Source:	CheonHyeong Sim (沈天珩 / ᠰᡳᠨ ᡨᡳᠶᠠᠨ ᡥᡝᠩ) (Revised according to the discussion with Liang Hai and Ken Lunde)
Title:	Proposal to Add VS3 for Sibe Quotation Marks
Date:	2025−01−08
Action:	To be considered by UTC and Script Ad Hoc (and maybe CJK/Unihan Group, Properties & Algorithms Group)

Proposal

This proposal requests to add Standardized Variation Sequences for four quotation marks for the Sibe form as VS3 as in the following table: **(Note that, VS1 and VS2 for these quotation marks have already been published from Unicode 16.0)**

Codepoint	VS1 (U+FE00) Western Form		VS2 (U+FE01) Eastern Form		VS3 (U+FE02) Sibe Form	
	Hori.	(Vert.)	Hori.	Vert.	(Hori.)	Vert.
U+2018	a'a	a'a	字'字	字字	᠊'᠊	ᡯ'ᡯ
U+2019	a'a	a'a	字'字	字字	᠊'᠊	ᡯ'ᡯ
U+201C	a"a	a"a	字"字	字字	᠊"᠊	ᡯ"ᡯ
U+201D	a"a	a"a	字"字	字字	᠊"᠊	ᡯ"ᡯ

The glyphs of the quotation marks are highlighted in orange. In western texts and Sibe texts, a quotation mark usually appears with a space character before or after it, those space characters are highlighted in grey; in eastern texts, the spaces before or after the quotation marks are provided by the quotation marks themselves, thus no additional space characters are needed.

1/8

The proposed names for the VSes are as follows (the current names for VS1 and VS2 for these quotation marks are listed together for reference):

VS / Base	U+FE00 (Current Names)	U+FE01 (Current Names)	U+FE02 (Proposed Names)
U+2018	non−fullwidth form	right−justified fullwidth form	Sibe form
U+2019	non−fullwidth form	left−justified fullwidth form	Sibe form
U+201C	non−fullwidth form	right−justified fullwidth form	Sibe form
U+201D	non−fullwidth form	left−justified fullwidth form	Sibe form

Note that, Ken Lunde is reluctant to use "Sibe" in the names of the proposed VSes, in case it turns out that they are useful for other languages or contexts; however, Cheon-Hyeong Sim thinks that the worry is redundant, since there will unlikely to be any other vertically written scripts using **such modern** punctuations. Liang Hai once suggested to name these VSes as "Sibe **fullwidth** form", but actually they are not fullwidth (see the figures in the main body of this proposal), so CheonHyeong Sim thinks that "Sibe form" is enough.

Also, this proposal suggests to change the Vertical_Orientation properties of these quotation marks from R to Tr, see the last chapter for detailed reasons.

2018	; Tr	# Pi	LEFT SINGLE QUOTATION MARK
2019	; Tr	# Pf	RIGHT SINGLE QUOTATION MARK
201C	; Tr	# Pi	LEFT DOUBLE QUOTATION MARK
201D	; Tr	# Pf	RIGHT DOUBLE QUOTATION MARK

Background

Dr. Ken Lunde suggested to register the VS1 and VS2 as in the above table in L2/23−212, and revised his proposal after the SAH meeting on Oct. 13th, 2023. VS1 and VS2 were accepted for publication during UTC#177 meeting in Nov., 2023, and were formally published from Unicode 16.0. This is an additional request to register the VSes for the Sibe (锡伯, ᠰᡳᠪᡝ) forms.

Since the quotation marks encoded in the General Punctuation block are widely used both in the West and in East Asia, it is impossible to disunify anyone to the new code-point due to possible bad social impacts. One of the main issues here seems to be the plain text layout, see the original proposal from Ken (aka L2/23−212 mentioned above) for reference. However, there is another issue that Ken did not mention, which is the issue from the Vertical_Orientation property. Since Sibe needs a vertical layout in any case, it seems to be quite important. The details will be described in the following texts.

In my previous proposal *Propose to Add Script_Extension for some CJK Punctuations* (aka L2/23−272), many "script−specific localizations" were mentioned; but in this case, **the font−level script−specific localizations still cannot solve all the problems**. The reason is, a character with a Script value "Zyyy" normally inherits the value of the preceding character unless there is no character preceded, see also L2/23−272 for reference. An example will be given in the next chapter.

The Sibe form of quotation marks

First let us see the evidence showing the quotation marks in Sibe form from Cabcal ([ʧʰapʧʰal]) Newspaper:

Fig.1 《察布查尔报》（锡伯文） 2008/05/28

According to Cabcal Newspaper, the quotation marks are shown in a curved form (弯引号 in Chinese) even if they appear in a vertical layout, which is very different from the Eastern form (aka VS2). In Hong Kong, Macao, Taiwan and Japan, the corner brackets (直角引号 in Chinese) are preferred as quotation marks, regardless of whether they are in horizontal layouts or vertical layouts; in the mainland of China, the curved forms are preferred in horizontal layouts, while the corner brackets are preferred in vertical layouts, thus, U+2018, U+2019, U+201C, U+201D should be substituted by the corner brackets in vertical layouts in default − we do not want to change the codepoints when horizontal text and vertical text convert to each other − just like, U+FE12 PRESENTATION FORM FOR VERTICAL IDEOGRAPHIC FULL STOP do exist, but we still use U+3002 IDEOGRAPHIC FULL STOP in the vertical layouts.

Also, although the Western scripts themselves do not use the vertical layouts in general, it does not mean that we could merge the quotation marks in Sibe form (which is only needed in vertical layouts) into VS1. When mixing with Han script or Mongolian script,

the vertical forms of the quotation marks in Western form are still needed. Generally, the quotation marks in Western form rotate 90 degrees clockwise in vertical layouts, as well as the digits and the letters (just see the table in the first page). Since the Sibe form is different from both the Western form and the Eastern form, registering them as VS3 might be necessary.

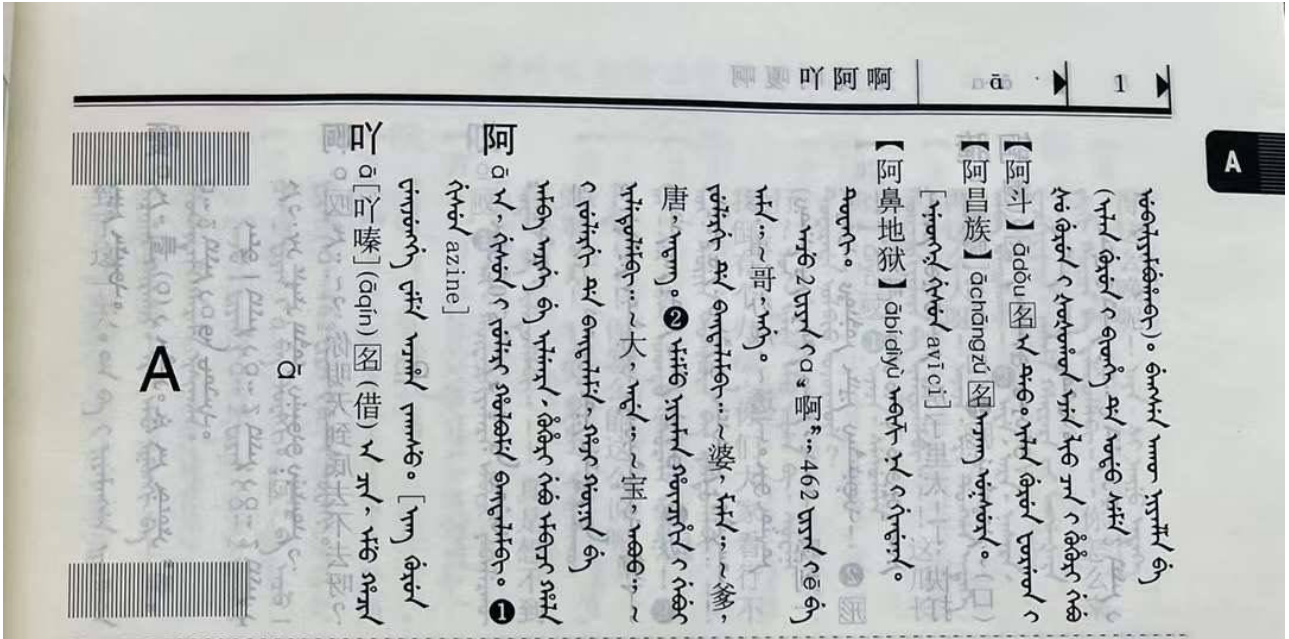The picture below shows another evidence of the quotation marks in Sibe form:

Fig.2 　《汉锡词典》P1

What is more, an issue related to the font–level script–specific localization is secreted in the picture. The left quotation mark follows a Latin character "a", while the right quotation mark follows a Han character "啊", which may lead to an inconsistency between the two quotation marks when the script–specific localization is active:

ᠰᡠᡩᡝ 2 ᠮᡠᡩᠠᠨ ᠪᡝ "阿" :: 462 ᠮᡠᡩᠠᠨ ᠪᡝ ᡝᡳᠯᡝ ᠋᠋ᠨᡝᡳᠮᡝ ᠊ ᠃

　　If the VSes are registered, then we can manually change the forms freely in any contexts.

Fullwidth or not?

In Fig.1, we could see that there are some extra spaces before the opening quote and after the closing quote. Generally, we should have extra space characters before the opening quotes and after the closing quotes in order to break the line correctly, just as what many Western texts (e.g. English texts) usually do. What is more, extra spaces are needed not only for quotes but also for other punctuations such as the full stop, the comma, etc., which makes the use of different punctuations more consistent, we could see that there are more spaces after such punctuations than before them from the **justified** Sibe texts in Fig.3:

Fig.3 《察布查尔报》（锡伯文） 2024/11/20

Also, in Fig.3, we could see that when a closing quote is followed by a comma, then an extra space is not needed after the quote but only needed after the comma; and, the quote is obviously NOT fullwidth without the extra space.

Meanwhile, we could have a look at how other writing systems of the Mongolian script use punctuations, for example the most common writing system − Hudum Mongolian. For example we may see the articles from mgyxw.cn (中国蒙古语新闻网, ᠦᠨᠳᠦᠰᠦᠲᠡᠨ ᠤ ᠮᠣᠩᠭᠤᠯ ᠬᠡᠯᠡᠨ ᠤ ᠰᠤᠨᠢᠨ ᠮᠡᠳᠡᠭᠡᠨ ᠤ ᠰᠦᠯᠵᠢᠶᠡ), mgl.bynr.gov.cn (巴彦淖尔市人民政府, ᠪᠠᠶᠠᠨᠨᠠᠭᠤᠷ ᠬᠤᠲᠠ ᠶᠢᠨ ᠠᠷᠠᠳ ᠤᠨ ᠵᠠᠰᠠᠭ ᠤᠨ ᠤᠷᠳᠤᠨ), etc. It is easy to discover that, extra space characters appear after the Mongolian commas (U+1802), after the Mongolian full stops (U+1803), before the Mongolian opening quotes (U+300A), and after the Mongolian closing quotes (U+300B). Note that, U+300A and U+300B are originally the book title marks from Chinese, and the Sibe orthography also uses them as the book title marks (see also Fig.3), only few writing systems such as the Hudum Mongolian or Tibetan treat them as quotes. Therefore it is also a consistency within the Mongolian script.

Design for Sibe form in horizontal layouts

Traditionally, the quotation marks in Sibe form never occur in horizontal layouts, however, we still need a design of them to adapt to the modern technology. Firstly, if the VSes are accepted, then we have to show their example glyphs in the CodeCharts, which have to be horizontal forms; Secondly, vertical plain texts are very infrequent (they **do** exist, e.g. Bainu, an instant messenger software designed for Mongols), most of the plain texts are horizontal, e.g. QQ, an instant messenger software widely used in China. When the quotation marks with VS3 are input in the horizontal plain texts, in any case, they have to be displayed. Here I suggest the unrotated forms (i.e. the shape in the first chapter "Proposal") to be used, just like the exclamation mark, the question mark, etc., as their default forms in horizontal layouts. Thus, if the VSes are accepted, a notation should be added into the CodeCharts in order not to confuse people since VS1 and VS3 look similar.

Issue related to vertical orientation

Currently, the Vertical_Orientation properties for U+2018, U+2019, U+201C, U+201D are all set to R, which means they should be rotated in any case in the vertical layouts. What is more, for these characters in the vertical layouts, **different shaping engines usually have inconsistent implementations to the same character**, which makes font designers very confused.

Harfbuzz

In Harfbuzz, the OpenType Layout feature "vert" only applies to the characters with Vertical_Orientation properties U, Tu or Tr in the vertical layouts. In other words, this feature would not apply to the quotation marks automatically, unless you manually enable the feature. What is more, even if the feature is manually enabled, the substituted glyphs will still maintain rotated (**consistent with current VO properties**). To make them display correctly, these are a part of the glyphs and the Single Substitution ("vert" feature) in my font:

$201C					$201D				
"	"	"	"	�ⸯ	"	"	"	"	ⸯ

Glyph	Name
" → 『	quotedblleft uni300E
" → 『	quotedblleft.2 uni300E
" → ⌐	quotedblleft.3 quotedblleft.3.vert
" → 』	quotedblright uni300F
" → 』	quotedblright.2 uni300F
" → ⌐	quotedblright.3 quotedblright.3.vert

The glyphs after substitution seems to be " 『 " and " 』 ", but actually they would be shown as " ⌐ " and " ⌐ " in the vertical layouts since they are always rotated.

DirectWrite

However, in contrast, in the environments using DirectWrite, for example Microsoft Office, the quotation marks are always unrotated even if they have the Vertical_Orientation properties as R (**inconsistent with current VO properties**; my personal view is that, these four code points in vertical layout were specially processed in this shaping engine), and meanwhile, the "vert" feature **can** be automatically applied to them. The problem here is, since they are always unrotated, the fonts which show correctly in Harfbuzz will become wrong ones in Microsoft Office:

$$『\\字\\』$$

On the contrary, the fonts which show correctly in Microsoft Office, e.g. SimSun, SimKai, SimHei, etc., will become wrong ones in Harfbuzz. In brief, no font can show the quotation marks in the vertical layouts correctly in both Harfbuzz and DirectWrite. **The implementations for these four codepoints are completely incompatible between different shaping engines.** Although that seems to be more likely an issue related to the shaping engine, changing the Vertical_Orientation properties from R to Tr would be better (then, Harfbuzz can automatically apply the "vert" feature, and DirectWrite no longer needs special treatment for these codepoints).

(End of document)