

Commas and Cedillas: Report on UTC 159-A81

Roozbeh Pournader

2026-09-02

During UTC meeting #159, I and Lisa Moore were assigned an action item to “Investigate feedback from David Corbett on Table 7-1 [Tue Apr 2 10:13:24 CDT 2019] in L2/19-124.” The feedback reads:

Table 7-1 “Preferred Rendering of Cedilla versus Comma Below” says that the comma below glyph of U+0327 COMBINING CEDILLA should be used with d, g, k, l, n, r, and t. The table should not recommend a comma glyph below t, because they are needlessly ambiguous with U+021B LATIN SMALL LETTER T WITH COMMA BELOW, and might encourage people to use the wrong diacritic in Romanian because it looks right in a particular font. The section “Latvian Cedilla” contradicts the table, saying that with g it should look like a rotated comma above. That section says “The uppercase form of the letter [G] is always shown with a cedilla” but the code chart shows it with a comma. The table should explain that U+0327 should look like a comma under certain capital letters, not just small letters.

I recommend accepting the feedback and making editorial corrections to the Core Spec.

I first looked at the document from which Table 7-1 originated. [L2/13-225](#) lists d, g, k, l, n, r—but not t—as having comma-shaped default glyphs. It explicitly says Romanian t should use U+0326 COMBINING COMMA BELOW because U+0327 COMBINING CEDILLA has a cedilla default with that letter. Table 7-1 was subsequently added under UTC action item [137-A75](#), apparently acquiring the extra t during editing.

Recommendations

T/t should move to the cedilla group. The current code charts renders U+0162 LATIN CAPITAL LETTER T WITH CEDILLA and U+0163 LATIN SMALL LETTER T WITH CEDILLA with a cedilla and annotates U+0163 to say that U+021B LATIN SMALL LETTER T WITH

COMMA BELOW should instead be used for Romanian. Independent feedback in [L2/19-325](#) from Denis Moyogo Jacquerye also documents actual cedilla use of T/t in Gagauz, Manjaku, Mankanya, and transliterations.

The Latvian G/g wording in the Core Spec is partially wrong. Contemporary Latvian typography uses a comma below capital G and a turned comma above lowercase g. The respective glyphs for U+0122 LATIN CAPITAL LETTER G WITH CEDILLA and U+0123 LATIN SMALL LETTER G WITH CEDILLA clearly show those preferences. The present statement that uppercase G is “always shown with a cedilla” contradicts the representative glyph in the latest Unicode charts, as well as evidence from various fonts and online sources.

Case needs to be explicit. The table's lowercase notation was evidently intended as a base-letter shorthand, but that is inadequate for G/g, whose preferred forms differ by case.

The Latvian paragraph has another omission: it lists k, n, and r but omits l. Modern Latvian uses comma-accented ģ, ķ, ļ, and ņ; ŀ is used in the historical Latvian orthography. The paragraph should include L/l and qualify R/r appropriately. Finally, because D/d with commas below are used in Livonian and not Latvian, it should mention Livonian.

Unicode 17 partially addressed the original complaint by changing the heading from “Cedilla versus Comma Below” to “Cedilla versus Comma”, allowing the g entry to cover an above-letter comma. But the table still contains “t” in the comma column, and the adjacent capital G prose remains incorrect.

Recommended table replacement

The table should be renamed to something like “Preferred Default Rendering of U+0327 by Base Letter” and make case and placement explicit:

Preferred default form	Base letters
Cedilla below	C/c, E/e, H/h, S/s, T/t

Comma below	D/d, G, K/k, L/l, N/n, R/r
Turned comma above	g

We should precede it with something like:

U+0327 COMBINING CEDILLA and characters canonically equivalent to sequences containing it may be displayed with either a traditional cedilla or a comma-shaped mark. In the absence of language information, the preferred default form depends on the base letter, as shown in Table 7-1. These defaults concern rendering, not the choice of character for encoding a particular language.

Recommended Latvian replacement

The subsection could be renamed “Latvian and Livonian Cedilla/Comma Accent”. The text should be replaced with something like:

In contemporary Latvian typography, letters encoded with characters named WITH CEDILLA are normally displayed with a comma-shaped diacritic. U+0122 LATIN CAPITAL LETTER G WITH CEDILLA is displayed with a comma below, whereas U+0123 LATIN SMALL LETTER G WITH CEDILLA is displayed with a turned comma above to avoid the descender of the lowercase letter. K/k, L/l, N/n, and historical R/r are displayed with a comma below. In Livonian, U+1E10 LATIN CAPITAL LETTER D WITH CEDILLA and U+1E11 LATIN SMALL LETTER D WITH CEDILLA are likewise displayed with a comma below. (Livonian similarly uses and displays L/l, N/n, and R/r with commas below.) Their names and canonical decompositions using U+0327 COMBINING CEDILLA are retained for compatibility. (“Normally” or “preferred contemporary form” is better than “always,” because historical Latvian printing had also used cedilla-like forms.)

Recommended Romanian clarification

The Romanian section should discuss both s and t, rather than illustrating only s:

Modern Romanian text should use U+0218/U+0219 LATIN CAPITAL/SMALL LETTER S WITH COMMA BELOW and U+021A/U+021B LATIN CAPITAL/SMALL LETTER T WITH COMMA BELOW, or the corresponding decomposed sequences

with U+0326 COMBINING COMMA BELOW. U+015E/U+015F and U+0162/U+0163, or sequences with U+0327, may occur in legacy Romanian data but should not be used for new Romanian text.

That distinction reflects the Romanian migration accepted by the UTC (see UTC Consensus [117-C9](#)) and prevents a font's rendering of U+0327 COMBINING CEDILLA from encouraging the wrong encoding.

Acknowledgement

Denis Moyogo Jacquerye kindly provided feedback on an earlier version of this document, leading to inclusion of Livonian in the proposed texts for the Core Specification.