with < V>.

Example: <彈V> for



Rule 4 (unknown characters): If you cannot identify the standard character of a character variant, treat it as unknown character.