Proposal to Encode Additional Phonetic Symbols in the UCS

**Date:** 2003-06-09  
**Author:** Peter Constable, SIL International  
**Address:** 7500 W. Camp Wisdom Rd.  
Dallas, TX 75236  
USA  
**Tel:** +1 972 708 7485  
**Email:** peter_constable@sil.org

### A. Administrative

1. **Title**  
   Proposal to Encode Additional Phonetic Symbols in the UCS

2. **Requester's name**  
   SIL International (contact: Peter Constable)

3. **Requester type**  
   Expert contribution

4. **Submission date**  
   2003-06-09

5. **Requester's reference**  

6a. **Completion**  
   This is a complete proposal

6b. **More information to be provided?**  
   Only as required for clarification.

### B. Technical—General

<table>
<thead>
<tr>
<th>1a. New Script? Name?</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b. Addition of characters to existing block? Name?</td>
<td>Yes — Phonetic Extensions</td>
</tr>
<tr>
<td>2. Number of characters in proposal</td>
<td>15</td>
</tr>
<tr>
<td>3. Proposed category</td>
<td>A</td>
</tr>
<tr>
<td>4. Proposed level of implementation and rationale</td>
<td>3 (some combining marks)</td>
</tr>
<tr>
<td>5a. Character names included in proposal?</td>
<td>Yes</td>
</tr>
<tr>
<td>5b. Character names in accordance with guidelines?</td>
<td>Yes</td>
</tr>
<tr>
<td>5c. Character shapes reviewable?</td>
<td>Yes</td>
</tr>
<tr>
<td>6a. Who will provide computerized font?</td>
<td>SIL International</td>
</tr>
<tr>
<td>6b. Font currently available?</td>
<td>Yes</td>
</tr>
<tr>
<td>6c. Font format?</td>
<td>TrueType</td>
</tr>
<tr>
<td>7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?</td>
<td>Yes</td>
</tr>
</tbody>
</table>
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?  Yes

8. Does the proposal address other aspects of character data processing?  Yes, suggested character properties are included (see section E).

---

### C. Technical—Justification

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Has this proposal for addition of character(s) been submitted before?</td>
<td>No</td>
</tr>
<tr>
<td>2a. Has contact been made to members of the user community?</td>
<td>Yes</td>
</tr>
<tr>
<td>2b. With whom?</td>
<td>Linguists</td>
</tr>
<tr>
<td>3. Information on the user community for the proposed characters is included?</td>
<td>Linguists</td>
</tr>
<tr>
<td>4. The context of use for the proposed characters</td>
<td>Linguistics text books, linguistic descriptions (books, journal publications, etc.); dictionaries.</td>
</tr>
<tr>
<td>5. Are the proposed characters in current use by the user community?</td>
<td>Yes</td>
</tr>
<tr>
<td>6a. Must the proposed characters be entirely in the BMP?</td>
<td>Preferably</td>
</tr>
<tr>
<td>6b. Rationale?</td>
<td>If possible, should be kept with other phonetic symbols in the BMP.</td>
</tr>
<tr>
<td>7. Should the proposed characters be kept together in a contiguous range?</td>
<td>Preferably together with other phonetic symbols</td>
</tr>
<tr>
<td>8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?</td>
<td>The character LATIN SMALL LETTER C WITH STROKE might possibly be conceived of as being represented by the sequence (&lt; U+0063, U+0338 &gt;).</td>
</tr>
<tr>
<td>8b. Rationale for inclusion?</td>
<td>We consider the use of the overlay character U+0338 for representing such abstract characters unacceptable. For further discussion, see §F.1.</td>
</tr>
<tr>
<td>9a. Can any of the proposed characters be considered to be similar (in appearance or function) to an existing character?</td>
<td>The character LATIN SMALL LETTER C WITH STROKE is similar in appearance to U+00A2 CENT SIGN.</td>
</tr>
<tr>
<td>9b. Rationale for inclusion?</td>
<td>Distinct characters (see the discussion in §F.1).</td>
</tr>
<tr>
<td>10. Does the proposal include the use of combining characters and/or use of composite sequences?</td>
<td>No.</td>
</tr>
<tr>
<td>11. Does the proposal contain characters with any special properties?</td>
<td>No.</td>
</tr>
</tbody>
</table>
D. SC2/WG2 Administrative

1. Relevant SC2/WG2 document numbers
2. Status (list of meeting number and corresponding action or disposition)
3. Additional contact to user communities, liaison organizations, etc.
4. Assigned category and assigned priority/time frame

Other comments

E. Proposed Characters

A code chart and list of character names are shown on a new page.
### E.1 Code Chart

<table>
<thead>
<tr>
<th>Code</th>
<th>Character</th>
</tr>
</thead>
<tbody>
<tr>
<td>xx00</td>
<td>LATIN SMALL LETTER C WITH STROKE</td>
</tr>
<tr>
<td>xx01</td>
<td>LATIN SMALL LETTER D WITH HOOK AND TAIL</td>
</tr>
<tr>
<td>xx02</td>
<td>LATIN SMALL LETTER DB DIGRAPH</td>
</tr>
<tr>
<td>xx03</td>
<td>LATIN SMALL CAPITAL LETTER I WITH STROKE</td>
</tr>
<tr>
<td>xx04</td>
<td>LATIN SMALL LETTER P WITH STROKE</td>
</tr>
<tr>
<td>xx05</td>
<td>LATIN SMALL LETTER QP DIGRAPH</td>
</tr>
<tr>
<td>xx06</td>
<td>LATIN SMALL LETTER S WITH SWASH TAIL</td>
</tr>
<tr>
<td>xx07</td>
<td>LATIN SMALL LETTER ESH WITH RETROFLEX HOOK</td>
</tr>
<tr>
<td>xx08</td>
<td>LATIN SMALL CAPITAL LETTER U WITH STROKE</td>
</tr>
<tr>
<td>xx09</td>
<td>LATIN SMALL LETTER UPSILON WITH STROKE</td>
</tr>
<tr>
<td>xx0A</td>
<td>LATIN SMALL LETTER Z WITH SWASH TAIL</td>
</tr>
<tr>
<td>xx0B</td>
<td>LATIN SMALL LETTER EZH WITH RETROFLEX HOOK</td>
</tr>
<tr>
<td>xx0C</td>
<td>LATIN LETTER SMALL CAPITAL I OVER SMALL SCHWA</td>
</tr>
<tr>
<td>xx0D</td>
<td>LATIN LETTER SMALL UPSILON OVER SMALL SCHWA</td>
</tr>
<tr>
<td>xx0E</td>
<td>COMBINING SNAKE BELOW</td>
</tr>
</tbody>
</table>

### E.2 Character Names

- xx00: LATIN SMALL LETTER C WITH STROKE
- xx01: LATIN SMALL LETTER D WITH HOOK AND TAIL
- xx02: LATIN SMALL LETTER DB DIGRAPH
- xx03: LATIN SMALL CAPITAL LETTER I WITH STROKE
- xx04: LATIN SMALL LETTER P WITH STROKE
- xx05: LATIN SMALL LETTER QP DIGRAPH
- xx06: LATIN SMALL LETTER S WITH SWASH TAIL
- xx07: LATIN SMALL LETTER ESH WITH RETROFLEX HOOK
- xx08: LATIN SMALL CAPITAL LETTER U WITH STROKE
- xx09: LATIN SMALL LETTER UPSILON WITH STROKE
- xx0A: LATIN SMALL LETTER Z WITH SWASH TAIL
- xx0B: LATIN SMALL LETTER EZH WITH RETROFLEX HOOK
- xx0C: LATIN LETTER SMALL CAPITAL I OVER SMALL SCHWA
- xx0D: LATIN LETTER SMALL UPSILON OVER SMALL SCHWA
- xx0E: COMBINING SNAKE BELOW
E.3 Unicode Character Properties

The character COMBINING SNAKE BELOW 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+0323 COMBINING DOT BELOW.

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.

F. Other Information

F.1 LATIN SMALL LETTER C WITH STROKE

The character LATIN SMALL LETTER C WITH STROKE is often used to represent a voiceless alveolar affricate, particularly by Americanist linguists.

Figure 1. From Brody (1986), p. 261.

2.2. in Kekchi and Pokomchi-Pokomam.

The second case of documented sound change I will consider involves the change of Proto-Quichean to in both Kekchi and Pokomam-Pokomchi.

Figure 2. From Campbell (1976), p. 124.

Modern K’iche’. Several hundred years later, in modern Totonicapan K’iche’, the ABS1SG in has spread to all parts of the paradigm except the POSSESSIVE; see figure 6.

• SUBJECT of INTRANSITIVE: s-in-kam-ik COMPL-ABS1SG-die-AFF.INTR ‘I died’
• OBJECT of INTRANSITIVE: s-in-a-fel-o COMPL-ABS1-ERG2SG-see-AFF.TR ‘you saw me’

Figure 3. From Robertson (1999), p. 457.

Note that this character has similar appearance to one of the glyph variants of U+00A2 CENT SIGN. That character has other glyph variants, however, such as “¢”, that are not acceptable for phonetic transcription. Moreover, the character properties of U+00A2 (e.g. General Category Sc) are not what are needed for phonetic characters.

Also, question 8a of section C above asks whether these characters can be considered presentation forms of existing character or character sequences. As mentioned, the LATIN SMALL LETTER C WITH STROKE might be conceived as being represented as a sequence involving the overlay character U+0338 COMBINING LONG SOLIDUS OVERLAY. I suggest, however, that this would be inappropriate and is irrelevant. Apart from certain mathematical operators that decompose into sequences using this overlay character, there is a clear precedent for Latin characters not to represent characters such as LATIN SMALL LETTER C WITH STROKE using sequences involving U+0338: there are several Latin characters with stroke encoded in the UCS, but none of them has a decomposition involving U+0338.
Therefore, insofar as existing characters with overlaid stroke are not considered presentation forms of existing sequences, it is suggested that the LATIN SMALL LETTER C WITH STROKE is likewise not to be considered a presentation form of some existing sequence.

F.2 LATIN SMALL LETTER D WITH HOOK AND TAIL

The character LATIN SMALL LETTER D WITH HOOK AND TAIL is not explicitly IPA-approved, but it is consistent with IPA conventions and is listed in the IPA Handbook (IPA 1999). It is used to represent a voiced retroflex implosive, a speech sound that is rare but that is attested in at least the Parkari language (Hoyle 2001).

F.3 The characters LATIN SMALL LETTER DB DIGRAPH and LATIN SMALL LETTER QP DIGRAPH

These characters are used to represent labiodental stops, which are known to occur in some Bantu languages. These characters have been used primarily by Africanists in language descriptions, but are also attested in general works on phonetics and phonology.

The subject of a Parkari Nonfinal form is normally the same as the subject of the main verb. However, there are exceptions, e.g.

“My farmwork”

11. pur-e pare on poh u-a m wale sad-o-h ponji.
fill in nonfinite result nonfinite and then that-G in again leave-P-pres water
After filling them in then I let the water into it.

The Affricates:

There are eleven affricate combinations in Zulu, of which six are ejective, i.e., accompanied by a simultaneous closure of the glottis, four are voiced and one radical.

mf (denti-labial affricates). These, which are phonetically !mf! and !mb!, are only found in conjunction with the denti-labial nasal.
F.4 The characters LATIN SMALL CAPITAL LETTER I WITH STROKE, LATIN SMALL CAPITAL LETTER U WITH STROKE and LATIN SMALL LETTER UPSILON WITH STROKE

The characters LATIN SMALL CAPITAL LETTER I WITH STROKE and LATIN SMALL CAPITAL LETTER U WITH STROKE are used by some Americanists to represent central lower-high vowels:

![American Usage Vowel Symbols]

Figure 10. From Pullum and Ladusaw (1996), p. 298.

The barred small capital I is also used in some recent Oxford dictionaries (though with a different meaning), as is the barred upsilon:

![in addition to these transcriptions of recent developments in RP, the two composite symbols, ī and ȳ, are used to represent /ɪ/ or /ʊ/ and /u/ or /u/ respectively (see p. x above and the discussion of vowel reduction below, p. xvii). The fol-]

Figure 12. From Upton et al (2003).
F.5 LATIN SMALL LETTER P WITH STROKE

In the Americanist tradition, barred stop symbols are often used to represent fricatives, with barred-p representing a voiceless bilabial fricative.

There are a series of fricatives.

of other languages (cf. Parker 1994). Sérgio Meira (personal communication [henceforth p.c.]) adds that [h] can also affect its own environment as well. Thus in some dialects of Tiriyo, /p/ can be realized as [p] following [h] (/pihpa/ ‘skin’ → [pihpa] ~ [pi:pa]), while a phonemic /k/ can be realized as [h] following a syllable-final [h]: /pahko/ ‘father’ → [pa:ko] ~ [pahho] ~
F.6  The characters LATIN SMALL LETTER S WITH SWASH TAIL and LATIN SMALL LETTER Z WITH SWASH TAIL

These characters have been used by Africanists to represent labialized alveolar fricatives. It should be noted that these are not glyph variants of s-retroflex hook and z-retroflex hook.

Figure 19. From IPA (1949), p. 14.

Figure 20. S/z-swash tail, distinct from retroflex-hook forms; from Doke(1967), p. 30.

Figure 21. Z-swash tail (red highlight) in contrast with z-retroflex hook (blue highlight); from Tucker (1971), p. 648.
F.7 The characters LATIN SMALL LETTER ESH WITH RETROFLEX HOOK and LATIN SMALL LETTER EZH WITH RETROFLEX HOOK

These characters are intended to represent retroflex counterparts to the palato-alveolar fricatives esh “ʃ” and ezh “ʒ”. These symbols are not IPA-approved, and their appropriateness is uncertain since the sounds represented by esh and ezh are “usually regarded as having the blade of the tongue raised towards the hard palate,” a gesture that would “preclude tongue tip retroflexion” (Peter Ladefoged, personal communication). Nevertheless, these symbols are, in fact, used by some linguists:

*Hindi*

- [ɲntʰɔɾ] ‘continuously’
- [paɾɛ:ʃiŋ] ‘labour, hard work’
- [kʰurɔɾ] ‘reason (for doing something)’
- [bʱɔli] ‘a (particular) festival’

Figure 22. From Laver (1994), p. 559.

*Norwegian (Southeastern, Larvik)*

- [vɔːfɔɾkrɑŋ] ‘overalls’
- [hɔɾfɔɾɡɔɾ] ‘hair colours’
- [sʰydɔr̝tʰɔp] ‘as thin as an awl’
- [ɔɾɔɾɔ?] ‘sweet and sour’

Figure 23. From Laver (1994), p. 560.

F.8 The characters LATIN LETTER SMALL CAPITAL I OVER SMALL SCHWA and LATIN LETTER SMALL UPSILON OVER SMALL SCHWA

These characters are used in the *Longman Dictionary of Contemporary English* and derivative titles.

**ca·pit·u·late** /ˈkæpəˌlət/ n [L] 1 formal to accept or agree to something that you have been opposing for a long time; 2 give in: Helen finally capitulated and let her son have a car. 2 formal to accept defeat by your enemies in a war; 3 *surrender* — *capitulation*

**cap·let** /ˈkæpˌlɛt/ n [C] a small smooth pill (= solid piece of medicine) with a shape that is slightly longer and narrower than a tablet (= a small round pill)

Figure 24. From Longman Publishing (2003), p. 217.

Note that the meaning assigned to these symbols is one of alternation between two pronunciations:

- /ɔː/ shows stress shift
- /aː/ means that some speakers use /u/ and some use /a/ 
- /a/ means that /a/ may or may not be used

Figure 25. From Longman Publishing (2003).

In principle, these characters could be seen as combining two symbols that might in general be arbitrarily chosen; in other words, there is a theoretical potential for a very large number of such paired-value characters. That might be
taken to suggest that a different approach (e.g. involving markup) may be in order. On the other hand, there are not a large number of such characters in use; there are only these two in the Longman dictionaries, and no others that I know of.

### F.9 COMBINING SNAKE BELOW

The COMBINING SNAKE BELOW is used by some in the Americanist tradition to indicate lenis (weak) articulation.

![Example of COMBINING SNAKE BELOW](image)

**Figure 26.** From Floyd (1981), p. 117.

| ḏ | fortis consonant |
| ḗ | lenis consonant |
| ḕ | extralenis consonant |

**Figure 27.** From Mills (1984), p. xxii.

| [ši, tšin] | for /šiN/ | 'ear' |
| [čan] | for /čan/ | 'yellow' |
| [bələ.la.səˈrən] | for /bələ.la.səˈrən/ | 'a hoe' |

**Figure 28.** From Lengyel (1991), p. 343.

---

### G. References


Mills, Elizabeth. 1984. *Senoufo phonology, discourse to syllable (a prosodic approach).* (Summer Institute of Linguistics publications in linguistics, 72.) Dallas: Summer Institute of Linguistics and University of Texas at Arlington.


