A. Title: Revised Proposal for Additional Latin Phonetic and Orthographic Characters

2. Requester's name: Lorna A. Priest

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution


5. Requester's reference (if applicable): L2/04-246R

6. Choose one of the following:
   a. This proposal is for a new script (set of characters):
      Proposed name of script: No
   b. The proposal is for addition of character(s) to an existing block: Yes
      Name of the existing block: Combining Diacritical Marks, Latin Extended, Phonetic Extensions

2. Number of characters in proposal: 21

3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary x B.1-Specialized (small collection) B.2-Specialized (large collection)
   C-Major extinct D-Attested extinct E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols

4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): 3 (some combining marks)
   Is a rationale provided for the choice? No

5. Is a repertoire including character names provided? Yes
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? Yes
   b. Are the character shapes attached in a legible form suitable for review? Yes

6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? SIL International
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

7. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? Yes
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? Yes

8. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? Yes, suggested character properties are included, as are lower case representations (see section D)

9. Additional Information:
   Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?  
   No

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?  
   Yes

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?  
   Yes

4. The context of use for the proposed characters (type of use; common or rare)  
   Common

5. Are the proposed characters in current use by the user community?  
   Yes

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?  
   Preferably

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?  
   Preferably together with other related blocks

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?  
   No

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?  
   Yes
   Characters with stroke or bar might be construed as precomposed forms of sequences with combining overlay marks.

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?  
    No

11. Does the proposal include use of combining characters and/or use of composite sequences?  
    No

12. Does the proposal contain characters with any special properties such as control function or similar semantics?  
    No

13. Does the proposal contain any Ideographic compatibility character(s)?  
    No

D. Proposed Characters

A code chart and list of character names are shown on a new page.
<table>
<thead>
<tr>
<th>D.1. Proposed Characters</th>
<th>D.1. Character Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DC 024</td>
<td>1DC4 COMBINING MACRON-ACUTE</td>
</tr>
<tr>
<td>1</td>
<td>1DC5 COMBINING GRAVE-MACRON</td>
</tr>
<tr>
<td>2</td>
<td>1DC6 COMBINING MACRON-GRAVE</td>
</tr>
<tr>
<td>3</td>
<td>1DC7 COMBINING ACUTE-MACRON</td>
</tr>
<tr>
<td>4</td>
<td>1DC8 COMBINING GRAVE-ACUTE-GRAVE</td>
</tr>
<tr>
<td>5</td>
<td>1DC9 COMBINING ACUTE-GRAVE-ACUTE</td>
</tr>
<tr>
<td>6</td>
<td>1DCA COMBINING LATIN SMALL LETTER R BELOW</td>
</tr>
<tr>
<td>7</td>
<td>0242 LATIN CAPITAL LETTER B WITH STROKE</td>
</tr>
<tr>
<td>8</td>
<td>• lowercase is 0180 b</td>
</tr>
<tr>
<td>9</td>
<td>0243 LATIN CAPITAL LETTER U BAR</td>
</tr>
<tr>
<td>A</td>
<td>• lowercase is 0289 u</td>
</tr>
<tr>
<td>B</td>
<td>0244 LATIN CAPITAL LETTER TURNED V</td>
</tr>
<tr>
<td>C</td>
<td>• lowercase is 028C ʌ</td>
</tr>
<tr>
<td>D</td>
<td>0245 LATIN CAPITAL LETTER E WITH STROKE</td>
</tr>
<tr>
<td>E</td>
<td>0246 LATIN SMALL LETTER E WITH STROKE</td>
</tr>
<tr>
<td>F</td>
<td>0247 LATIN CAPITAL LETTER J WITH STROKE</td>
</tr>
<tr>
<td></td>
<td>0248 LATIN SMALL LETTER J WITH STROKE</td>
</tr>
<tr>
<td></td>
<td>0249 LATIN CAPITAL LETTER SMALL Q WITH HOOK TAIL</td>
</tr>
<tr>
<td></td>
<td>024A LATIN SMALL LETTER Q WITH HOOK TAIL</td>
</tr>
<tr>
<td></td>
<td>024B LATIN CAPITAL LETTER R WITH STROKE</td>
</tr>
<tr>
<td></td>
<td>024C LATIN SMALL LETTER R WITH STROKE</td>
</tr>
<tr>
<td></td>
<td>024D LATIN CAPITAL LETTER Y WITH STROKE</td>
</tr>
<tr>
<td></td>
<td>024E LATIN SMALL LETTER Y WITH STROKE</td>
</tr>
<tr>
<td></td>
<td>024F LATIN SMALL LETTER V WITH CURL</td>
</tr>
</tbody>
</table>
D.3. Unicode Character Properties

1DC4..1DC9 should have a general category of Mn, and a canonical combining class of 230. Other properties should match those of similar characters, such as U+0300 COMBINING GRAVE ACCENT.

1DCA should have a general category of Mn, and a canonical combining class of 220. Other properties should match those of similar characters, such as U+0300 COMBINING GRAVE ACCENT.

0242..0245, 0247, 0249, 024A, 024B, and 024D should have a general category of Lu. Other properties for these characters should match those of similar characters, such as U+0041 LATIN CAPITAL LETTER A.

Other characters should have a general category of Ll. Other properties for these remaining characters should match those of similar characters, such as U+0061 LATIN SMALL LETTER A.

E. Other Information

E.1 Combining Diacritical Marks

The characters COMBINING MACRON-ACUTE, COMBINING GRAVE-MACRON, COMBINING MACRON-GRAVE, COMBINING ACUTE-MACRON are phonetic symbols which represent contour tone.

COMBINING MACRON-ACUTE and COMBINING GRAVE-MACRON are attested in the IPA Handbook. The other four contour-tone diacritics are not cited in the IPA handbook, but are widely used by linguists.

Figure 1. COMBINING MACRON-ACUTE and COMBINING GRAVE-MACRON (IPA, 1999, p. 23).

Figure 2. COMBINING MACRON-ACUTE and COMBINING MACRON-GRAVE (Gilley, 1992, p. 51).
Figure 3. COMBINING GRAVE-MACRON, COMBINING MACRON-ACUTE and COMBINING ACUTE-MACRON (Kutsch Lojenga, 1993, p. xii).

Figure 4. COMBINING GRAVE-MACRON and COMBINING ACUTE-MACRON (Kutsch Lojenga, 1993, p. 408).

Figure 5. COMBINING ACUTE-MACRON (Gilley, 1992, p. 53).
In addition to being used phonetically, COMBINING GRAVE-MACRON and COMBINING MACRON-GRAVE are also used orthographically in the Bette language of Nigeria.

Figure 6. COMBINING GRAVE-MACRON and COMBINING MACRON-GRAVE (IBS, 1982, p. 198).

COMBINING GRAVE-ACUTE-GRAVE and COMBINING ACUTE-GRAVE-ACUTE are phonetic symbols which represent contour tone.

Figure 7. COMBINING GRAVE-ACUTE-GRAVE and COMBINING ACUTE-GRAVE-ACUTE (Burquest, 1998, p. 234).

COMBINING LATIN SMALL LETTER R BELOW is used orthographically in four languages in Indonesia, namely Mongondow, Sangir, Siau and Talaud.

Figure 8. COMBINING GRAVE-ACUTE-GRAVE and COMBINING ACUTE-GRAVE-ACUTE (Goldsmith, 1995, p. 446)

Buké i Matius muluhabaré si kité, Habaré’u Ralúasé, humonéú i Yesus Kristus ko Datung Salamaté konipudiandi-ngu Mawu Ruata. Bou i Yesus éé Mawu Ruata nungganapé apang ko nipudiandi-Né su raljungu Pudariandi Tebé su manga umaté-É. Maningu i Yesus nipuhana bou taumatangu Yahudi, kaiso Habaré’u Ralúasé éé mang bañíñébé ketang gunangu tau Yahudi, kaiso mang lai gunangu kase-lahé’u dunia.

Figure 9. COMBINING LATIN SMALL LETTER R BELOW (Lembaga Alkitab Indonesia, 2002, p. 1).
E.2 Latin Extended

LATIN CAPITAL LETTER B WITH STROKE is an orthographic character used for at least two languages of Vietnam (Jorai and Katu). Lower case is represented in Unicode as U+0180 LATIN SMALL LETTER B WITH STROKE.
LATIN CAPITAL LETTER U BAR is used orthographically for a number of languages around the world. These include the Mesem and Melpa (see Figures 15 and 16) languages of Papua New Guinea, Sayula Popoluca of Mexico, the Badwe’e language of Cameroon (see Figure 17), the Budu language of Democratic Republic of Congo (see Figure 14), Comanche (see Figure 18) and Arhuaco of Colombia (see Figure 23). Lower case is represented in Unicode as U+0289 LATIN SMALL LETTER U BAR.
Figure 16. LATIN CAPITAL LETTER U BAR (WHBL, 1988, Cover).

Figure 17. LATIN CAPITAL LETTER U BAR (WBT, 2003, Back of front cover).
LATIN CAPITAL LETTER TURNED V is used orthographically in the Nankina language of Papua New Guinea (see Figure 19) and North Tepehuan of Mexico (see Figure 20). Lower case is represented in Unicode as U+028C LATIN SMALL LETTER TURNED V.

The characters LATIN SMALL LETTER E WITH STROKE and LATIN CAPITAL LETTER E WITH STROKE are used orthographically in the Southeastern Tepehuan language of Mexico.
Although there are currently no publications using LATIN CAPITAL LETTER E WITH STROKE, it should be encoded as it is an orthographic character. A dictionary is currently in the process of being published which uses LATIN CAPITAL LETTER E WITH STROKE.

The characters LATIN SMALL LETTER J WITH STROKE and LATIN CAPITAL LETTER J WITH STROKE are used orthographically in the Arhuaco (Ethnologue code: ARH) language of Colombia. It is being used for a voiced alveopalatal affricate like the j of "just" in English. It has been used for perhaps the last 10 years, but little literature has been produced to date. The letter "j" also occurs in Arhuaco for borrowed words from Spanish.
LATIN SMALL LETTER Q WITH HOOK TAIL and LATIN CAPITAL LETTER SMALL Q WITH HOOK were introduced by early Lutheran missionaries for the Numanggang language of Papua New Guinea (1930s or 1940s). After a number of decades of use, in 2002 the community decided to discontinue using LATIN SMALL LETTER Q WITH HOOK TAIL and LATIN CAPITAL LETTER SMALL Q WITH HOOK, but they are found in existing publications such as song books and liturgy materials. They are needed for electronic publishing of archived documents. They are also found in Kâte language (distantly related to Numanggang) literature. As shown in Figure 24 and 26 “q” also appears in the literature and could not be considered an alternate form.

Figure 24. LATIN SMALL LETTER Q WITH HOOK TAIL (BFBS, 1965, p. 105)

**Wofunj Jesu Kristo**

**Ere Miti Qari Zi**

Figure 25. LATIN CAPITAL LETTER SMALL Q WITH HOOK TAIL (BFBS, 1965, Title page).
LATIN SMALL LETTER R WITH STROKE and LATIN CAPITAL LETTER R WITH STROKE are used in the Kanuri orthography of Niger.
The characters LATIN SMALL LETTER Y WITH STROKE and LATIN CAPITAL LETTER Y WITH STROKE are used orthographically in the Lubuagan Kalinga language of the Philippines. The Lubuagan people are strong in their determination to keep this symbol. No examples of the LATIN CAPITAL LETTER Y WITH STROKE are available: the y-stroke is never word-initial in Kalinga text, and so the capital would not occur in most body text. Like any Latin characters used orthographically, however, it can occur in document elements set in full caps, and so must also be encoded.
Figure 30. LATIN SMALL LETTER Y WITH STROKE (Pateuweg, 1998, p. 17).
E.3 Phonetic Extensions

LATIN SMALL LETTER V WITH CURL is a phonetic symbol used to indicate a labiodental flap. It is not approved IPA usage, yet it has been widely used, and is in current use, particular among Africanist linguists.

Voiced labiodental flap, made by moving the lower lip inwards behind the upper teeth and then flapping it against the upper teeth outwards (phonetic symbol ʁ). This sound seems to occur in ideophones only, e.g. binbo (describing sud-

Figure 31. LATIN SMALL LETTER V WITH CURL (Hoffman, 1963, p. 25).

In Niger-Kordofanian almost all known examples are from Adamawa-Eastern. Once more Tucker and Bryan describe it as most common in the Ndoogo-Sere group, examples being given from Sere, Mundu, Ndoogo and Bai and once more in common words. Particularly interesting is its occurrence in Ndoogo, Sere and Bai in the word of 'child': an obvious cognate of the widespread Niger-Congo root *k. Richardson gives an example from an Adamawa language Kapere, a southern dialect of Mbum in which it occurs in the second person plural pronoun. He also notes its occurrence in Ngba Mabo, a form of Maka Limba, and in the Yangere dialect of Banda, both of these once more languages of the Eastern division of Adamawa-Eastern.

Figure 32. LATIN SMALL LETTER V WITH CURL (Greenberg, 1983, p. 11).

Sounds (DGBAKA MABO)

There are five phonemes ɨ, ɛ, ɔ, ʊ with subsidiary members in unstressed syllables, also ɔ, ə. Amongst the consonants the following are noteworthy: kp, gb, t (flapped r), (flapped v), and v with syllabic value, e.g. kpâː nê 'sleep', ndzg̥eː nê 'woman', ɲını 'nine', ʊv̥ɛ 'five'.

Tone. Both nominals and verbs may be categorized according to tone, which often distinguishes between lexical entities, e.g. kpâː nê 'branch', kpâː nê 'leaf'.

Word shape. No closed syllables occur. Simple words are generally of the shape CV or CV.CV. In the latter case, when the second consonant is l or r, the two vowels are the same. VCV is very rare, e.g. gwâ 'fire'. Compound words are very common, e.g. kpâː nê 'branch' (hand' + 'tree').

Figure 33. LATIN SMALL LETTER V WITH CURL (Richardson, 1957, p. 91).

321. ɨ. The characteristic consonant sounds kp and gb are found everywhere; 'b and 'd are found everywhere except in Zande. In addition a flapped l-sound is heard throughout (usually as a variant of l or r), while Ndogo and Kreish pronounce a flapped v-sound.

<table>
<thead>
<tr>
<th>MORU</th>
<th>BONGO</th>
<th>NDODO</th>
<th>ZANDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>kp</td>
<td>ekpi (to cough)</td>
<td>ekpi (to help)</td>
<td>kpi (sour)</td>
</tr>
<tr>
<td>gb</td>
<td>tɔmbɔ (canoe)</td>
<td>'dugba (to catch)</td>
<td>gbi (to strike)</td>
</tr>
<tr>
<td>'b</td>
<td>'ba (home)</td>
<td>'be (home)</td>
<td>'bá (home)</td>
</tr>
<tr>
<td>'d</td>
<td>la'di (to cook)</td>
<td>fi'di (to cook)</td>
<td>'di (to speak)</td>
</tr>
<tr>
<td>ʊ</td>
<td>kiri' (charcoal)</td>
<td>tiŋgbɔ (rhinoceros)</td>
<td>ti'r (sweet)</td>
</tr>
<tr>
<td>ʁ</td>
<td>(boy)</td>
<td>(boy)</td>
<td>(boy)</td>
</tr>
</tbody>
</table>

1 See vocabulary in Fr. Ribero's Elementi di Lingua Ndogo.
2 Unless, of course, drum signalling has been borrowed from a di-tonal people and adapted to Zande. Note that the Banda drum signalling analysed by the author in the Western District is tonal.
3 Relatively rare in Mpur-Madi, however.
4 It is the normal sound in Zande, alternating with r and l. It is spelt r.
5 In Sere, where ɨ is mostly to be found.
6 Pronounced tinjapa.

Figure 34. LATIN SMALL LETTER V WITH CURL (Tucker, 1940, p. 65).
F. References


Ramírez Solís, Cornelio, et al. (in press). Diccionario tepehuano de Santa María Ocotán, Durango. (Versión Preliminar.) D.R. por el Instituto Lingüístico de Verano, A.C., México, D.F.


