

Proposal to correct General_category for a pair of ornamental guillemets

For consideration by Unicode Technical Committee

2019-04-22

Marcel Schneider (charupdate@orange.fr)

*“We must always say what we see.
Above all we must always
— which is the most difficult —
see what we see.”*

Alain Finkielkraut quoting Charles Péguy

Proposal History

This proposal targets the upcoming version 12.1 of the Standard, and is part of an on-going effort to help fix more errors in NamesList.txt, but couldn't be composed prior to L2/19-112 *Proposal to define a space character as a group separator*, its dependencies (L2/19-114, L2/19-115, L2/19-116) and its forerunner L2/19-113. In an attempt to meet the close deadline of UTC meeting #159 scheduled to start on April 30, 2019, exceptionally this paper has been composed over the Easter holidays. The goal is also that it may join these papers in the feedback hopper for simultaneous consideration, both to help mitigate the general tone of L2/19-112 and L2/19-114 through L2/19-116, and with respect to the following paragraph.

Among these preceding proposals, L2/19-113 through L2/19-116 were actually submitted in the evening of April 18 when Unicode had received L2/19-112. None of them arrived before. However, they were virtually submitted each one on the day of its datestamp. Consistently, they were left as-is when being submitted actually, although some of them could use updates and formatting fixes. The reason why things grew complicated is that altering the due course and the logical submission order would have been if not irresponsible, so at least unbearable, while on the other hand, diplomatic reasons discouraged from submitting L2/19-112 anyhow else than as first-of-series. To be clear, I'm afraid that if L2/19-112 were not submitted first, Asmus Freytag could take offense given he had suggested the topic on the Public Mailing List. Please keep that in mind when reading that these proposals posted on April 19 are referred to as having been submitted earlier.

Problem

Although they are encoded amidst paired punctuation marks whose Bidi_Paired_Bracket_Type property value is "open" or "close", U+276E HEAVY LEFT-POINTING ANGLE QUOTATION MARK ORNAMENT and U+276F HEAVY RIGHT-POINTING ANGLE QUOTATION MARK ORNAMENT are different in that they cannot have the same General_category property values as the surrounding characters. Their Gc values should be equal to those of the ordinary punctuation marks U+2039 SINGLE LEFT-POINTING ANGLE QUOTATION MARK and U+203A SINGLE RIGHT-POINTING ANGLE QUOTATION MARK that they are ornamental variants of.

Background

The General_category values "Pi" and "Pf" are designed for quotation marks with locale-dependent semantics. They are given to comma-shaped quotation marks (vertically aligned on caps height or baseline), but also to angular quotation marks since these don't have fixed semantics neither, as of occurring at the start or the end of quotations. A mark that is opening in some locales may be closing in some other locales, and conversely. Also, in some locales the same mark is used both at the start and at the end of quotations.

For example, U+201A SINGLE LOW-9 QUOTATION MARK and U+201E DOUBLE LOW-9 QUOTATION MARK are Gc=Ps because they may never occur at the end of quotations. By contrast, U+201B SINGLE HIGH-REVERSED-9 QUOTATION MARK and U+201F DOUBLE HIGH-REVERSED-9 QUOTATION MARK are Gc=Pi because although traditionally their glyphs were used to open quotations, some convention might use them as well to close quotations, so that opening semantics must not be taken for granted.

The semantics of U+276E HEAVY LEFT-POINTING ANGLE QUOTATION MARK ORNAMENT and U+276F HEAVY RIGHT-POINTING ANGLE QUOTATION MARK ORNAMENT is even less stable across locales as is that of those reversed quotation marks. As a consequence, they cannot be Gc=Ps and Gc=Pe, but must be Gc=Pi and Gc=Pf respectively.

On a side note: The Gc values "Punctuation, initial" ("Pi") and "Punctuation, final" ("Pf") per se don't reflect the fact that the semantics of the attributes "initial" and "final" is locale-dependent. But they are helpful in that they reflect the most common use case.

Proposed changes

1. In [UnicodeData.txt](#): Change Gc from "Ps" to "Pi" for U+276E HEAVY LEFT-POINTING ANGLE QUOTATION MARK ORNAMENT, and from "Pe" to "Pf" for U+276F HEAVY RIGHT-POINTING ANGLE QUOTATION MARK ORNAMENT.
2. In [BidiBrackets.txt](#): Remove U+276E and U+276F.
3. In [NamesList.txt](#): Add a subheading "Ornamental quotation marks" before U+276E; repeat the subheading "Ornamental brackets" before U+2770.

Acknowledgments

Thanks to everyone who directly or indirectly helped put this paper together.

Thanks to Microsoft for Word Online and OneDrive.