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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Title</td>
<td>Revised Proposal to Encode Additional Phonetic Symbols in the UCS</td>
</tr>
<tr>
<td>2. Requester’s name</td>
<td>SIL International (contact: Jonathan Kew), Peter Constable</td>
</tr>
<tr>
<td>3. Requester type</td>
<td>Expert contribution</td>
</tr>
<tr>
<td>4. Submission date</td>
<td>2004-02-01</td>
</tr>
<tr>
<td>5. Requester’s reference</td>
<td>L2/03-190</td>
</tr>
<tr>
<td>6a. Completion</td>
<td>This is a complete proposal</td>
</tr>
<tr>
<td>6b. More information to be provided?</td>
<td>Only as required for clarification.</td>
</tr>
</tbody>
</table>

B. Technical—General

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. New Script? Name?</td>
<td>No</td>
</tr>
<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>9</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>
</tbody>
</table>
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.
<table>
<thead>
<tr>
<th></th>
<th>Character Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>LATIN SMALL LETTER C WITH STROKE</td>
</tr>
<tr>
<td>1</td>
<td>LATIN SMALL LETTER DB DIGRAPH</td>
</tr>
<tr>
<td>2</td>
<td>LATIN SMALL CAPITAL LETTER I WITH STROKE</td>
</tr>
<tr>
<td>3</td>
<td>LATIN SMALL LETTER IOTA WITH STROKE</td>
</tr>
<tr>
<td>4</td>
<td>LATIN SMALL LETTER P WITH STROKE</td>
</tr>
<tr>
<td>5</td>
<td>LATIN SMALL LETTER QP DIGRAPH</td>
</tr>
<tr>
<td>6</td>
<td>LATIN SMALL CAPITAL LETTER U WITH STROKE</td>
</tr>
<tr>
<td>7</td>
<td>LATIN SMALL LETTER UPSILON WITH STROKE</td>
</tr>
<tr>
<td>8</td>
<td>COMBINING SNAKE BELOW</td>
</tr>
</tbody>
</table>
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.

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, 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:

<table>
<thead>
<tr>
<th>DEC</th>
<th>HEX</th>
<th>LATIN CHARACTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>00D8</td>
<td>Ø</td>
<td>LATIN CAPITAL LETTER O WITH STROKE</td>
</tr>
<tr>
<td>00F8</td>
<td>ø</td>
<td>LATIN SMALL LETTER O WITH STROKE</td>
</tr>
<tr>
<td>0141</td>
<td>L</td>
<td>LATIN CAPITAL LETTER L WITH STROKE</td>
</tr>
<tr>
<td>0142</td>
<td>ł</td>
<td>LATIN SMALL LETTER L WITH STROKE</td>
</tr>
<tr>
<td>019B</td>
<td>Ł</td>
<td>LATIN SMALL LETTER LAMBDA WITH STROKE</td>
</tr>
<tr>
<td>01FE</td>
<td>Ô</td>
<td>LATIN CAPITAL LETTER O WITH STROKE AND ACUTE</td>
</tr>
<tr>
<td>01FF</td>
<td>Ï</td>
<td>LATIN SMALL LETTER O WITH STROKE AND ACUTE</td>
</tr>
<tr>
<td>1D0C</td>
<td>Ṉ</td>
<td>LATIN LETTER SMALL CAPITAL L WITH STROKE</td>
</tr>
<tr>
<td>1D13</td>
<td>Ṕ</td>
<td>LATIN SMALL LETTER SIDEWAYS O WITH STROKE</td>
</tr>
</tbody>
</table>

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. -vilk- ‘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. -d'ar- ‘shine’, -bar- ‘give birth to’. In POKA (21b) there is a ‘v’ without friction, which is Figure 5. From Guthrie (1967), p. 61.

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 т the vclariazion effect is not very strong and may conveniently be ignored for present purposes. It should be noted that the word дыр, quoted as an example above, is very often unstressed and pronounced бр.

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

### I. VOWELS AND DIPHTHONGS

The i-phoneme:  

<table>
<thead>
<tr>
<th>i</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>13</td>
</tr>
<tr>
<td>i contrasted with i</td>
<td>14</td>
</tr>
<tr>
<td>the unstressed vowel i</td>
<td>14, 15</td>
</tr>
<tr>
<td>the unstressed vowel ı</td>
<td>16, 17</td>
</tr>
<tr>
<td>ı contrasted with ı</td>
<td>17</td>
</tr>
</tbody>
</table>

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]  
i  as in пыл  [pil]  
t  as the first vowel in игл à  [t'gl'a]  
\[\text{Diacritical mark for extra-long vowel}

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:

\begin{itemize}
\item The barred small capital I is also used in some recent Oxford dictionaries (though with a different meaning), as is the barred upsilon:
\end{itemize}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{american_usage_vowel_symbols}
\caption{American Usage Vowel Symbols}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{american_usage_vowel_symbols}
\caption{American Usage Vowel Symbols}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{american_usage_vowel_symbols}
\caption{American Usage Vowel Symbols}
\end{figure}
In addition to these transcriptions of recent developments in RP, the two composite symbols, ɬ and ɮ, are used to represent [t] or [d] and [u] or [o] respectively (see p. xv above and the discussion of vowel reduction below, p. xvii). The fol-

Figure 12. From Upton et al (2003).

beautiful
BR 'bjuːtɪf(ə)l
AM 'bjudəf(ə)l
beautifully
BR 'bjuːtɪf(ə)li,
'bjʊtlɪ
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.

![Frictional Chart]

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

- **p** = voiceless bilabial fricative
- **b** = voiced bilabial fricative

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

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

There are a series of fricatives.

<table>
<thead>
<tr>
<th>Fricatives</th>
<th>ɬ</th>
<th>f</th>
<th>θ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vl. flat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vd. flat</td>
<td>h</td>
<td>v</td>
<td>d</td>
</tr>
</tbody>
</table>

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 Tiriyyo, /p/ can be realized as [ph] following [h] (/phipa/ ‘skin’ → [phipa] ~ [p:i:pa]), 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.

\[
\begin{align*}
/\text{p}a\text{d}a/ & \quad \text{(no aspiration)} & \text{’grind’} \\
/\text{p}a\text{d}a/ & \quad \text{(slight aspiration)} & \text{’ocean, sea’} \\
/\text{p}_{\text{h}}\text{a}\text{d}a/ & \quad \text{(much aspiration)} & \text{’digging’}
\end{align*}
\]

\(/\text{n}/ \quad \text{(fortis) is produced with tight lip closure and no aspiration.}

\(/\text{n}/ \quad \text{(lenis) is produced with a slightly looser lip closure and is}

\text{released with a minimum of aspiration, yet it is much weaker than}

\text{fully aspirated /p}^{/}\text{n}/.

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

\[
\begin{align*}
\text{(b)} & \quad \text{fortis consonant} \\
\text{(l)} & \quad \text{lenis consonant} \\
\text{(b)} & \quad \text{extralenis consonant}
\end{align*}
\]

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

\[
\begin{align*}
[\text{Si}’\text{tʃi}’\text{n}.] & \quad \text{for } /\text{ʃi}’\text{cin}/ \quad \text{’ear’} \\
[\text{kwa}’\text{n}.] & \quad \text{for } /\text{q}’\text{an}/ \quad \text{’yellow’} \\
[\text{b-}a\text{lai}’a\text{’a}’\text{roŋ}.] & \quad \text{for } /\text{wa}’\text{ɬi}’a\text{’asaroŋ}/ \quad \text{’a hoe’}
\end{align*}
\]

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

Brody, Jill. 1986. “Repetition as a rhetorical and conversational device in Tojolobal (Mayan).”

Campbell, Lyle. 1977. Quichean linguistic prehistory. (University of California publications in


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.


