Proposal for the inclusion of the DELETE SIGN for proofreading & discussion of the intended use and behavior of already encoded signs
Eduardo Marín Silva
04/Oct/21

Introduction. This is a proposal to include the proofreading mark that indicates a character, word or section to be deleted. The mark was formally included in ISO 5776 (1983) which in turn derives from BS 5261 (1976), but they precede those standards in their informal use.

Furthermore I would like to bring attention to the lack of documentation of the intended use of the characters in question (2038 ⸯ CARET, 2041 ⁁ CARET INSERTION POINT, 2050 ⁐ CLOSE UP).

Function, name and glyph. The symbol, like its name implies, indicates deletion of a section of text. The mark can either be in the margins (with the relevant portion with a strike-through line), by itself or enclosed by 2050 ⁐ [see figures 3, 4 and 8], overlayed over a single character [see figures 7 and 8], overlayed over an entire word [see figure 8] or above the line itself (for words and longer portions).

There are some names to consider according to Wikipedia: DELE, DELEATUR, DELETE; of those the last one is the better suited, and the other two can be informative aliases.

The glyph seems to be highly variable; the preferred form would presumably be the one used by ISO and BS, which is recreated below. Do note that other glyphs include a loop with a rising or curled end or a glyph similar to that of the german penny sign.

Code chart entry. The codepoint 2065 is the last open codepoint in the "General Punctuation" block, which is where the other 3 marks are located, so that seems like the best possible place. The cross references are added due to the graphic similarity to other variants. Here is a mock-up of the proposed inclusion.

2065 ⁊ DELETE SIGN
> = dele, deleatur
> • used as a proofreader's mark to indicate portions to be deleted
> • known to have a lot of glyphic variation; the representative glyph correspond to the form preferred by ISO and BS
> → 20B0 ₰ german penny sign
> → 27B0 ➰ curly loop
> → 1F397 🎗 reminder ribbon

On the zero-advance nature of the proofreading marks in Unicode. There's a concerning lack of documentation on the purpose of the proofreading marks (2038 ⸠ CARET, 2041 ⅄ CARET INSERTION POINT, 2050 ◡ CLOSE UP). Particularly, it is not stated whether this marks are zero-advance or not. On actual use in proofreading, they are zero advance, because they are meant to be added in an *ad-hoc* way (after the document is written), so they should not affect the spacing of the text being checked.

      In other contexts, like documents explaining the meaning of the marks, a presentation with width may be preferred, but this can be achieved with some padding with blank-spaces around the mark; the close up mark may enclose other proofreading marks (like the delete mark here), which can only happen if the "close up" itself is zero-advance.

      If they are meant to be mere symbols, that aren't intended for actual proofreading, then it should be stated as so in the spec and the code charts; because currently some fonts treat them as zero-advance and others as spacing, which can create conflict if a user uses white-spaces as padding when it was zero-advanced, and the font changes when it is spacing.

      Currently the general category of the marks is 'Punctuation_other' which would suggest they are spacing, given all other punctuation marks have width. However, characters being zero-advance, need not conflict with this general category, so it can remain the same. If they are meant to be symbols though, then they don't serve the purpose of punctuation, and so their general category should be changed to 'So'.

      In any event, the delete mark, as proposed here, should be treated as a spacing symbol; that's because in actual proofreading, there are many ways to indicate deletion on the line itself. One way is a simple strikethrough, another adds a loop at the end of the strikethrough (which is why a simple loop shape is associated with it), and yet another way adds a specific symbol above the line instead; ISO and BS even require that the strikethrough be delimited by vertical bars at both sides (this may require two additional combining marks, but that's beyond the scope of this document). This proposal covers the cases where a separate symbol is used (either above the line or in the margins), not those ones with overlayed strokes.

On the two caret characters. The two caret characters in proofreading have more complex behavior apart from their zero-advance nature. Firstly, the simple arrowhead caret has two forms that may be used concurrently. One is the current representative glyph (an up arrowhead below the baseline) and another is a down arrowhead above the ascender line. One form is used over another out of convenience; for instance if there is more room above the line than below the line. It is not clear if both forms should be treated as separate characters, because it would complicate the rendering if they use the same codepoint.

      A very similar behavior applies for the "Caret Insertion Point", because it can sometimes be considered a glyphic variant of the former character i.e. the extra stroke then just serves to better indicate the point of insertion by being overlayed on the line itself, rather than merely above or below it. Therefore it also has an inverted variant, used in the same way as the normal caret.

      However the caret insertion point is sometimes used with a very different semantic. In this use the representative glyph form, is used to indicate that the character to be inserted must be in subscript presentation, while the inverted form indicates that the character must be inserted in superscript presentation [see figures 5 and 6]. Therefore, the case for separately encoding an "Inverted Caret Insertion Point" is indeed stronger, since it also has a semantic distinction apart from a practical justification.

On other characters of ambiguous spacing. Some other punctuation characters (203F ‿ UNDERTIE, 2040 ⁀ CHARACTER TIE, 2054 ⁔ INVERTED UNDERTIE) seem to be intended to be spacing variants of other combining characters, however some fonts fail to display them like so, making them zero-advance. Again, this is a problem for legibility (arguably worse than the proofreaders marks) so better documentation regarding their spacing should be likewise added on the 'General Punctuation' (and the analogous block for the combining versions) block section. In addition, this documentation must also mention the distinction with other graphically similar combining characters:

Punctuation	Combining Character(s)
203F ‿ UNDERTIE	035C ͜ COMBINING DOUBLE BREVE BELOW or FE27 ︧ + FE28 ︨
2040 ⁀ CHARACTER TIE	0361 ͡ COMBINING DOUBLE INVERTED BREVE or FE20 ︠ + FE21 ︡
2054 ⁔ INVERTED UNDERTIE	1DFC ᷼ COMBINING DOUBLE INVERTED BREVE BELOW

Similar observations apply to some ancient Greek punctuation marks, since they are all supposed to be zero-advance: 2E0F ⸏ PARAGRAPHOS, 2E10 ⸐ FORKED PARAGRAPHOS, 2E11 ⸑ REVERSED FORKED PARAGRAPHOS and 2E12 ⸒ HYPODIASTOLE.

Figures.
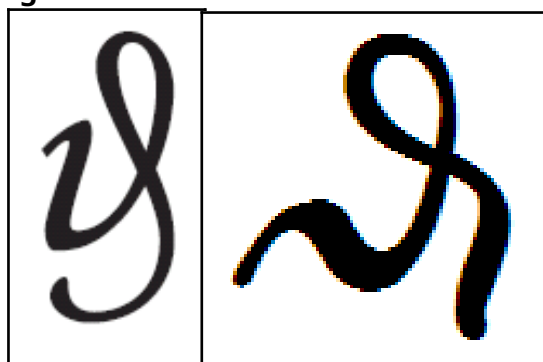
Figures 1 & 2. Representations of the delete sign, very similar to the German Penny sign (20B0 ₰). It may be acceptable in certain cases to use that codepoint instead of the character proposed here. https://de.wikipedia.org/wiki/Deleatur (1) https://en.wikipedia.org/wiki/Dele (2)
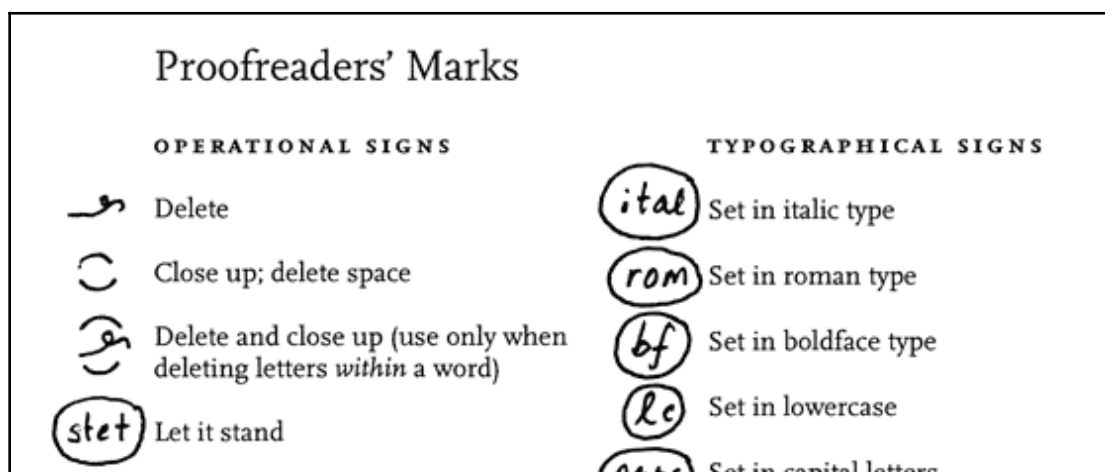

Figure 3. Excerpt the Chicago Manual of Style, showing the looped glyph and the interaction with 2050 ⁐ (https://www.chicagomanualofstyle.org/help-tools/proofreading-marks.html)

PROOFMARKS - How to communicate your text changes without ambiguity

In 1976 a British Standard for proofreading marks, BS 5261, was established. Although perhaps not so well known today, these marks remain an excellent way of conveying your text changes to someone else, whether you are editing words within your organisation or communicating with Lancing Press.

It is important that everyone work together with a common set of symbols to avoid misunderstandings. This chart shows some of the more commonly-used marks. A mark is usually placed within the body of the copy at the relevant place within the text with clarification in the margin.

Instruction	Textual mark	Marginal mark	Instruction	Textual mark	Marginal mark
Delete	/ through character or ⊢ through characters to be deleted	9/ (use oblique to separate multiple corrections in same line)	Change italic to roman type	circle character(s) to be changed	4/
			Change bold to roman type	circle character(s) to be changed	~/
Delete and close up	as above, but ⌐ or	9	Transpose characters or words	⌐⌐ between characters or words	⌐⌐/
Insert in text new matter indicated in the margin	⋏	new matter followed by ⋏	Insert space between words	Y between each word requiring spacing	Y/
Insert in text new punctuation indicated in	⋏	punctuation circled	Insert space between lines	─(between lines requiring spacing	textual mark extends into margin

Figure 4. Excerpt from the Lancing Press, describing the British Standard BS 5261. Notice the glyph used as well as other symbols like "transpose" (http://www.lancingpress.co.uk/printing-information/factfile_proofmarks_apr14.pdf)

Instruction	Textual mark	Marginal mark
Insert space between words	Y between each word requiring spacing	Y/
Insert space between lines	─(between lines requiring spacing	textual mark extends into margin
Reduce space between words or characters or lines	as above but ⌐ between each word or character or ⟶ between lines	⊤/ textual mark extends into margin
Delete space between words or characters	◠ between words or characters	◠/
Indent	⌐ round matter to be indented	⌐
Centre matter	⌐ ⌐ round matter to be centred	⌐ ⌐/
Start new paragraph	⌐ before first word of paragraph	⌐/
Run on	◠⌐	◠⌐/
Insert apostrophe	⋏ Where required	ỷ
Insert quotation mark	⋏ Where required	ỷ or ỷỷ and/or ỷ or /
Insert superior character	⋏ Where required	eg ²ỷ
Insert inferior character or figure	⋏ Where required	eg ⌐₂

Figure 5. Yet another excerpt from the same document. Notice the different shapes for 2041 ⋏ when they go in the margins as well as "run on"

Group 1.2 — Deletion, insertion, substitution, and appearance				
1.2.1	Delete and close up. In text, use vertical rule through single character and horizontal rule through multiple characters affected	ǀ or ⊢⊣	⌐	⌐ annd → and ⌐ pass the the salt → pass the salt ⌐ pass the sa lt → pass the salt
1.2.2	Substitute on baseline; to be used after any change that is not an insertion or deletion. In text, use vertical rule through single character and horizontal rule through multiple characters affected	ǀ or ⊢⊣	/	n/ abd → and cold/ It was hot. → It was cold.
1.2.3	Substitute superior character; to be used after any change that is not an insertion or deletion. In text, use vertical rule through single character and horizontal rule through multiple characters affected	ǀ or ⊢⊣	Y	^{2}Y m^3 → m^2 ^{23}Y 10^{123} → 10^{23}
1.2.4	Substitute inferior character; to be used after any change that is not an insertion or deletion. In text, use vertical rule through single character and horizontal rule through multiple characters affected	ǀ or ⊢⊣	$\wedge$	$\wedge_2$ vitamin B$_4$ → vitamin B$_2$ $\wedge_6$ vitamin B$_{12}$ → vitamin B$_6$

Figure 6. Excerpt from a preview version of ISO 5776:2016(E). Notice the glyph used, as well as the distinction on the inverted caret insertion point (https://www.sis.se/api/document/preview/920359/)

For general instructions

Instruction	Mark in margin	Mark in copy
Let the text stand (ignore any marks made to it & use the text as originally written)	STET	– – – – – under the characters to be left unchanged
Remove all non-textual or other marks	✕	Circle marks to be removed
Query to author: Is this as is supposed to be?	⑦	Circle text to be queried It was done by (NAFTA in 2011.)
Push down a work-up*	℔	Lance through an unintended mark

Figure 7. American proofreading marks. Notice the glyph used, as well as the simple overlayed looped version (https://learnenglishorstarve.wordpress.com/proofread/)
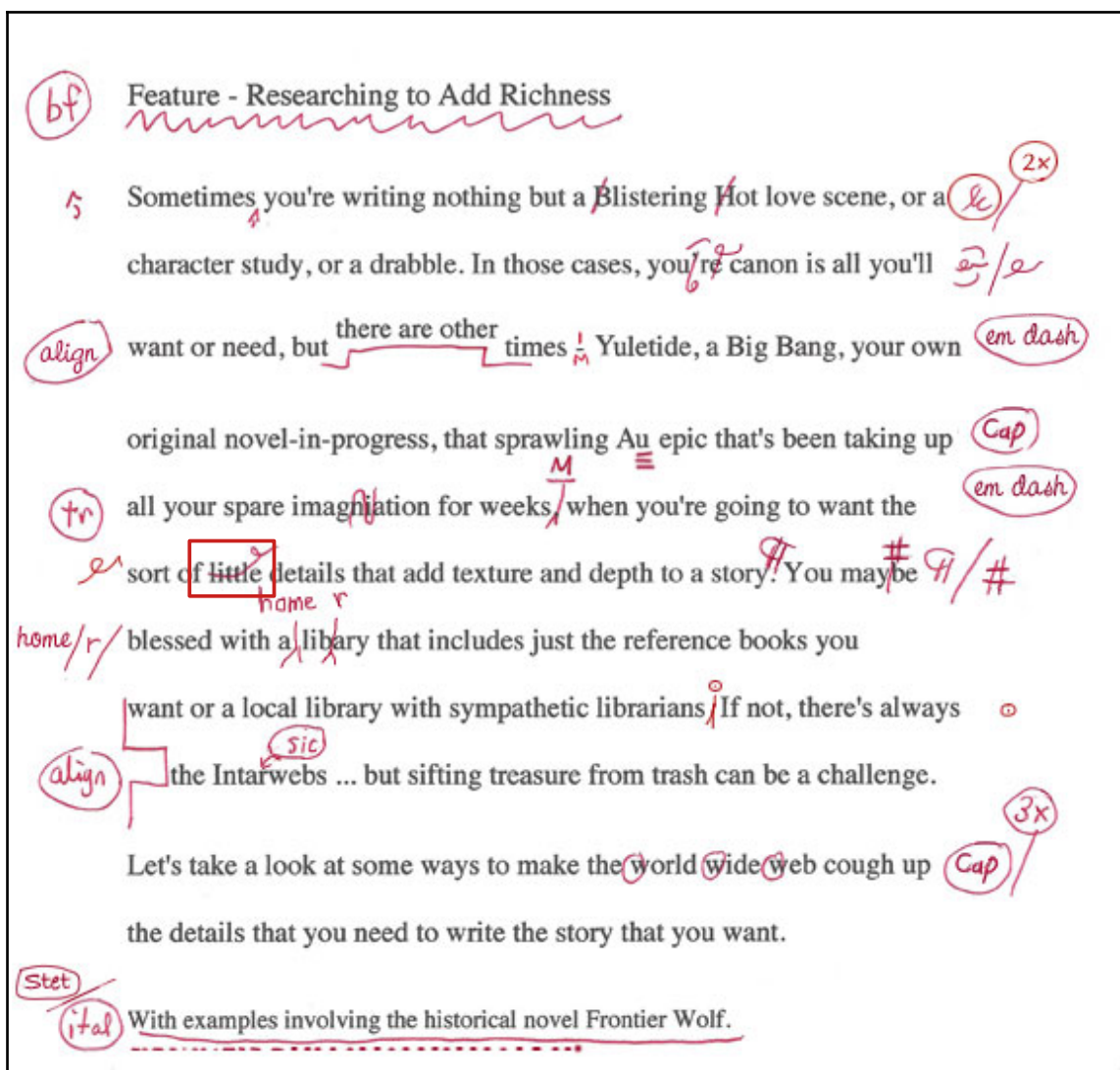
Figure 8. Example of marks being used in text. Notice the part enclosed in the red square, demonstrating the mark can also be used for entire words (https://fandom-grammar.livejournal.com/143792.html)

Using proofreading marks

Instruction	Textual mark	Margin mark
Delete	Strike through characters to be deleted del	9
Delete and close up	Above and below letter to be deleted	9
Transpose	Line between characters words or	trs
Change to lower case	Strike through characters to be changed	l.c.

Figure 9. Yet another attestation of the proposed glyph (https://www.nikkimgroup.com.au/features/pdf-proofreading-marks/using-proofreading-marks.pdf)