

**ISO/IEC JTC 1/SC 2/WG 2**  
**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS**  
**FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest *Roadmaps*.

**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
- or, 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
- Proposed name of script: \_\_\_\_\_
- 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  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
- If Yes, reference: \_\_\_\_\_
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
- b. Are the character shapes attached in a legible form suitable for review? 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.

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain _____	<u>No</u>
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.)? If YES, with whom? <u>Linguists, LIBTRALO in Liberia and Judeo-Tat experts.</u> If YES, available relevant documents: <u>Examples in document were provided by them.</u>	<u>Yes</u>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: <u>See comments in Section E</u>	_____
4. The context of use for the proposed characters (type of use; common or rare) Reference: <u>See examples in Section E</u>	<u>common</u>
5. Are the proposed characters in current use by the user community? If YES, where? Reference: <u>See comments in Section E</u>	<u>Yes</u>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: <u>If possible, should be kept with other related blocks in the BMP.</u>	<u>Yes</u>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? <u>Preferably together with other related blocks</u>	_____
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
11. Does the proposal include use of combining characters and/or use of composite sequences? 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: _____	<u>No</u>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary) _____	<u>No</u>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference: _____	<u>No</u>

**D.1. Proposed Characters**

	A78	052
0		
1		
2		
3		
4		
5		
6		ḥ
7		ḥ
8		
9		
A		
B		
C		
D	ҫ	
E	ҫ̣	
F		

**D.1. Character Names**

A78D LATIN CAPITAL LETTER TURNED H

- lowercase is 0265 ȷ

A78E LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT

0526 CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER

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.

```
0265;LATIN SMALL LETTER TURNED H;Ll;0;L;;;;N;;;A78D;;A78D
A78D;LATIN CAPITAL LETTER TURNED H;Lu;0;L;;;;N;;;0265;
A78E;LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT;Ll;0;L;;;;N;;;;;
0526;CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER;Lu;0;L;;;;N;;;0527;
0527;CYRILLIC SMALL LETTER SHHA WITH DESCENDER;Ll;0;L;;;;N;;;0526;;0526
```

### 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).

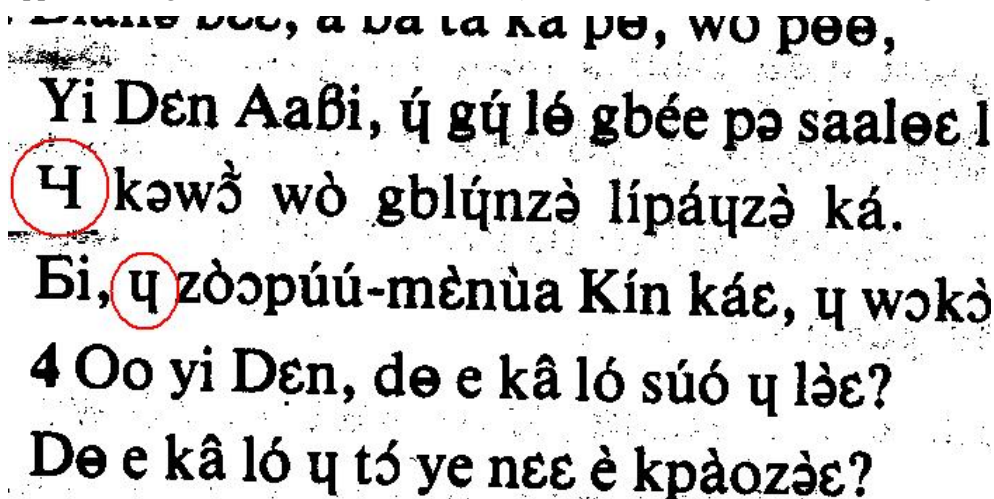


Figure 1. LATIN CAPITAL LETTER TURNED H (Gio NT, 1989, p.380).

# Dú

Debe e kozə dú ka.

η tó e tóə dəe?

N tóə Gblɔwɛa-dú-sú, n bɔu tóə sou-kɛla.

A lo yi saa mɛ.

Figure 2. LATIN SMALL LETTER TURNED H (Dunah, 2001, p.38).

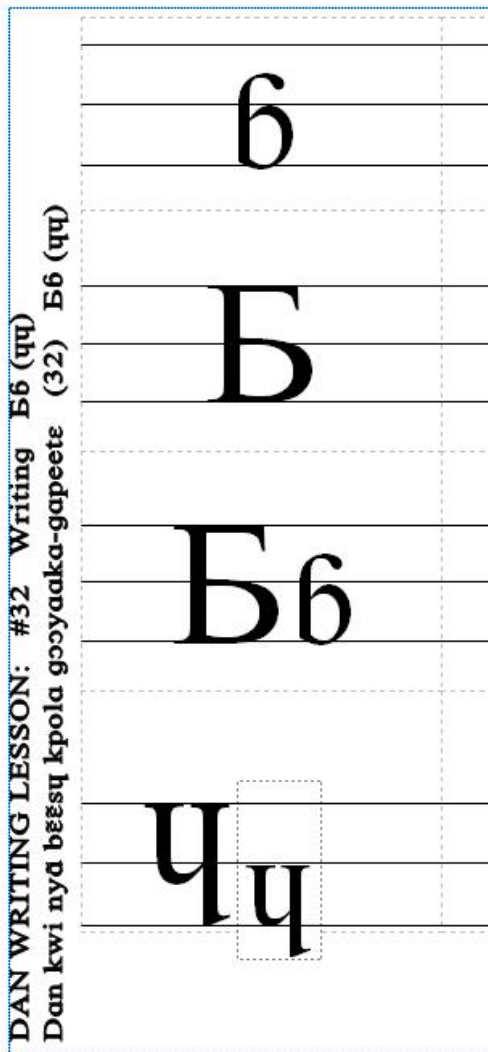


Figure 3. LATIN CAPITAL LETTER TURNED H (The Dan Literacy Association, 2001, p.64).

LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT is a phonetic symbol representing a voiceless lateral retroflex fricative. Lagedoged 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: **it̥** ‘flour’, **kits̥** ‘link hands’ ~ **i:t̥** ‘spear’; **nyts̥** ‘broken rice’ ~ **py:t̥** ‘summer’; **its̥** ‘to reach’ ~ **pit̥** ‘flesh’, **wit̥f** ‘large lizard’; **put̥** ‘stirring stick’ ~ **pu:t̥** ‘eighteen’, **ut̥** ‘anthill’ ~ **u:ts̥** ‘throw away water’; **et̥** ‘where’ ~ **æ:t̥** ‘to take’, **t̥et̥** ‘wrap garment around waist’ ~ **t̥æ:t̥** ‘to fold’, **pæ:t̥** ‘delivery’; **ot̥** ‘to climb’ ~ **o:t̥** ‘to boil’; **pot̥** ‘ten’, **kots̥** ‘brass vessel’ ~ **mo:t̥** ‘language’, **mo:t̥** ‘change direction’; **t̥ot̥** ‘churning vessel’ ~ **pa:t̥** ‘cockroach’.

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

<b>poq̥</b>	large fly sp.
<b>po̥l̥ o:t̥</b>	chief priet of the Ti dairy
<b>po̥ly</b>	sacred dairy
<b>po̥lt̥</b>	white
<b>pu̥t̥</b>	beating
<b>pu̥t̥c̥</b>	blue-gray
<b>pu̥t̥f̥</b>	flower
<b>pu̥:my</b>	earth
<b>pu̥:t̥</b>	eighteen
<b>pu̥f̥</b>	worm
<b>pum̥</b>	fruit
<b>pu̥t̥</b>	stirring stick
<b>py̥:t̥</b>	summer
<b>pu̥:ɹ̥</b>	flesh, meat, muscle
<b>pu̥:l̥</b>	price
<b>nu̥:l̥</b>	crevice in wall

Figure 5. LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT (Shalev, 1993, p.124).

PLACE ⇒ MANNER ↓	LABIAL	DENTAL	POST-DENTAL	ALVE-OLAR	RETRO-FLEX	PALATO-ALVEOLAR	VELAR
STOP (AND) AFFRICATE	p b	t̪ d̪	ts dz	t̪ d̪	t̪̺ d̪̺	tʃ dʒ	k ɡ
NASAL	m			n	ɳ		
FRICATIVE	f	θ	ɬ	s	ʂ	ʃ	x (χ)
TRILL			ʀ	r	ɽ		
APPROX-IMANT						j	w
LATERAL			l̪ l		l̪̺ l̪̺		

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 t̪ ts t̪̺ tʃ t̪̺̺ k  
 invariable voiced consonants: b d̪ dz d̪̺ d̪̺̺ ɡ l̪ l̪̺ ʒ ʀ w  
 variable voiceless consonants: f θ x r̪ r̪̺ t̪̺̺̺ s̪ s̪̺  
 variable voiced consonants: m n ɳ j

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

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 guttural plosive (like Hebrew 'ayin.'). See Figures 7-9.

хуб мәһлүм һисдиһо чәдвәл ән әлфбой  
 ә йор ишму вәнкәнүм, ишму бәгдә и 2 чә

Аа Бб Вв Гг Дд Ее Чч Зз Ии Йй  
 Кк Кк Лл Мм Нн Оо Пп Рр Сс Тт  
 һһ Һһ Һһ Уу Фф Хх Чч Шш Әә Үү

Figure 7. CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER and CYRILLIC SMALL LETTER SHHA WITH DESCENDER (Gurshumov, 2006, p. 114).

Дэсэ-дэс дит, эйэки шолуми сохит.  
Бүтүн **ш**иломэ эводу сохит.  
Күрпино дэшэнит, рэһно дуз сохит.  
Инсонһорэ эйэки шолуми сохит.

Figure 8. CYRILLIC SMALL LETTER SHHA WITH DESCENDER (Агаронова, 2006, p. 126).

**ш**овунэ кор нэсох, мисвојэ кор сох,  
Э худојтү, э тиројтү бовор сох.

Сэр нэзэнит э дуркунэ худоһо,  
Данитки јэкини э дүнјоһ Худо.

Нум худорэ нэкирит э нуһоги,  
Борухэшим тирозуј эну һэги.

Бурч түни һүрмэт сох, э бэбэ, дэдэј,  
Тэ эхир **ш**умүртү, урэ нисд тэ кэј.

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

## F. References

- Бат Гилил (Агаронова) Овгоил. 2006. «Шолум Сохим», Мирвори - Литературно-публицистический альманах No. 1 (© Иядательский дом «МИРВОРИ» Израиль, 2006).
- The Dan Literacy Association. 2001. *Dan Wo Lonosy Wāa a Bēesy Gēgbaā Kwi*. (First Step to Reading and Writing Dan). Liberian Bible Translation & Literacy Organization (LIBTRALO). Monrovia, Liberia.
- Dunah, Menmon Paul Z. 2001. *Ko Dan Kwi Lono: Kwi Be I* (Dan Graded Reader Book I). Dan Literacy and Translation Association. Monrovia, Liberia.
- Gurshumov, E. 2006. «Эз товнэй алифбой дэрбэнди» Мирвори - Литературно-публицистический альманах No. 1 (© Иядательский дом «МИРВОРИ» Израиль, 2006).
- Ladefoged, Peter. 1993. A Course in Phonetics, Third Edition. Fort Worth: Harcourt Brace College Publishers, pp. 275 - 280. These pages are found here:  
<http://www.es.unizh.ch/jstraessler/Ladef%20pp%20275-280.pdf> (accessed 30-Oct-07).
- Spajić, Siniša et al. Aug. 1994. “The rhotics of Toda”. In *Fieldwork Studies of Targeted Languages II*. UCLA WPP 87, pp 35-66.
- Shalev, Michael et al. June 1993. “Phonetics of Toda”. In *Fieldwork Studies of Targeted Languages*. UCLA WPP 84, pp 89-126.
- Shamayev, R. 2006. *Shoots from One Branch*. Self-published. Brooklyn, NY.
- Trinitarian Bible Society. 1989. *Wodowo Dou* (The New Testament in the Gio Language of the Republic of Liberia). Trinitarian Bible Society. London, England.