### A. Administrative

1. **Title:** Proposal to Encode Additional Latin and Cyrillic Characters  
2. **Requester's name:** Lorna A. Priest  
3. **Requester type (Member body/Liaison/Individual contribution):** Individual  
4. **Submission date:** 23 April 2008  
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. **Choose one of the following:**  
   - a. This proposal is for a new script (set of characters): No  
   - b. The proposal is for addition of character(s) to an existing block: Yes  
     Name of the existing block: Latin Extended-D and Cyrillic Extended-B  
2. **Number of characters in proposal:** 4  
3. **Proposed category (select one from below - see section 2.2 of P&P document):**  
   - A-Contemporary X  
   - 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  
4. **Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):** 1  
   - Is a rationale provided for the choice? No  
5. **Is a repertoire including character names provided?** Yes  
   - a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? Yes  
6. **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:  
7. **References:**  
   - a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? Yes  
   - b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? Yes  
8. **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)? Suggested character properties are included.  

---

### 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?**  
   Linguists, LIBTRALO in Liberia and Judeo-Tat experts.

   **If YES, available relevant documents:**  
   Examples in document were provided by them.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?  
   Reference: See comments in Section E

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

   **Reference:** See examples in Section E

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

   **If YES, where? 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?  
   Yes

   **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?  
    No

    **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:**
<table>
<thead>
<tr>
<th>A78</th>
<th>052</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>♹</td>
</tr>
<tr>
<td>7</td>
<td>♹</td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>♹</td>
</tr>
<tr>
<td>E</td>
<td>♹</td>
</tr>
<tr>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

**D.1. Proposed Characters**

<table>
<thead>
<tr>
<th>A78D</th>
<th>LATIN CAPITAL LETTER TURNED H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lowercase is 0265 Ꭓ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A78E</th>
<th>LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>0526</th>
<th>CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER</th>
</tr>
</thead>
</table>

| 0527 | CYRILLIC SMALL LETTER SHHA WITH DESCENDER        |
D.3. Unicode Character Properties

A78D should have a general category of Lu. Other properties should match those of similar characters, such as U+0041 LATIN CAPITAL LETTER A. Lower Case Equivalent is U+0265. The properties for 0265 should change as well. See below.

A78E 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.

0526 should have a general category of Lu. Other properties should match those of similar characters, such as U+04BA CYRILLIC CAPITAL LETTER SHHA.

0527 should have a general category of Ll. Other properties for this character should match those of similar characters, such as U+04BB CYRILLIC SMALL LETTER SHHA.

E. Other Information

E.1 Latin Extended

LATIN CAPITAL LETTER TURNED H is used in the Dan/Gio ((ISO 639-3:daf) orthography in Liberia (however, it is not in the Dan/Gio orthography used in Cote d’Ivoire). There are up to 200,000 Dan/Gio speakers in Liberia. Figure 1 shows the uppercase in the Gio New Testament. In recent years the Gio have been producing books, but they have not had any fonts with the uppercase character. They have sometimes just used the lowercase letter in place of an uppercase (as in the case of Figure 2), and sometimes they have used character formatting to make the uppercase larger and raised to the baseline (see MS Publisher screenshot in Figure 3).

![Figure 1. LATIN CAPITAL LETTER TURNED H (Gio NT, 1989, p.380).]
Debe e koza dú ka.
qù tó e too dòe?
N tò. Gblseqa-dú-sù, n bòu too sou-kela.
A lo yi saa mú.

Figure 2. LATIN SMALL LETTER TURNED H (Dunah, 2001, p.38).
LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT is a phonetic symbol representing a voiceless lateral retroflex fricative. Ladefoged discusses lack of IPA coverage and ad hoc solutions in “A Course in Phonetics” (page 278). He and others (Shalev and Spajic) have used LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT to transcribe Toda (ISO 639-3:tcx). See Figures 4-6.

**Phonemic vowel length**

In languages which distinguish between short and long vowels, there appears to be no single durational ratio by which this phonological distinction is implemented phonetically. The ratio of short to long vowel durations (V/V): “is close to 50%, but may vary a great deal” (Lehiste, 1960:34). Table 3 shows mean durations of vowels in the data set: it ‘flour’, kits ‘link hands’ ~ iit ‘spear’; nys ‘broken rice’ ~ pyt ‘summer’; its ‘to reach’ ~ pif ‘flesh’, wif ‘large lizard’; put ‘stirring stick’ ~ put ‘eighteen’, ut ‘anthill’ ~ nts ‘throw away water’; et ‘where’ ~ at ‘to take’, tef ‘wrap garment around waist’ ~ tæt ‘to fold’, pæt ‘delivery’; et to climb’ ~ ost ‘to boil’; pot ‘ten’, kots ‘brass vessel’ ~ moqt ‘language’, moqt ‘change direction’; jot ‘churning vessel’ ~ pot ‘cockroach’.

![Figure 4. LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT (Shalev, 1993, p.93).](image1)

---

![Figure 5. LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT (Shalev, 1993, p.124).](image2)
There are three alphabets used in Tati (Muslim Tat, code TTT, pop: 26,000) and Juhuri (Judeo-Tat, code JDT, pop: 101,000). Both dialects are written commonly in Cyrillic, Latin (an earlier form used in Azerbaijan for both Tats, 1921-39), and an “Azerbaijani Cyrillic” alphabet (used in Azerbaijan, 1939-1991). We are proposing the addition of the CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER and CYRILLIC SMALL LETTER SHHA WITH DESCENDER which are used in the “Azerbaijani Cyrillic” alphabet. Phonetically, the character represents a voiceless gutteral plosive (like Hebrew ‘ayin.’). See Figures 7-9.

Figure 2. The Toda consonant inventory

Previous accounts of the status of voicing in Toda are not completely clear. Emeneau (1984:14) divides the consonants into the following sets (using our symbols):

invariable voiceless consonants: p t s s s f x
invariable voiced consonants: b d z d z d o l i z s f
variable voiceless consonants: x
variable voiced consonants: m n j

Figure 6. LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT (Spajić, 1994, p. 36).

хуб мәңгум һисдиңо чәдәлән ән әлфбой ә йәр ишмү вәнкәнүм, ишмү бәгде и 2 чә

Аа Бб Вв Гг Дд Ее Чч Зз Ии Йй Кк Лл Мм Нн Оо Пп Рр Сс Тт Уу Яя Ӑ Ӓ Ӕ Ә ӽ Ӽ Ӥ ӹ
Figure 8. CYRILLIC SMALL LETTER SHHA WITH DESCENDER (Агаронова, 2006, p. 126).

Figure 9. CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER and CYRILLIC SMALL LETTER SHHA WITH DESCENDER (Shamayev, 2006, p. 195).

F. References


Shalev, Michael et al. June 1993. “Phonetics of Toda”. In Fieldwork Studies of Targeted Languages. UCLA WPP 84, pp 89-126.
