

UC Berkeley

Proposals from the Script Encoding Initiative

Title

Final proposal to encode the Ol Chiki script in the UCS

Permalink

<https://escholarship.org/uc/item/70g5x287>

Author

Everson, Michael

Publication Date

2005-09-21

Peer reviewed

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Final proposal to encode the Ol Chiki script in the UCS

Source: Michael Everson

Status: Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2005-09-21

This is a revision of N1956 and N2505 and contains the proposal summary form.

Introduction

The Ol Chiki script, also called Ol Cemet', Ol Ciki, or simply Ol, was invented by Pandit Raghunath Murmu in the first half of the 20th century CE to write Santali, a Munda language of India. Ol Chiki is alphabetic, sharing none of the syllabic properties of the other Indic scripts. The Ol Chiki script has received some official recognition and Raghunath has been honoured by the Orissan government. Ol Chiki has recently been promoted by some Santal organizations, with uncertain success, for use in certain other Munda languages in the Chota Nagpur area as well as the Dravidian Kuḍux language.

According to the Ethnologue, Santali's various dialects are spoken by 5.8 million people with 25% to 50% literacy, mostly in India with a few in Nepal and Bangladesh. The Ol Chiki script is used for the southern dialect of Santali as spoken in the Orissan Mayurbhañj district. While this dialect has only six vowels, the Santal Parganas dialect has eight or nine vowels. The extra Santal Parganas vowels are made by combining three vowels with the diacritic GAAHLAA TTUDDAAG.

Glyph placement

Ol Chiki is written from left to right, and consists of letters which represent consonants or vowels. In addition to these, a number of modifier letters are used to indicate tone, nasalization, length, and deglottalization. There are no combining characters. When both the nasalizing <'> MU TTUDDAAG and the modifying <.> GAAHLAA TTUDDAAG are applied to a syllable, a third character, <.'> MU-GAAHLAA TTUDDAAG, is used. Ol Chiki keyboards have keys for all three (see Figure 9); there is no advantage to adding rendering complexity to Ol Chiki rendering by combining these two.

Digits and punctuation

Digits have distinctive forms. In the samples viewed at present, European COMMA, EXCLAMATION MARK, and QUESTION MARK are used, as are "English quotation marks". The FULL STOP is not used, doubtless because it can be confused with <.> GAAHLAA TTUDDAAG. Both | PUNCTUATION MUCAAD (ᱫᱷ ᱠᱟᱨᱢᱟᱫᱽ) *mucăd* [mucət'] and || PUNCTUATION DOUBLE MUCAAD are used in poetry; PUNCTUATION MUCAAD is also used in prose. In Figure 4, a sort of high ellipsis can be seen; this should be added to the standard, but it is not certain whether this should be script-specific or generic punctuation. (Probably the latter.)

Gottalization and aspiration

The AHAD character is used to deglottalize consonants, a unique feature of the writing system which, as Zide 1996 says: "certainly increases the efficiency of writing Santali.... This neatly preserves the

morphophonemic relationships between the glottalized and voiced equivalents: the former occurs word-finally and at certain word-internal preconsonantal junctures, the latter prevocally, but never morpheme-initially in these alternations.” The letter 𑄎 AG therefore represents [kʰ] when written in word-final position and [g] when written word initially, or when written 𑄎 in word-final position. Aspiration of consonants is indicated by following a consonant with 𑄎 OH: thus 𑄎 [tʰ], 𑄎 [gʰ], 𑄎 [kʰ], 𑄎 [jʰ], 𑄎 [cʰ], 𑄎 [dʰ], 𑄎 [pʰ], 𑄎 [tʰ], 𑄎 [dʰ], and 𑄎 [bʰ].

Glyph variants

In handwriting and script fonts, the letters the letters which combine with AHAD ligate with it. So for 𑄎, instead of 𑄎, 𑄎 is written; for 𑄎, instead of 𑄎, 𑄎 is written; for 𑄎, instead of 𑄎, 𑄎 is written; for 𑄎, instead of 𑄎, 𑄎 is written (this is rare, and 𑄎 itself is handwritten 𑄎); and for 𑄎, instead of 𑄎, 𑄎 is written. Apart from handwriting fonts, these ligatures have not acquired typographic forms. Should such be required, ZWJ could be used for force them, since it is unlikely to be obligatory (it is obligatory in handwriting fonts, though, and should be built-in to them).

Names and ordering

Characters are arranged in a 5 by 6 matrix, named in a conventional way as shown in the names list. The first characters in each row (LA, LAA, LI, LU, LE, LO) are vowels. Given here is the UCS name, transliteration according to Zide 1996, the transliteration from the *ALA Romanization Handbook*, and the phonetic value(s) of the letters. In Figure 2 transliteration into Devanagari, Bangali, and Oriya are given. The UCS names reflect the Devanagari transliterations as rendered according to UCS convention; the modifier letters are not taken into account in transcribing the names (so 𑄎 [a] and 𑄎. [ə] are both AA, and so on).

𑄎 _{LA} , <i>la, la</i> , [ɔ]	𑄎 _{AT} , <i>at, at</i> , [t]	𑄎 _{AG} , <i>ak', ag</i> , [k', g]	𑄎 _{ANG} , <i>ani, ani</i> , [ɲ]	𑄎 _{AL} , <i>al, al</i> , [l]
𑄎 _{LAA} , <i>lā, lā</i> , [a]	𑄎 _{AAK} , <i>āk, āk</i> , [k]	𑄎 _{AAJ} , <i>āc', aj</i> , [c', dʒ]	𑄎 _{AAM} , <i>ām, ām</i> , [m]	𑄎 _{AAW} , <i>āw, āw</i> , [wv]
𑄎 _{LI} , <i>li, li</i> , [i]	𑄎 _{IS} , <i>is, is</i> , [s]	𑄎 _{IH} , <i>ih, ih</i> , [h, ʔ]	𑄎 _{INY} , <i>iñ, iñ</i> , [ɲ]	𑄎 _{IR} , <i>ir, ir</i> , [r]
𑄎 _{LU} , <i>lu, lu</i> , [u]	𑄎 _{UC} , <i>uc, uc</i> , [c]	𑄎 _{UD} , <i>ut', ud</i> , [t', d]	𑄎 _{UNN} , <i>uṇ, uṇ</i> , [ɲ]	𑄎 _{UY} , <i>uy, uy</i> , [h]
𑄎 _{LE} , <i>le, le</i> , [e]	𑄎 _{EP} , <i>ep, ep</i> , [p]	𑄎 _{EDD} , <i>eḍ, eḍ</i> , [ḍ]	𑄎 _{EN} , <i>en, en</i> , [n]	𑄎 _{ERR} , <i>eṛ, eṛ</i> , [ɾ]
𑄎 _{LO} , <i>lɔ, lɔ</i> , [o]	𑄎 _{OTT} , <i>ot, ot</i> , [t]	𑄎 _{OB} , <i>op', ob</i> , [p', b]	𑄎 _{OV} , <i>oṽ, oṽ</i> , [ṽ]	𑄎 _{OH} , <i>oh, oh</i> , [ʰ]

Processing

There are orthographic restrictions as to what characters some of the diacritical modifiers can follow.

- 1 The nasalization mark <̃> MU TTUDDAAG 1C78 (𑄎 𑄎𑄎𑄎𑄎.𑄎) *mū tuḍäg* [mū tuḍəkʰ] can follow any vowel, long or short. In the sources consulted, I have found 𑄎, 𑄎, 𑄎, 𑄎, and 𑄎.
- 2 The vowel modifier <-> GAAHLAA TTUDDAAG 1C79 (𑄎.𑄎𑄎. 𑄎𑄎𑄎.𑄎) *gāhlā tuḍäg* [gəhlə tuḍəkʰ] follows 𑄎 1C5A *a*, 𑄎 1C5F *ā*, and 𑄎 1C6F *e*. In the sources consulted, I have found all three: 𑄎. ǎ [ɔ], 𑄎. ǎ̃ [ə], and 𑄎. ǎ̃̃ [ɛ].
- 3 The nasalization mark <̃> MU TTUDDAG 1C78 and the vowel modifier <-> GAAHLAA TTUDDAAG 1C79 when used together form the mark <-̃> MU-GAAHLAA TTUDDAG 1C7A (𑄎 𑄎.𑄎𑄎. 𑄎𑄎𑄎.𑄎) *mū gāhlā tuḍäg* [mū gəhlə tuḍəkʰ] in the text stream. Example: 𑄎. ǎ̃̃̃.
- 4 The length mark ~ RELAA 1C7B (𑄎𑄎𑄎) *relā* [rela:] may combine with any oral or nasal vowel.
- 5 The glottal protector ˀ PHARKAA 1C7C (𑄎𑄎𑄎𑄎) *phārkā* [pʰa:rka:] follows the four glottal consonants when preceding a consonant or vowel (otherwise the glottal consonant is deglottalized by position, so 𑄎-𑄎 is [kʰɔ] and 𑄎𑄎 is [gɔ]).

6 The deglottalizing Ɂ AHAD 1C7D (ᱠᱡᱢᱦ) *ahad* [ɔhətʰ] can only follow four consonants: Ɂ *k*' (becomes -ɁɁ -g), Ɂ *c*' (becomes ɁɁ -j), Ɂ *t*' (becomes ɁɁ -d), Ɂ *p*' (becomes ɁɁ -b).

Implementations

R. C. Hansdah and N. C. Murmu have made a number of Ol Chiki fonts available; these all map the Ol Chiki characters to ASCII characters. Revising these to UCS fonts will be easily accomplished. Ol Chiki presents no implementation problems.

Collating order

The collation would appear to be alphabetic and to follow the order of the characters in the code table. I have not seen an ordered wordlist.

Unicode Character Properties

```

1C50;OL CHIKI DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
1C51;OL CHIKI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
1C52;OL CHIKI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
1C53;OL CHIKI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
1C54;OL CHIKI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
1C55;OL CHIKI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
1C56;OL CHIKI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
1C57;OL CHIKI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
1C58;OL CHIKI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
1C59;OL CHIKI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
1C5A;OL CHIKI LETTER LA;Lo;0;L;;;;;N;;;;;
1C5B;OL CHIKI LETTER AT;Lo;0;L;;;;;N;;;;;
1C5C;OL CHIKI LETTER AG;Lo;0;L;;;;;N;;;;;
1C5D;OL CHIKI LETTER ANG;Lo;0;L;;;;;N;;;;;
1C5E;OL CHIKI LETTER AL;Lo;0;L;;;;;N;;;;;
1C5F;OL CHIKI LETTER LAA;Lo;0;L;;;;;N;;;;;
1C60;OL CHIKI LETTER AAK;Lo;0;L;;;;;N;;;;;
1C61;OL CHIKI LETTER AAJ;Lo;0;L;;;;;N;;;;;
1C62;OL CHIKI LETTER AAM;Lo;0;L;;;;;N;;;;;
1C63;OL CHIKI LETTER AAW;Lo;0;L;;;;;N;;;;;
1C64;OL CHIKI LETTER LI;Lo;0;L;;;;;N;;;;;
1C65;OL CHIKI LETTER IS;Lo;0;L;;;;;N;;;;;
1C66;OL CHIKI LETTER IH;Lo;0;L;;;;;N;;;;;
1C67;OL CHIKI LETTER INY;Lo;0;L;;;;;N;;;;;
1C68;OL CHIKI LETTER IR;Lo;0;L;;;;;N;;;;;
1C69;OL CHIKI LETTER LU;Lo;0;L;;;;;N;;;;;
1C6A;OL CHIKI LETTER UC;Lo;0;L;;;;;N;;;;;
1C6B;OL CHIKI LETTER UD;Lo;0;L;;;;;N;;;;;
1C6C;OL CHIKI LETTER UNN;Lo;0;L;;;;;N;;;;;
1C6D;OL CHIKI LETTER UY;Lo;0;L;;;;;N;;;;;
1C6E;OL CHIKI LETTER LE;Lo;0;L;;;;;N;;;;;
1C6F;OL CHIKI LETTER EP;Lo;0;L;;;;;N;;;;;
1C70;OL CHIKI LETTER EDD;Lo;0;L;;;;;N;;;;;
1C71;OL CHIKI LETTER EN;Lo;0;L;;;;;N;;;;;
1C72;OL CHIKI LETTER ERR;Lo;0;L;;;;;N;;;;;
1C73;OL CHIKI LETTER LO;Lo;0;L;;;;;N;;;;;
1C74;OL CHIKI LETTER OTT;Lo;0;L;;;;;N;;;;;
1C75;OL CHIKI LETTER OB;Lo;0;L;;;;;N;;;;;
1C76;OL CHIKI LETTER OV;Lo;0;L;;;;;N;;;;;
1C77;OL CHIKI LETTER OH;Lo;0;L;;;;;N;;;;;
1C78;OL CHIKI MU TTUDDAG;Lm;0;L;;;;;N;;;;;
1C79;OL CHIKI GAAHLAA TTUDDAAG;Lm;0;L;;;;;N;;;;;
1C7A;OL CHIKI MU-GAAHLAA TTUDDAAG;Lm;0;L;;;;;N;;;;;
1C7B;OL CHIKI RELAA;Lm;0;L;;;;;N;;;;;
1C7C;OL CHIKI PHAARKAA;Lm;0;L;;;;;N;;;;;
1C7D;OL CHIKI AHAD;Lm;0;L;;;;;N;;;;;
1C7E;OL CHIKI PUNCTUATION MUCAAD;Po;0;L;;;;;N;;;;;
1C7F;OL CHIKI PUNCTUATION DOUBLE MUCAAD;Po;0;L;;;;;N;;;;;

```

References



Hembram, S. M., et al. 1972. Adibasi Ol script = ... Hembram, Smt. Mary. 1993. Ol Ita. Konay Lal Tudu. 1978. Ol cemed. Murmu, Nilamani,ed. 1998. Bhañj Parāyni. Murmu, Raghunath. 1972. Ranar. Zide, Norman. 1996. "Scripts for Munda languages", in Peter T. Daniels and William Bright, eds. The world's writing systems.

Acknowledgements

This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley).

Figure 1: Sample from Raghunath Murmu's Ranar showing the alphabet. The page is split into two columns. The left column contains text in Santali script with Devanagari comments. The right column contains a grid of Santali characters with their corresponding Devanagari equivalents written below them. At the top right, there is a handwritten note: 'Munda Adibasi Grant Languages'.

Figure 1. Sample from Raghunath Murmu’s Ranar, showing the alphabet; the Devanagari comments are written in by a previous owner of the booklet.


 ଓଡ଼ିଆ ଗ୍ରନ୍ଥକରଣ
 
OL CHIKI

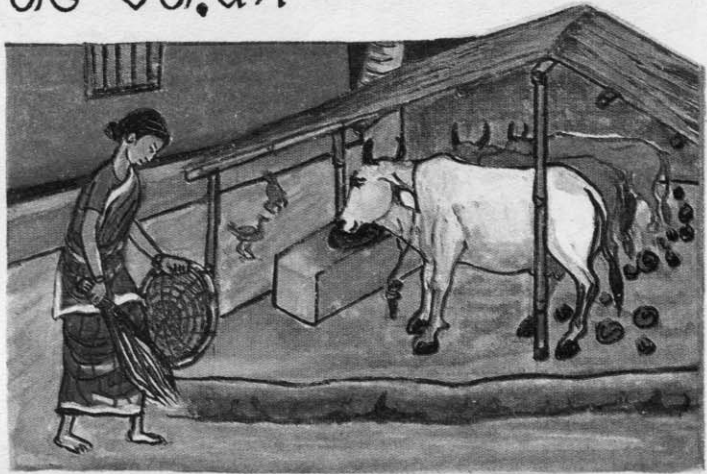
ୠ	ୡ	ୢ	ୣ	୤
ଅ (ଅତ୍) ଅ (ଅଗ୍) ମ (ଅଂ) ଲ (ଲ୍)				
ଞ (ଞତ୍) ଢ (ଞଗ୍) ଣ (ଞଂ) ଲ୍ (ଲ୍)				
ଠ (ଠତ୍) ଡ (ଠଗ୍) ଣ (ଠଂ) ଲ୍ (ଲ୍)				
O (O) OT (T) OG (G) ON (N) OL (L)				
ୠ	ୡ	ୢ	ୣ	୤
ଆ (ଆକ୍) କ (ଆଜ୍) ଞ (ଆମ୍) ଶ (ଆବ୍) ବ				
ଞା (ଞାକ୍) କ (ଞାଜ୍) ଞ (ଞାମ୍) ଶ (ଞାବ୍) ଶ				
ଞଠ (ଞଠକ୍) କ (ଞଠାଜ୍) ଞ (ଞଠାମ୍) ଶ (ଞଠାବ୍) ଶ				
AA (A) AAK (K) AAJ (J) AAM (M) AAW (W)				
ୠ	ୡ	ୢ	ୣ	୤
ଝ (ଝି) ଞ (ଝି) ଞ (ଝି) ଞ (ଝି) ଞ (ଝି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
IS (S) IH (H) IMG (G) IR (R)				
ୠ	ୡ	ୢ	ୣ	୤
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
U (U) UCH (CH) UDD (D) UUN (N) UY (Y)				
ୠ	ୡ	ୢ	ୣ	୤
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
AE (E) AP (P) AD (D) AN (N) AR (R)				
ୠ	ୡ	ୢ	ୣ	୤
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି) ଞ (ଞି)				
O (O) OT (T) (B) (B) ONG (G) ONH (H)				

Printed at : PAPPU PRINTEERS Rairangpur, Phone : 2162

Figure 2. Sample from the magazine *Bhanj Parayni*, showing alphabet with Devanagari, Bengali, Oriya, and Latin transliterations.

උරුමයේ වගකීම

උරුමයේ වගකීම
 වගකීම වගකීම
 වගකීම වගකීම වගකීම
 වගකීම වගකීම
 වගකීම වගකීම
 වගකීම වගකීම වගකීම



වගකීම වගකීම
 වගකීම වගකීම වගකීම
 වගකීම වගකීම වගකීම
 වගකීම වගකීම වගකීම
 වගකීම වගකීම වගකීම

වගකීම වගකීම
 වගකීම වගකීම වගකීම
 වගකීම වගකීම වගකීම වගකීම
 වගකීම වගකීම වගකීම
 වගකීම වගකීම වගකීම
 වගකීම වගකීම වගකීම



Figure 3. Sample from the children's primer *Al Itā*, showing use of PUNCTUATION MUCAAD and PUNCTUATION DOUBLE MUCAAD.

オル・チキ文字 アルファベット表

翻字	o	a	i	u	e	o	t	k'	n
音価	o	a	i	u	e	o	t	k/g	ŋ
翻字	k	c'	m	w	s	h	ñ	r	c
音価	k	c/dʒ	m	w/v	s	h-ʔ	ɲ	r	c
翻字	ɲ	y	p	d	n	r	t	p'	ɰ
音価	ɲ	j	p	d	n	ɾ	t	p/b	ɰ

:: 補助記号 ::

破裂音を伴わない 喉頭化音の有声閉 鎖音化を表わす	母音の次に下点で 書き母音が低音調 であることを示す	母音の次に上点で 書き母音を鼻母音 化する	母音の長母音化	破裂音を伴わない 喉頭化音とその他 の音との境界を示す

Figure 5. Discussion of Ol Chiki from the *Senseido Encyclopaedia of Linguistics*.

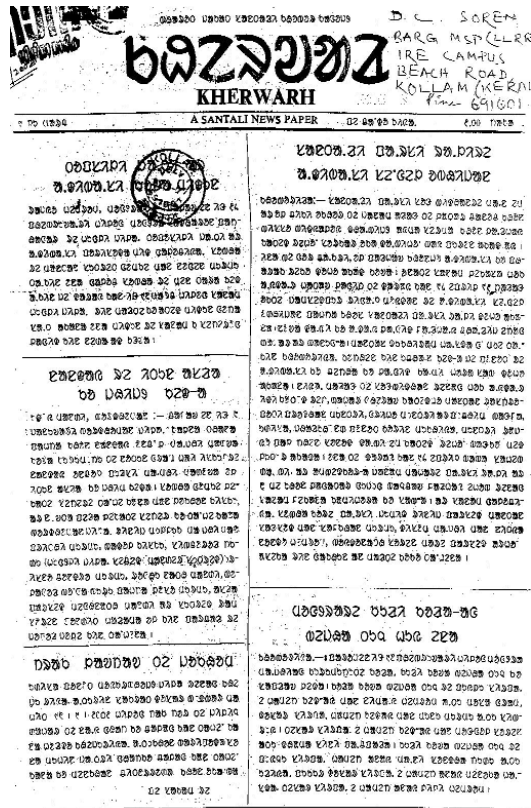


Figure 6. Sample from the newspaper Kherwarh.

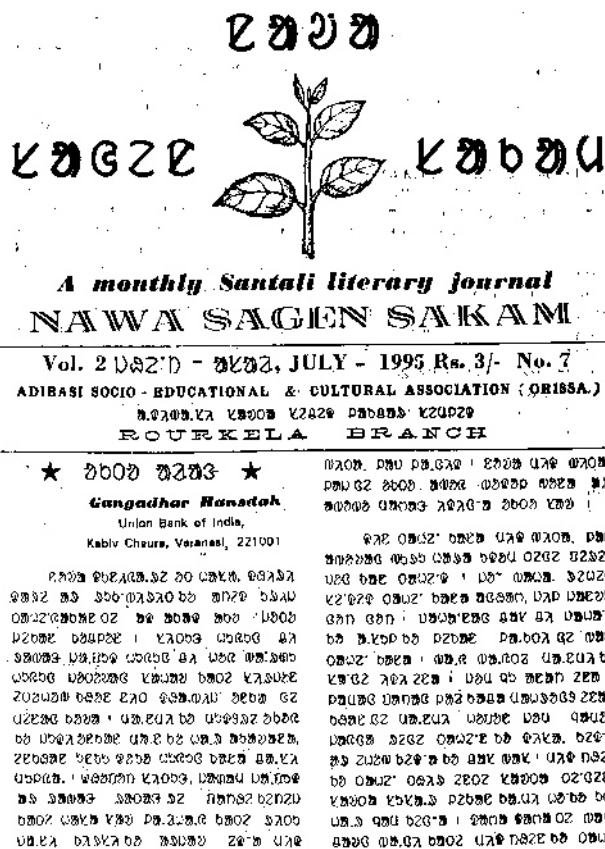


Figure 7. Sample from the literary journal Nawa Sagan Sakam.

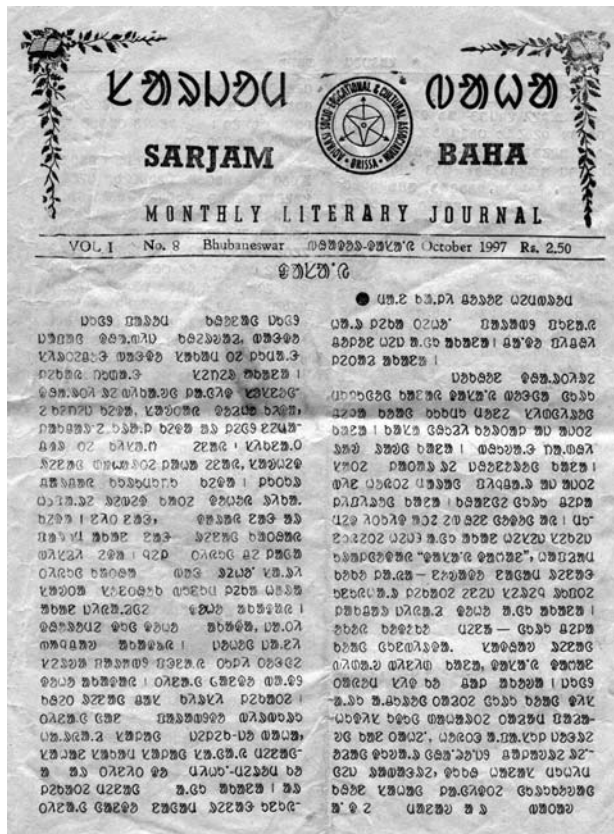


Figure 8. Sample from the literary journal *Sarjam Baha*.

Key Board Mapping for ዎፆ ዌገባገ

~ re '	! 1፻	@ 2፯	# 3፯	\$ 4፭	% 5፭	^ 6፻	& 7 2	* 8፯	(9፭) 00	_ ph -	+ =	←
Tab	ባ Q	ዎ W	ጊ E	ጃ R	ዐ T ስ	ፎ Y	ኃ U	ገ I	ፎ O ፆ	ፆ P	{ [}]	
Caps Lock	ገ A	ጊ S	ፎ D ፆ	ፎ F	ፆ G	ፆ H ፆ	ፆ J	ፆ K	ፆ L	: mg ;	" ,	↵	
Shift	ገ Z	ዎ X	ፆ C	ዎ V	ፆ B	mu N ፯	ፆ M ፆ	< ,	> . ga	? /	Shift		
Ctrl	Alt	Wesanthals E-Group http://www.wesanthals.org						Alt	Ctrl				

Note: ga = gəhlə / gəhle, mg = mu- gəhlə / gəhle, re = rela, ph = pharka

Figure 9. Ol Chiki keyboard layout.

TABLE XXX - Row 1C: OL CHIKI

	1C5	1C6	1C7
0	0	b	⓪
1	ƚ	ɔ	ɛ
2	ɛ	ɔ	ɔ
3	ɛ	ɔ	ɔ
4	ɛ	ɔ	ɔ
5	ɛ	ɔ	⓪
6	ɛ	ɔ	ɛ
7	ɛ	ɔ	ɛ
8	ɛ	ɔ	.
9	ɛ	ɔ	.
A	ɛ	ɔ	:
B	0	⓪	~
C	ɛ	ɔ	-
D	ɛ	ɔ	ɔ
E	ɔ	ɛ	ɛ
F	ɛ	ɔ	

G = 00
P = 00

TABLE XXX - Row 1C: OL CHIKI

hex	Name
50	OL CHIKI DIGIT ZERO
51	OL CHIKI DIGIT ONE
52	OL CHIKI DIGIT TWO
53	OL CHIKI DIGIT THREE
54	OL CHIKI DIGIT FOUR
55	OL CHIKI DIGIT FIVE
56	OL CHIKI DIGIT SIX
57	OL CHIKI DIGIT SEVEN
58	OL CHIKI DIGIT EIGHT
59	OL CHIKI DIGIT NINE
5A	OL CHIKI LETTER LA
5B	OL CHIKI LETTER AT
5C	OL CHIKI LETTER AG
5D	OL CHIKI LETTER ANG
5E	OL CHIKI LETTER AL
5F	OL CHIKI LETTER LAA
60	OL CHIKI LETTER AAK
61	OL CHIKI LETTER AAJ
62	OL CHIKI LETTER AAM
63	OL CHIKI LETTER AAW
64	OL CHIKI LETTER LI
65	OL CHIKI LETTER IS
66	OL CHIKI LETTER IH
67	OL CHIKI LETTER INY
68	OL CHIKI LETTER IR
69	OL CHIKI LETTER LU
6A	OL CHIKI LETTER UC
6B	OL CHIKI LETTER UD
6C	OL CHIKI LETTER UNN
6D	OL CHIKI LETTER UY
6E	OL CHIKI LETTER LE
6F	OL CHIKI LETTER EP
70	OL CHIKI LETTER EDD
71	OL CHIKI LETTER EN
72	OL CHIKI LETTER ERR
73	OL CHIKI LETTER LO
74	OL CHIKI LETTER OTT
75	OL CHIKI LETTER OB
76	OL CHIKI LETTER OV
77	OL CHIKI LETTER OH
78	OL CHIKI MU TTUDDAG
79	OL CHIKI GAAHLAA TTUDDAAG
7A	OL CHIKI MU-GAAHLAA TTUDDAAG
7B	OL CHIKI RELAA
7C	OL CHIKI PHAARKAA
7D	OL CHIKI AHAD
7E	OL CHIKI PUNCTUATION MUCAAD
7F	OL CHIKI PUNCTUATION DOUBLE MUCAAD

hex	Name
-----	------

A. Administrative

1. Title

Final proposal to encode the Ol Chiki script in the UCS.

2. Requester's name

Michael Everson

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2005-09-21

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

Proposed name of script

Ol Chiki.

1b. The proposal is for addition of character(s) to an existing block

No.

1b. Name of the existing block

2. Number of characters in proposal

48

3. Proposed category (see section II, Character Categories)

Category A.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 1.

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Simple alphabetic script.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson (Everttype). TrueType.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson (Everttype). Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. 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.

9. 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.

See Unicode properties above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes. See N1956 and N2505.

2a. 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.

2b. If YES, with whom?

Jeyakumar C. K. is a user himself, and has been in touch with Mr Kubendiran, editor of *ᱛᱟᱹᱨᱩᱵᱟᱹᱨᱟᱹ Bhāṣābhīmāni*.

2c. If YES, available relevant documents

N/A.

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

Yes. Speakers of the Santali language, whose population is 5,800,000, with 25%–50% literacy, according to the SIL *Ethnologue*.

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

Common. To write the Santali language. Latin, Devanagari, Bengali, and Oriya scripts have also been used to write Santali.

4b. Reference

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

Yes.

5b. If YES, where?

In primary and adult education (general use).

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Contemporary use and accordance with the Roadmap.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference