

Proposal to synchronize seven glyphs in the Code Charts

For consideration by Unicode Technical Committee

2020-01- 13 (revised; first submitted 2020-01-06)
Marcel Schneider (charupdate@orange.fr)

*We should always say what we see.
Above all we should always
—which is most difficult—
see what we see.*

Charles Péguy

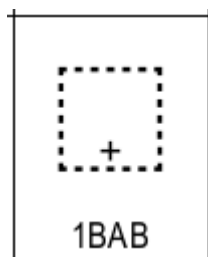
This proposal adds to the response to Action item 161-A1 in that it aims at synchronizing the Core Specification with changes already effected to other parts of the Unicode Standard or suggested in *Proposal to make material changes to UAX #14*, or *Proposal suggesting formal edits to UAX #14*, submitted simultaneously. A minor part of this proposal relies on *Proposal to extend support for abbreviations*, scheduled for simultaneous submission.

By coincidence, this proposal is also part of [Unicode 13.0 beta](#) feedback.

These suggestions are sorted by importance and concomitant likelihood of acceptance rather than by code points.

1 SUNDANESE SIGN VIRAMA U+1BAB

Change from:



Change to:



Rationale:

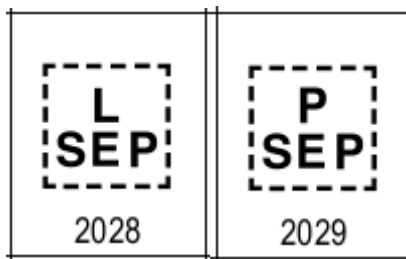
Among the eleven invisible stackers encoded so far, listed in the table below (where brackets stand for a dashed box), Sundanese *virama* is the only one whose reference glyph is lacking the dotted circle U+25CC above the (combining) plus sign below.

In the Change-to picture above, the COMBINING PLUS SIGN BELOW its DOTTED CIRCLE in a dashed box has been copy-pasted from the reference glyph of U+1039 MYANMAR SIGN VIRAMA.

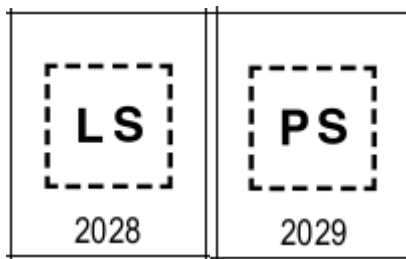
cp	Name	Chart glyph
1039	MYANMAR SIGN VIRAMA	[◌]
17D2	KHMER SIGN COENG	[◌]
1A60	TAI THAM SIGN SAKOT	[◌]
1BAB	SUNDANESE SIGN VIRAMA	[◌]
AAF6	MEETEI MAYEK VIRAMA	[◌]
10A3F	KHAROSHTHI VIRAMA	[◌]
11133	CHAKMA VIRAMA	[◌]
11A47	ZANABAZAR SQUARE SUBJOINER	[◌]
11A99	SOYOMBO SUBJOINER	[◌]
11D45	MASARAM GONDI VIRAMA	[◌]
11D97	GUNJALA GONDI VIRAMA	[◌]

2 LINE SEPARATOR..PARAGRAPH SEPARATOR U+2028..U+2029

Change from:



Change to:



Rationale:

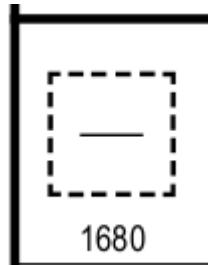
The abbreviations “LS” and “PS”, used in the Core Specification, are consistent with the following control characters and their Control Picture glyphs in the Code Charts at U+241C..U+241F, while the mnemonics in the reference glyphs of LS and PS, actually LSEP and PSEP, need to be synched.

Templates backing the two-letter abbreviations of LINE SEPARATOR and PARAGRAPH SEPARATOR:

cp	ISO 6429 name	Alias	Abbr	Glyph	Ctrl Pict
001C	INFORMATION SEPARATOR FOUR	FILE SEPARATOR	FS →	 001C	 241C
001D	INFORMATION SEPARATOR THREE	GROUP SEPARATOR	GS		
001E	INFORMATION SEPARATOR TWO	RECORD SEPARATOR	RS		
001F	INFORMATION SEPARATOR ONE	UNIT SEPARATOR	US		

3 OGHAM SPACE MARK U+1680

Change from:





Change to:



Rationale:

By adding “OG” and “SP” above and below the stemline of U+1680, this glyph may be completed according to the wide consensus as it appears in the 2007 Action Item 113-A15. It’s about synching the glyph with those of the other space characters. Preferred was the fourth glyph depicted in [L2/08-142](#), with the stemline and the four Latin capitals, after Michael Everson explained in [L2/07-392](#) that the Ogham space behaves like any other white space in that it disappears at line end, and that the stemline is unessential. Nevertheless, the Ogham space stemline is an important feature of this space’s reference glyph. Only the mnemonic of this space character is missing to date. It should be added in accordance with a longstanding practice of space character reference glyph design in the Unicode Standard.

Space

1680  OGHAM SPACE MARK
 • glyph is blank in “stemless” style fonts
 → 0020  space

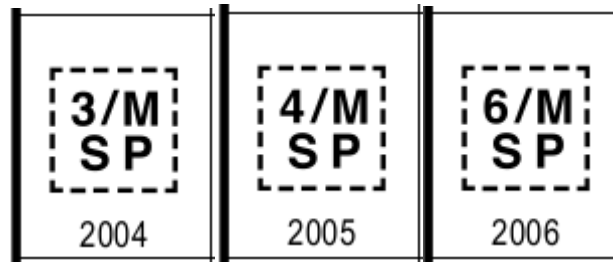
Annotation in the Code Charts stating that there may be no line in the glyph of OGHAM SPACE MARK.

If [L2/08-318](#), section 9.14, fell short of completing that change, making a new attempt now is appropriate because omitting the four Latin capitals in the reference glyph of the Ogham space is contradicting the claim about the optional status of the stemline. Since Ogham can be written stemless, and stemless Ogham fonts do exist on the marketplace, a bare stemline can by no means be a distinctive sign of the OGHAM SPACE MARK reference glyph. When we think of the stemline as missing, we’ll end up today with a blank. The abbreviation inside the dashed box reminds us that the stemline may not be there.

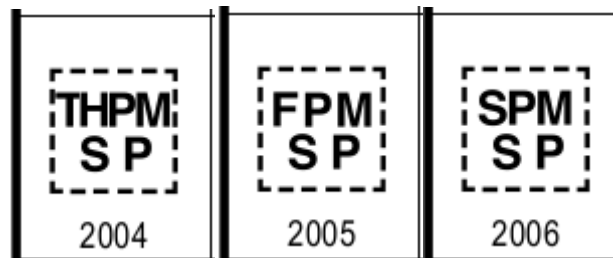
Initially there was no dashed box either. When that flaw was on the table, things were ready to get “OG” and “SP” in alongside. Now that all spaces are hopefully going to have their abbreviation standardized (as suggested in *Proposal to extend support for abbreviations*), this move also includes OGHAM SPACE MARK. There seems to be a good occasion for making Ogham script cease standing out, by adding mnemonic letters to make it a fully-fledged Code Charts space character reference glyph.

4 THREE-PER-EM SPACE..SIX-PER-EM SPACE U+2004..U+2006

Change from:



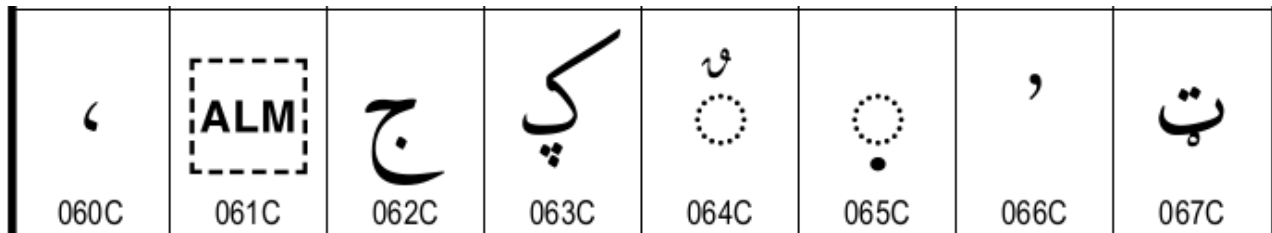
Change to:



Rationale:

Replacing “3/MSP” with “THPMSP”, “4/MSP” with “FPMSP”, and “6/MSP” with “SPMSP”, in the glyphs of three-, four- and six-per-em spaces is optional and only aims at consistency. Although glyph mnemonics do not need to comply to namespace constraints—some are even using scripts other than Latin—Unicode may wish to consistently provide a standard abbreviation for every space character, just like it does provide a name for every character, and would then use those abbreviations throughout.

Conversely indeed reference glyph mnemonics can be freely chosen since they don’t need to stick with the locale of the block, as exemplified by U+061C ARABIC LETTER MARK:



This demonstrates that when Unicode feels the need to use a standard abbreviation composed of Latin capitals consistently with Unicode character namespace constraints, add it to NameAliases.txt and make it show up inside the dashed box of the reference glyph in the Code Charts, there is always a means to do so even when the script is non-Latin, Arabic for instance.

This contrasts with the prepended concatenation marks called sub- or supertending marks ARABIC NUMBER SIGN, ARABIC SIGN SANAH, ARABIC FOOTNOTE MARKER, ARABIC SIGN SAFHA, ARABIC SIGN SAMVAT and ARABIC NUMBER MARK ABOVE encoded at block start:



Acknowledgments

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

Thanks to Microsoft for Word Online, OneDrive, VS Code and MSKLC.

Thanks to Google for Google Chrome, Google Search, Google Books, Google Translate and Gmail.