

# Revised Proposal to Encode Additional Phonetic Symbols in the UCS

*Date:* 2004-02-01

*Author:* Peter Constable, Microsoft

*Address:* One Microsoft Way  
Redmond, WA 98052  
USA

*Tel:* +1 425 722 1867

*Email:* petercon@microsoft.com

---

## A. Administrative

---

- |                                      |   |
|--------------------------------------|---|
| 1. Title                             | Revised Proposal to Encode Additional Phonetic Symbols in the UCS |
| 2. Requester's name                  | SIL International (contact: Jonathan Kew), Peter Constable        |
| 3. Requester type                    | Expert contribution   |
| 4. Submission date                   | 2004-02-01  |
| 5. Requester's reference             | L2/03-190   |
| 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                 | 9                         |
| 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?      | An earlier proposal (L2/03-190) was reviewed by UTC (meeting #95). That proposal contained consonant symbols with retroflex hook, and it was asked that these be moved to a different proposal; this has been done. Concern was expressed over two text elements that were felt to be best represented as rich text; these have been removed from this proposal. Concern was also expressed over two symbols as it was not adequately demonstrated that they were not presentation forms of existing characters; these also have been removed from the proposal. One new character (iota-stroke) has been added. |
| 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; see §F.1.  |
| 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	ɔ̃
2	ɪ
3	ɹ
4	ɸ
5	ɸ
6	ʊ
7	ʊ
8	ʊ̥
9	
A	
B	
C	
D	
E	
F	

## E.2 Character Names

xx00	LATIN SMALL LETTER C WITH STROKE
xx01	LATIN SMALL LETTER DB DIGRAPH
xx02	LATIN SMALL CAPITAL LETTER I WITH STROKE
xx03	LATIN SMALL LETTER IOTA WITH STROKE
xx04	LATIN SMALL LETTER P WITH STROKE
xx05	LATIN SMALL LETTER QP DIGRAPH
xx06	LATIN SMALL CAPITAL LETTER U WITH STROKE
xx07	LATIN SMALL LETTER UPSILON WITH STROKE
xx08	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.

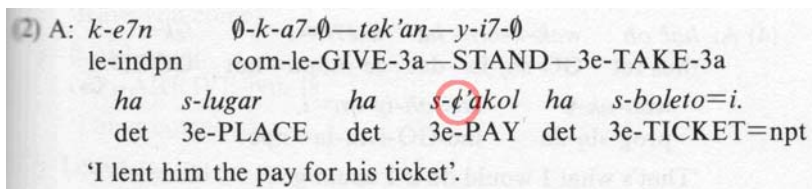


Figure 1. From Brody (1986), p. 261.

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

The second case of documented sound change I will consider involves the change of Proto-Quichean \*ɕ to s 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-ɕ'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 “ɕ”, 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, particularly given that phonetic symbols are often adopted for orthographic uses.

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:

00D8	Ø	LATIN CAPITAL LETTER O WITH STROKE
00F8	ø	LATIN SMALL LETTER O WITH STROKE
0141	Ł	LATIN CAPITAL LETTER L WITH STROKE
0142	ł	LATIN SMALL LETTER L WITH STROKE
019B	λ	LATIN SMALL LETTER LAMBDA WITH STROKE
01FE	Ŏ	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE
01FF	ŏ	LATIN SMALL LETTER O WITH STROKE AND ACUTE
1D0C	Ł	LATIN LETTER SMALL CAPITAL L WITH STROKE
1D13	Ϸ	LATIN SMALL LETTER SIDEWAYS O WITH STROKE

Table 1. Latin characters in the UCS with diagonal stroke but no decomposition to sequences with U+0338

Therefore, insofar as existing characters with overlaid stroke are not considered presentation forms of existing sequences, LATIN SMALL LETTER C WITH STROKE likewise should not be considered a presentation form of some existing sequence.

## F.2 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.

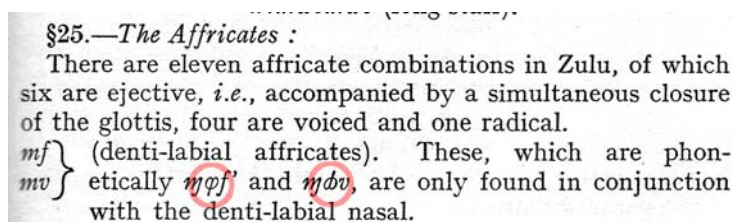


Figure 4. From Doke (1950), p. 17.

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. **-ɓar-** 'shine', **-bar-** 'give birth to'. In POKA (21b) there is a 'v' without friction, which is

Figure 5. From Guthrie (1967), p. 61.

VOICELESS UNASPIRATED	<b>papa</b> 'cloud'	<b>ɸu</b> 'finished'	<b>tiŋɸfuβu</b> 'hippos'	<b>ɲfutsu</b> 'tortoise'
VOICELESS ASPIRATED	<b>pʰapʰaɲani</b> 'butterfly'		<b>ŋɸfʰuka</b> 'distance'	
VOICED	<b>kuba</b> 'to hit'	<b>kuβaβa</b> 'to be painful'	<b>ʃileɸvu</b> 'chin'	<b>kuvumba</b> 'to guess'
BREATHY VOICED	<b>jimbʰo</b> 'ostrich'		<b>ŋɸvʰuβu</b> 'tree (sp.)'	<b>kuvʰɛɲa</b> 'to scratch'

Figure 6. From Ladefoged and Maddieson (1996), p. 18.

### F.3 The character LATIN SMALL LETTER IOTA WITH STROKE

This character is used by Slavic linguists in descriptions of Russian:

15.9 Before unstressed **ɨ** the velarization effect is not very strong and may conveniently be ignored for present purposes. It should be noted that the word **был**, quoted as an example above, is very often unstressed and pronounced **bɨl**.

Figure 7. From Jones and Ward (1969), p. 81.

I. VOWELS AND DIPHTHONGS	
The i-phoneme: <b>i</b>	I2
<b>ɨ</b>	I3
<b>i</b> contrasted with <b>ɨ</b>	I4
the unstressed vowel <b>ɨ</b>	I4, I5
the unstressed vowel <b>ɨ</b>	I6, I7
<b>ɨ</b> contrasted with <b>ɨ</b>	I7

Figure 8. From Ward (1966), p. 3.

The following symbols from the IPA are used in the Introduction for the phonetic transcription of Russian words.

### Vowels

- i as in ил [il]
- ɨ as in пыл [pɨl]
- ɪ as the first vowel in игла [ɪ'gla]
- ɨ̞ as the first vowel in дыра [dɨ̞'ra]

Figure 9. From Wade (2000), p. 2.

#### 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	ʉ	ɨ̞	u
	Lower	ɪ	ʊ	ɨ̞	ʉ̞	ɨ̞̞	ʊ̞
Mid	Higher	e	ö	ə		ɘ	o
	Lower	ɛ	ɔ̥	ʌ		ɔ	
Low		æ		a/ɑ			

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

- ɨ̞ SS good, stood; NZ, SA sit
- ɨ̞̞] SS lip
- ʉ̞̞] fluid, ruin

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, **[ɪ]** and **[ə]**, are used to represent [ɪ] or [ə] and [ʊ] or [ə] respectively (see p. x above and the discussion of vowel reduction below, p. xvii). The following 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	<b>[p̥]</b>	f	θ	s
Voiced	<b>[b̥]</b>	v	ð	z

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

**[p̥]** = voiceless bilabial fricative  
**[b̥]** = voiced bilabial fricative

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

FRICATIVES			
Vl. flat	<b>[p̥]</b>	f	θ
Vd. flat	<b>[b̥]</b>	v	ð

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

There are a series of fricatives.

f	<b>[p̥]</b>	[wa <sup>3</sup> li <sup>3</sup> pa <sup>3</sup> rɪn <sup>2</sup> su <sup>2</sup> ]	'wild manioc'
s	[s]	[a <sup>2</sup> su <sup>3</sup> su <sup>2</sup> ]	'bone'
h	[h]	[hot <sup>3</sup> su <sup>2</sup> ]	'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 [p̥] following [h] (/pihpə/ 'skin' → [pihp̥ə] ~ [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 COMBINING SNAKE BELOW

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

/p̥ada/ (no aspiration)	'grind'
/p̥ada/ (slight aspiration)	'ocean, sea'
/p <sup>h</sup> ada/ (much aspiration)	'digging'

/p/ (fortis) is produced with tight lip closure and no aspiration.  
 /p̥/ (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<sup>h</sup>/.

Figure 19. From Floyd (1981), p. 117.

(b)	fortis consonant
(b̥)	lenis consonant
(b <sup>h</sup> )	extralenis consonant

Figure 20. From Mills (1984), p. xxii.

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

Figure 21. 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.

- 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.
- Jones, Daniel, and Dennis Ward. 1969. *The phonetics of Russian*. 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.
- Lengyel, Thomas E. 1991. "Toward a dialectology of Ixil Maya: variation across communities and individuals." *International Journal of American Linguistics* 57.330–64.
- 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.
- Upton, Clive; William Kretzschmar; and Rafal Konopka. 2003. *The Oxford Dictionary of Pronunciation for Current English*. Oxford: Oxford University Press.
- Wade, Terence. 2000. *A Comprehensive Russian Grammar, 2nd edn*. Oxford: Blackwell.
- Ward, Dennis. 1966. *Russian pronunciation illustrated*. Cambridge: Cambridge University Press.