

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

- | | | |
|-----|---|---------------------------|
| 1a. | New Script? Name? | No |
| 1b. | Addition of characters to existing block? Name? | Yes — Phonetic Extensions |
| 2. | Number of characters in proposal | 15 |
| 3. | Proposed category | A |
| 4. | Proposed level of implementation and rationale | 3 (some combining marks) |
| 5a. | Character names included in proposal? | Yes |
| 5b. | Character names in accordance with guidelines? | Yes |
| 5c. | Character shapes reviewable? | Yes |
| 6a. | Who will provide computerized font? | SIL International |
| 6b. | Font currently available? | Yes |
| 6c. | Font format? | TrueType |
| 7a. | Are references (to other character sets, dictionaries, descriptive texts, etc.) provided? | Yes |

- | | | |
|-----|---|---|
| 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

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

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

	xx0
0	ç
1	ç̣
2	ɔ̥
3	ɪ̥
4	ɸ
5	ɸ̣
6	ʒ̣
7	ʃ̣
8	ʊ̥
9	ʊ̥̄
A	ʒ̣̄
B	ʒ̣̄̄
C	ɪ̥̄̄
D	ʃ̣̄̄
E	◌̚
F	

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.

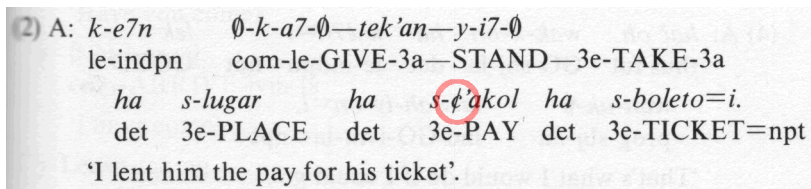


Figure 1. From Brody (1986), p. 261.

2.2. /ç/ in Kekch' and Pokomch'-Pokomam.

The second case of documented sound change I will consider involves the change of Proto-Quichean *ç to ç in both Kekch' and Pokomam-Pokomch'.

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

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

- SUBJECT OF INTRANSITIVE: *š-in-kam-ik* COMPL-ABS1SG-die-AFF.INTR 'I died'
- OBJECT OF TRANSITIVE: *š-in-a-ç'et-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 “Ç”, 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).

◌ɖ	Hooktop right-tail D	Voiced retroflex implosive <i>Not explicitly IPA approved</i>	219	-----	E219
----	----------------------	--	-----	-------	------

Figure 4. From IPA (1999), p. 179.

Labial	Dental	Dent/Al	Alveol	Post-al	R'flex	Palatal	Velar	Uvular	Total
[ɸ]	ɖ	ɖ ~ ɖ	ɖ	ɖ	◌ɖ	ʃ	ʒ	ʁ]	

Figure 5. From Laver (1994), p. 582.

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 pər-e ən pəh u-a m wəle **sɑɖ-õ-h** pəŋi.
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.

Figure 6. From Hoyle (2001), p. 254.

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 character have been used primarily by Africanists in language descriptions, but are also attested in general works on phonetics and phonology.

§25.—*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 phon-
mv } etically **mpf** and **mpv**, are only found in conjunction with the denti-labial nasal.

Figure 7. From Doke (1950), p. 17.

21. The languages of this zone are notable for the different voiced labial sounds that occur in them. In MANDA (11) there is a labio-dental semi-vowel, e.g. **-vik-** ‘put’, where the first consonant appears to be distinct from **-w-**. In TODGA (15) there is a labio-dental plosive which is distinct from the bilabial plosive, e.g. **-ɖar-** ‘shine’, **-bar-** ‘give birth to’. In POKA (21b) there is a ‘v’ without friction, which is

Figure 8. From Guthrie (1967), p. 61.

VOICELESS UNASPIRATED	papa 'cloud'	φu 'finished'	tiŋɸfuβu 'hippos'	ŋfutsu 'tortoise'
VOICELESS ASPIRATED	pʰapʰatani 'butterfly'		ŋɸfʰuka 'distance'	
VOICED	kuba 'to hit'	kuβaβa 'to be painful'	ʃileɸvu 'chin'	kuvumba 'to guess'
BREATHY VOICED	jimbʰo 'ostrich'		ŋɸvʰuβu 'tree (sp.)'	kuvʰeɸa 'to scratch'

Figure 9. From Ladefoged and Maddieson (1996), p. 18.

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 vocoids:

American Usage Vowel Symbols

		Front		Central		Back	
		Unround	Round	Unround	Round	Unround	Round
High	(Higher)	i	ü	ɨ	ɥ	ɨ̈	u
	Lower	ɪ	ʊ	ɨ̄	ɥ̄	ɨ̄̈	U
Mid	Higher	e	ö	ə		ë	o
	Lower	ɛ	ɔ̈	ʌ		ɔ	
Low		æ	a/ɑ				

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

ɨ̄	SS <u>good</u> , <u>stood</u> ; NZ, SA <u>sit</u>
ɨ̄̈	SS <u>lip</u>
ɥ̄̈	v ^ə] <u>fluid</u> , <u>ruin</u>

Figure 11. From Bailey (1985), p. xxiii.

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, **ɨ̄** and **ɥ̄**, are used to represent [ɪ] or [ə] and [ʊ] or [ɔ̈] respectively (see p. x above and the discussion of vowel reduction below, p. xvii). The fol-

Figure 12. From Upton et al (2003).

beautiful
 BR 'bjʊtɪf(ə)l
 AM 'bjudəfəl
beautifully
 BR 'bjʊtɪf(ə)li,
 'bjʊtɪflɪ
 AM 'bjudəf(ə)li

Figure 13. 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.

Fricatives:				
Voiceless	ɸ	f	θ	ɬ
Voiced	ɸ̣	v	ð	ʒ

Figure 14. From Brewster and Brewster (1976), p. 279.

ɸ = voiceless bilabial fricative
 ɸ̣ = voiced bilabial fricative

Figure 15. From Campbell (1977), p. 4.

FRICATIVES			
Vl. flat	ɸ	f	θ
Vd. flat	ɸ̣	v	ð

Figure 16. From Smalley (1989), p. 454.

There are a series of fricatives.

f	ɸ	[wa³li³ pa³rɪn²su²]	'wild manioc'
s	[s]	[a²su³su²]	'bone'
h	[h]	[hot³su²]	'monkey'

Figure 17. From Kroeker (2001), p. 78.

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 [ɸ] following [h] (/pihpə/ 'skin' → [pihpə] ~ [pi:pə]), while a phonemic /k/ can be realized as [h] following a syllable-final [h]: /pahko/ 'father' → [pa:ko] ~ [pahho] ~

Figure 18. From Parker (2001), p. 109.

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.

Labialized θ , δ or s , z : σ , ς (ξ , ζ) of the orthography for Shona recommended by Professor Doke). One may, however, represent these sounds by digraphs such as sf , zv or sq , zq , in

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

	Bilabial	Denti-labial	Dental	Alveolar	Retroflex	Labio-alveolar	Prepalatal	Palatal	Prevelar	Velar	Glottal
EXPLOSIVE	rad. p eject. p' asp. ph voic. b		t' tʰ d	t t' th d	t' tʰ ɖ			c' ch ɟ		k k' kh g	
IMPLOSIVE	ɓ			ɗ							
NASAL	contin. m syll. m̃	(m̃)	n ñ	n ñ	ɳ	(ɳ)	ɲ ɲ̃	(ɲ)		ŋ ŋ̃	
ROLLED	contin. syll. r			r r̃	ɽ						
FLAPPED	voic. ɾ			ɽ							
LATERAL	contin. syll. l		ɭ	l l̃	ɭ						
FRICATIVE	med. unv. ɸ voic. β	f v		s z		ɬ ɮ	ʃ ʒ		h	x ɣ	ħ ʕ
FRICATIVE	lat. unv. ɸ voic. β			ɬ ɮ							
AFFRICATE	med. rad. pf eject. pɸ' asp. pɸh voic. bv dɸv			ts ts' tsh dz		tɬ dɮ	tʃ tʃ' tʃh dʒ			kx' kxh	
AFFRICATE	lat. eject. tɬ' asp. tɬh voic. dɮ			tɬ' tɬh dɮ						(kɬ') kɬh	
SEMI-VOWEL	(w)	β̃					j			w	

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

Alveolar voiced

fricative	z	z	z	z	z
[retroflex		ɮ	ɮ	ɮ	z]
palatalized	zy	z̟	z̠	z̡	
labialized	z̠			z̠	zv, zw, z̠
[velarized		z̠, θ̠	z̠	z̠	z̠]

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 “j” and ezh “z”. 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
 [nɛɾɔ̃ntʰɜɾ.] ‘continuously’
 [paɾɛjːɟ̃m] ‘labour, hard work’
 [kʰɔɾɟ̃tʰ] ‘reason (for doing something)’
 [hʷɔ̃li] ‘a (particular) festival’

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

Norwegian (Southeastern, Larvik)
 [vɑ̃tɛzkrɑ̃ŋ] ‘overalls’
 [hʷɔ̃rfaɾvɛɾ] ‘hair colours’
 [sʷyːlʷtʷɛ̃n] ‘as thin as an awl’
 [sʷɟ̃lʷɔ̃tʷ] ‘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 /kəˈpɪtʃⁱleɪt/ v [l] **1** formal to accept or agree to something that you have been opposing for a long time; **give in**: *Helen finally capitulated and let her son have a car. **2** formal to accept defeat by your enemies in a war; **surrender** — **capitulation** /kəˈpɪtʃⁱleɪʃən/ n [C,U]
cap·let /kæpⁱlɪt/ 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:

/◀/ shows stress shift
 /ɪ̯/ means that some speakers use /ɪ/ and some use /ə/
 /ʊ̯/ means that some speakers use /ʊ/ and some use /ə/
 /ə/ means that /ə/ 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.

/p _̄ ada/ (no aspiration)	'grind'
/p _̄ ada/ (slight aspiration)	'ocean, sea'
/p ^h ada/(much aspiration)	'digging'

/p/ (fortis) is produced with tight lip closure and no aspiration.
 /p_̄/ (lenis) is produced with a slightly looser lip closure and is released with a minimum of aspiration, yet it is much weaker than fully aspirated /p^h/.

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

(b)	fortis consonant
(b _̄)	lenis consonant
(b _̄ ^h)	extralenis consonant

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

[.ʃi.'tʃiŋ.]	for /ʃičin/	'ear'
[.kʷaŋ.]	for /q'an/	'yellow'
[.b̄al.la.sa.'roŋ.]	for /waʔl ʔasaro'n/	'a hoe'

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

G. References

- Bailey, Charles-James N. 1985. *English phonetic transcription*. (Summer Institute of Linguistics Publications in Linguistics, 74.) Dallas: Summer Institute of Linguistics and University of Texas at Arlington.
- Brewster, E. Thomas, and Elizabeth S. Brewster. 1976. *Language acquisition made practical: Field methods for language learners*. Colorado Springs, CO: Lingua House.
- Brody, Jill. 1986. "Repetition as a rhetorical and conversational device in Tojolobal (Mayan)." *International Journal of American Linguistics* 52.255-74.
- Campbell, Lyle. 1977. *Quichean linguistic prehistory*. (University of California publications in linguistics, 81.) Berkeley, CA: University of California Press.
- Clark, John, and Colin Yallop. 1995. *An introduction to phonetics and phonology, 2nd edn*. (Blackwell textbooks in linguistics.) Oxford: Blackwell.
- Doke, Clement M. 1950. *Text-book of Zulu grammar*. London: Longmans, Green & Co.
- Floyd, Rick. 1981. *Manual for articulatory phonetics*. Dallas: Summer Institute of Linguistics
- Guthrie, Malcolm. 1967. *The classification of the Bantu languages*. London: International African Institute.

- Hoyle, Richard A. 2001. *Scenarios, discourse and translation: The scenario theory of Cognitive Linguistics, its relevance for analysing New Testament Greek and modern Parkari texts, and its implications for translation theory*. University of Surrey Roehampton PhD thesis.
- International Phonetic Association. 1949. *The principles of the International Phonetic Association*. London: International Phonetics Association.
- . 1975. "The Association's alphabet." *Journal of the International Phonetic Association* 5:52–58.
- . 1999. *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Kroeker, Menno. 2001. "A descriptive grammar of Nambikuara." *International Journal of American Linguistics* 67.1–87.
- Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell Publishers.
- Lass, Roger. 1984. *Phonology: an introduction to basic concepts*. (Cambridge textbooks in linguistics.) Cambridge: Cambridge University Press.
- Laver, John. 1994. *Principles of phonetics*. (Cambridge textbooks in linguistics.) Cambridge: Cambridge University Press.
- Lengyel, Thomas E. 1991. "Toward a dialectology of Ixil Maya: variation across communities and individuals." *International Journal of American Linguistics* 57.330–64.
- Longman Publishing. 2003. *Longman dictionary of contemporary English*.
- 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.
- Parker, Steve. 2001. "On the phonemic status of [h] in Tiriyo." *International Journal of American Linguistics* 67.105–18.
- Pullum, Geoffrey K., and William A. Ladusaw. 1996. *Phonetic symbol guide, 2nd edn*. Chicago: University of Chicago Press.
- Robertson, John S. 1999. "The history of first-person singular in the Mayan languages." *International Journal of American Linguistics* 65.449–65.
- Tucker, A.N. 1971. "Orthographic systems and conventions in Sub-Saharan Africa." *Current trends in linguistics, volume 7: Linguistics in Sub-Saharan Africa*, ed. by Thomas A. Sebeok, 618–53. The Hague: Mouton.
- Upton, Clive; William Kretschmar; and Rafal Konopka. 2003. *The Oxford Dictionary of Pronunciation for Current English*. Oxford: Oxford University Press.