## A. Administrative

1. **Title:** Proposal to Encode Additional Latin Phonetic and Orthographic Characters
2. **Requester's name:** Lorna A. Priest, Peter G. Constable
3. **Requester type (Member body/Liaison/Individual contribution):** Individual contribution
4. **Submission date:** 31 March 2005
5. **Requester's reference (if applicable):**
6. **Choose one of the following:**
   - This is a complete proposal: Yes
   - More information will be provided later: No

## B. Technical – General

1. **Number of characters in proposal:** 13
2. **Proposed category (select one from below - see section 2.2 of P&P document):**
   - A-Contemporary
   - 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
3. **Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):** 3
   - Is a rationale provided for the choice? No
   - If Yes, reference:
4. **Is a repertoire including character names provided?** Yes
   - Are the names in accordance with the “character naming guidelines” in Annex L of P&P document? Yes
5. **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:
6. **References:**
   - Are references to other character sets, dictionaries, descriptive texts etc.) provided? Yes
   - Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? Yes
7. **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)

---

C. Technical - Justification

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

   If YES explain ____________________________

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

   If YES, with whom? ____________________________

   If YES, available relevant documents: Email correspondence. See also References.

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

   Reference: See comments in Section E

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

   Reference: Orthographic characters are used in literacy materials, liturgical books and general literature. Upstep and downstep characters are used in linguistic writing.

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

   Reference: See comments in Section E

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

   If YES, is a rationale provided? ____________________________

   If YES, reference: If possible, should be kept with other related blocks in the BMP.

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

   If YES, is a rationale for its inclusion provided? ____________________________

   If YES, reference: ____________________________

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No

   If YES, is a rationale for its inclusion provided? ____________________________

   If YES, reference: ____________________________

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

    If YES, is a rationale for its inclusion provided? ____________________________

    If YES, reference: ____________________________

11. Does the proposal include use of combining characters and/or use of composite sequences? Yes

    If YES, is a rationale for such use provided? ____________________________

    If YES, reference: ____________________________

    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

    If YES, reference: ____________________________

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

    If YES, describe in detail (include attachment if necessary) ____________________________

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

    If YES, is the equivalent corresponding unified ideographic character(s) identified? ____________________________

    If YES, reference: ____________________________

D. Proposed Characters

A code chart and list of character names are shown on a new page.
D.1. Proposed Characters

<table>
<thead>
<tr>
<th>xx00</th>
<th>0</th>
<th>!</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>i</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>†</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>:</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>=</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>α</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>m</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>ε</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>W</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>W</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

D.1. Character Names

<table>
<thead>
<tr>
<th>xx00</th>
<th>MODIFIER LETTER AFRICANIST DOWNSHIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>xx01</td>
<td>MODIFIER LETTER AFRICANIST UPSTEP</td>
</tr>
<tr>
<td>xx02</td>
<td>MODIFIER LETTER RAISED UP ARROW</td>
</tr>
<tr>
<td>xx03</td>
<td>MODIFIER LETTER RAISED DOWNSHIFT</td>
</tr>
<tr>
<td>xx04</td>
<td>MODIFIER LETTER COLON</td>
</tr>
<tr>
<td>xx05</td>
<td>MODIFIER LETTER SHORT EQUAL</td>
</tr>
<tr>
<td>xx07</td>
<td>LATIN CAPITAL LETTER ALPHA</td>
</tr>
<tr>
<td></td>
<td>• lowercase is U+0251 α LATIN SMALL LETTER ALPHA</td>
</tr>
<tr>
<td>xx08</td>
<td>LATIN CAPITAL LETTER M WITH HOOK</td>
</tr>
<tr>
<td>xx09</td>
<td>LATIN LETTER TRESILLO</td>
</tr>
<tr>
<td>xx0A</td>
<td>LATIN LETTER CUATRILLO</td>
</tr>
<tr>
<td>xx0B</td>
<td>LATIN CAPITAL LETTER W WITH HOOK</td>
</tr>
<tr>
<td>xx0C</td>
<td>LATIN SMALL LETTER W WITH HOOK</td>
</tr>
<tr>
<td>xx0E</td>
<td>COMBINING DOUBLE DIAERESIS</td>
</tr>
</tbody>
</table>
D.3. Unicode Character Properties

xx00..xx05 should have a general category of Lm. Other properties should match those of similar characters, such as U+02BC MODIFIER LETTER APOSTROPHE.

xx07, xx08, and xx0B should have a general category of Lu. Other properties should match those of similar characters, such as U+0041 LATIN CAPITAL LETTER A.

xx0C should have a general category of Ll. Other properties for this character should match those of similar characters, such as U+0061 LATIN SMALL LETTER A.

xx09 and xx0A should have a general category of Lo. Other properties for this character should match those of similar characters, such as U+01C0 LATIN LETTER DENTAL CLICK.

xx0E should have a general category of Mn. Other properties for this character should match those of similar characters, such as U+035D COMBINING DOUBLE BREVE.

E. Other Information

E.1 Modifier Letters

IPA includes two symbols to indicate tonal downstep (see Figure 1, Figure 3 and Figure 4) and tonal upstep (see Figure 1 and Figure 5). These are IPA symbols 517 and 518 respectively.

The IPA Handbook cites these, and indicates that they can be represented in Unicode using the characters U+2193 DOWNWARDS ARROW and U+2191 UPWARDS ARROW:

However, the UCS characters U+2193 and U+2191 are full-height arrows, whereas the IPA symbols for downstep and upstep are raised, half-height arrows. IPA has full-height arrows as distinct symbols, used to represent ingressive versus egressive airflow in disordered speech; these are IPA symbols 661 and 662. These are also cited in the IPA Handbook, as shown in Figure 2. Note that in the IPA set these full arrows for indicating airflow have distinct names and numeric identifiers from the raised, half-height arrows used for downstep and upstep.

Accordingly, we consider it necessary that separate characters for the two pairs of arrow be encoded in the UCS. The existing characters, U+2193 and U+2191, are appropriate characters for representing IPA symbols 661 and 662. (Note that Unicode includes annotations for these characters indicating their use in IPA to indicate egressive and ingressive airflow.) Thus, we propose the addition of two new characters to represent the raised, half-height arrows used in IPA for tonal downstep and upstep: MODIFIER LETTER RAISED UP ARROW and MODIFIER LETTER RAISED DOWN ARROW.
The following addition samples illustrate these modifier letters in attested usage:

**Transcription of recorded passage**

iku₃u ùgụọ na aⁿigung naaⁿįịtụ a jìka ónọ a jìka ìbe jà iké mgbẹ hà nịmụ nụtụ ónọ ónọ iɗụ nwa ñu we ụguọ já náhịa. hà kwa-kwa-iụta na ónọ ëbụụ üzọ mèé ka ónọ iɗụ àfu jụpụ uwe ja ka a gi ewe dị ka ónọ ka ìbe jà iké. iku₃u ùgụọ wéé mànità fèè, fèè, fèè, ôtu iké ịjà hà; ma ka ó na efè ka ónọ iɗụ àfu na ndị ndị a jà ịké. jà fêkàtâ hâpụ. mgbẹ àfu a’ǹwụ wéé ụfụtụ, ụfụtụ-ké mèé ka ebe mìi- lié kpo’ọ ọ’kụ ná-àtụfùrị ogé ónọ iɗụ àfu jụpụ uwe ịjà ọké a me ẹzụ iku₃u ùgụọ kwe nà a’ǹwụ ka jà iké.

Figure 3. MODIFIER LETTER RAISED DOWN ARROW (IPA, 1999, p. 110)

**Grammatical contrast**

Progressive aspect

àákon à dép ọgụn. ‘Akon is buying (a) dress.’
àákon à wèt ọgwet. ‘Akon is writing (a) book/letter.’

Compleitive aspect

àákon à ’a dép ọgụn. ‘Akon has bought (a) dress.’
àákon à ’a wèt ọgwet. ‘Akon has written (a) book/letter.’

Basic tones on the verbs used to illustrate grammatical tone in Ibibio are H and L, thus: dép ‘buy’ and wèt ‘write’.

Verbs have two tone classes H and L. Disyllabic verbs have predictable high tone on the final syllable in isolation and in simple declarative sentences. Tone patterns in nouns are more variable. High (‘), Low (‘) and the downstep feature (‘) are marked. Where long vowels have identical tones, only the first vowel is marked with a tone.

Figure 4. MODIFIER LETTER RAISED DOWN ARROW (Urue, 2004, p. 108)

[y’c’ y’c’]

M M H U  ‘the dust will blow away’

The derivation for this example parallels that of the preceding example except that in this case, the application of Hi-Spread brings two hs into adjacency on the TBU tier, and upstep occurs.

One possible alternative to this analysis is that the floating Hi tone simply upsteps everything to its right, i.e., the surface tone of y’c’ (125) is simply the sequence Lo-Hi upstepped. However, if what Pike and Wistrand

Figure 5. MODIFIER LETTER RAISED UP ARROW (Snider, 1999, p. 111)

Africanist linguists have their own preferred conventions for indicating upstep and downstep, which are different from the IPA-recommended symbols. Tonal downstep is indicated by Africanists using a superscript exclamation mark, as illustrated in Figure 6 to Figure 11. For upstep, an inverted exclamation mark is used, though in some publications this is superscripted (see Figure 8), while in others it is
subscripted (see Figures 6–7 and Figures 9–11). Because of these glyph variants, we have proposed character names that avoid reference to the visual position of the inverted exclamations. An alternative would be to encode three characters: MODIFIER LETTER RAISED EXCLAMATION MARK, MODIFIER LETTER RAISED INVERTED EXCLAMATION MARK, and MODIFIER LETTER LOW INVERTED EXCLAMATION MARK.

Figure 6. MODIFIER LETTER AFRICANIST DOWNSTEP and MODIFIER LETTER AFRICANIST UPSTEP (Snider, 1993, p. 13)

I go on to explore the applicability of my total downstep treatment to two tonally well-documented languages which resemble Akan in having (at least) the three-way contrast [0 down]/[1 down]/[2 down] after a high tone, but which differ from Akan in also having a three-way contrast [-2 down]/[-1 down]/[0 down] after a low tone: Akan, of course, while it has the three-way contrast [HH][0 down]/[HH][1 down]/[HH][2 down] after [H], has only the two-way contrast [LH][1 down]/[LL][0 down] after [L]. The first of these two languages is Dschang, which, like Akan, is a downstep language by any definition, and the second Ebrié, which, although it has many of the characteristics of a downstep language, appears on the surface to be a discrete level tone language with three tone levels.

I suggest that the tone systems of these languages differ from that of Akan basically in that they do not have the tonal segment structure condition which in Akan disallows the [+high, +stepping] segment h, and that their [-2 down] is analysable as LhH \( \text{[L,H]} = \text{upstep} \) just

Figure 7. MODIFIER LETTER AFRICANIST DOWNSTEP and MODIFIER LETTER AFRICANIST UPSTEP (Snider, 1993, p. 187)

Pitch lowering can be contrastive (and is then termed downstep), as in Kenyang (Mamfe Bantu: Cameroon), where the site of pitch lowering is notated with \( \text{\textbullet} \).

(28) \text{\textbullet} \text{éb\textbullet y \textbullet mé\textbullet m\textbullet w\textbullet é} \quad \text{"it hurts me"}

Upstep (pitch raising) also exists. In Kimantuumbi (Bantu: Tanzania), adjacent H tones are separated by upstep (notated with \( \text{\textbullet} \)).

(29) \text{b\textbullet a\textbullet t\textbullet ī\textbullet yā \textbullet \textbullet k\textbullet ū\textbullet n\textbullet d\textbullet ū\textbullet ē} \quad \text{"they ate the birds"}

Figure 8. MODIFIER LETTER AFRICANIST DOWNSTEP and MODIFIER LETTER AFRICANIST UPSTEP (Odden, 1995, p. 452)
MODIFIER LETTER COLON and MODIFIER LETTER SHORT EQUAL are orthographic characters used for writing tone. Both are used in one or more languages of Congo (Budu); MODIFIER LETTER COLON is also used in at least one language in Kenya (Sabaot) and at least one language in Papua New Guinea (Kunzi-Boazi).

These characters are used in the orthographies of these languages to write grammatical tones. For instance, in the Budu language, MODIFIER LETTER COLON indicates past tense, whereas MODIFIER LETTER SHORT EQUAL indicates future tense; in the Sabaot language, the MODIFIER LETTER COLON functions as a subject marker.

These characters are distinct from the similar punctuation colon and equal sign characters. They are used as word-forming characters; so, for instance, they should never indicate a word boundary. Also, their graphic appearance is specifically distinct from the colon and equal sign:
Compared with a typical colon glyph, the dots of the MODIFIER LETTER COLON should be closer together, and the side bearings should be narrower. (Note that the sample in Figure 15 is not as good an examplar in this regard as the other samples.)

Compared with a typical glyph for the equal sign, the MODIFIER LETTER SHORT EQUAL is rather shorter.

The following figures demonstrate attested usage of these characters:

![Figure 12](image1.png)

**Ezoqa TiBi enqoni**


**Figure 12. MODIFIER LETTER COLON** (Fumey, 2001, p. 14) [Kuni-Boazi language].


Geté bazaflagah ndøgo, no qæzømbemanqatum, taqá tozomboékhaékewagát, ta

**Figure 13. MODIFIER LETTER COLON** (Fumey, 2004, p. 1) [Kuni-Boazi language].

The Sabaoit Language Committee has agreed to use the marker : as a 'subject mark' to be added to the noun, when it functions as subject in the sentence. Therefore, when we write kâwan, it means that 'his father' did something. But when we write just kâwan, it means that something was done to 'his father'. The two sentences above can now be written clearly as and exactly as follows:

1. Kâmwoochi kâwan.
2. Kâmwoochi :kâwan.

Now try the following exercise:

**Exercise 11:** Translate the following sentences into English.

1. Kìkas bëlyoontët wëlit koosye.
2. Yu kaka bëlyoontët, kulambíyi keey :boontëet.

**Figure 14. MODIFIER LETTER COLON** (BTL, 1990, p. 15) [Sabaoit language].
E.2 Latin Extended

LATIN CAPITAL LETTER ALPHA is a phonetic symbol in the Americanist tradition for a voiceless, low central vowel (see Figure 16).

Although Pike (1947) and Floyd (1981) do not illustrate the use of this symbol, both make general
reference to the use of capital letters for transcribing voiceless vocoids:

“Voiceless vocoids may be written with capital letters, or letters extra-large in size, such as ‘A’ and ‘ơ’.” (Pike 1947, p. 5)

“Symbolization: VI. vocoids are symbolized by capital letters: [A I O E …] or by large letters: [ I ơ E …].” (Floyd 1981, p. 37).

This character is also used orthographically in the Fe’fe’e language of Cameroon. Note that LATIN CAPITAL LETTER A is also used in this orthography, so it cannot be considered a glyph variant of LATIN CAPITAL LETTER A.

LATIN CAPITAL LETTER M WITH HOOK is used in phonetic transcription by some in the Americanist tradition for transcribing a voiceless labiodental nasal stop.

Figure 18. LATIN CAPITAL LETTER ALPHA (Tcheulahie, 1984 p. 50) [Fe’Fe’e language]

Figure 19. LATIN CAPITAL LETTER M WITH HOOK (Pike, 1976, p.70)
LATIN LETTER TRESILLO and LATIN LETTER CUATRILLO were used by Jesuit missionaries in the XVIth century to transcribe certain sounds in Mayan languages of Guatemala that do not occur in Spanish. (It is unclear whether the Jesuit missionaries were the inventors of these characters or if they had prior use elsewhere.) In addition to older linguistic descriptions by the Jesuits, these characters are also found in recent Mayanist publications when citing data from Jesuit sources (see also: http://scripts.sil.org/RecentCuatrilloUse).

As can be seen in Figure 21, there may be additional characters that were used by the Jesuit linguists that will need to be added to the UCS. The need for the two proposed here is clear, however.

LATIN CAPITAL LETTER W WITH HOOK and LATIN SMALL LETTER W WITH HOOK are used in the orthographies of certain languages spoken in Burkina Faso: they are in current use in the orthography of the Puguli language, and the language committee for the Lobiri language also has plans to use this character in their orthography. The orthography guide for Lobiri is in the process of being finalized. These languages have a rare contrast between egressive and ingressive labial semivowels ("w"); hence, these characters are used in these orthographies contrastively with "W" and "w".
E.3 Combining Diacritical Marks

COMBINING DOUBLE DIAERESIS is used in the Teribe/Naso language of Panama, which uses it centered over a double “el.” This orthography was established in November, 2000. A description of the process for decision-making follows [SBP, translation of p. 5]:

Introduction to the Naso alphabet

In the past, many people tried to create an alphabet for the Naso people, but none of these was sanctioned by the internal government of the Naso people as the official alphabet. For this reason, there were several different alphabets.

With the goal of coming up with a single official alphabet that would meet the needs of all of the Naso people, a meeting was held in November of 2000. This meeting was sanctioned by King Tito Santana V. and the leadership of the Naso counsel. Participants were invited from a variety of communities including Dluy, Kûy, Loma Bandera, Santa Rosa, Siey Kjing, Siey Llik, Solong and Changuinola.

During three days, this group studied, discussed and defined the official Naso alphabet. Afterwards, they decided to elect The Naso Language Committee, which continued to work on the Naso dictionary.

The Naso alphabet that we have today is the result of the work done at this meeting and at several meetings of this committee. This committee also approved the spelling of each Naso word that is used in this book.

COMBINING DOUBLE DIAERESIS can be seen above the double-el in Figures 24 and 25. It is also interesting to note in Figure 25 that COMBINING DIAERESIS is used over a single “el”.

Figure 23. LATIN CAPITAL LETTER W WITH HOOK (SIL, 2000, cover) [Puguli language]

Figure 24. LATIN SMALL LETTER W WITH HOOK (ANTBA, 2004, p. 2) [Puguli language]
Lección 8: La letra LL

Indicaciones de la lección:
En esta lección estudiamos la letra LL. Se llama “doble naso” o “doble de naso”.

La letra LL:

Práctica 1:
Lean estas palabras tres veces. Primero lean individualmente en voz baja. La segunda vez lean juntos en grupo en voz alta. Luego voluntarios pueden leer uno por uno en voz alta.

dling    patoca

dling    bejco real

llbo    algo

llwek    enderezar

Figure 25. COMBINING DOUBLE DIAERESIS (SBP, 2003, 2005, p. 31) [Naso/Teribe language of Panama]

Kjok wen soré, e lanyo

1 Kjok ére dámê, llbêye îjêm obishko ga kjok uum îj sháryara Sbôré. 2 Klung ére igovy îjêm obi, dlung rong, kjing yêr oró. Eshko ga Sbô Sêyara wir pjang dlung kjing erâ, nîlën welen.

Dbar kjwara


Figure 26. COMBINING DOUBLE DIAERESIS (SBP, 2004, p. 1) [Naso/Teribe language of Panama]
F. References


Bible Translation and Literacy. 1990. *Korooryo ku taay* (Reading and writing Sabaot). Nairobi, Kenya: Bible Translation and Literacy (E.A.)


