

Proposal to encode Tolong Siki in Unicode

Anshuman Pandey

pandey@umich.edu
pandey.github.io/unicode

January 5, 2023

1 Introduction

Tolong Siki is an alphabet used for writing Kurukh (ISO 639-3: *kru*), a Dravidian language spoken by 2.28 million persons in India, primarily in the states of Jharkhand, Bihar, and West Bengal. It was invented in 1988 by Dr. Narayan Oraon of Gumla district, Jharkhand. The script was developed over the years in collaboration with Francis Ekka, the former director of the Central Institute of Indian Languages (CIIL), Mysore; Padma Shri Ram Dayal Munda, the former Vice Chancellor of Ranchi University; and Dr. Nirmal Minz, founder principal, Gossner College, Ranchi. Tolong Siki was formally published on May 15, 1999.

In 2010, I submitted “Preliminary Proposal to Encode the Tolong Siki Script in the UCS” (N3811 L2/10-106) to the Unicode Technical Committee. At that time, relatively little was known about Tolong Siki outside of the Kurukh community. So, the purpose of that document was to inform an international audience about the existence and usage of Tolong Siki. It was not a formal request to include the script in The Unicode Standard. Also, as Tolong Siki is a newly-invented script, there was a need to measure the suitability of including the script in Unicode. Tolong Siki is one of more than twenty scripts that have been invented in the past thirty years. Some of these scripts are used today, while many are not. Also, those in use have display varying degrees of the following attributes that help to determine suitability of encoding a script in Unicode: stability of the script and character repertoire; acceptance by the user community; usage in publications; technological development; and institutional support or official recognition.

I have monitored the usage and development of Tolong Siki over the past twelve years. The script’s creator, Narayan Oraon, has provided updates through the years. I have been contacted regarding the status of Tolong Siki in Unicode by numerous members of the user community, among whom Ashwin Kumar Kispotta has provided me with substantial document of the script. Based on this information, it is clear today that Tolong Siki has achieved the aforementioned criteria.

1. *Stability* Tolong Siki has remained highly stable in its structure and basic repertoire from the time of its official release in 1999. The vowel and consonant letters, as well as digits, are unchanged in both inventory, semantics, and form. Latin punctuation continue to be used. Changes have been made to the set of diacritics signs, namely the addition of new signs that extend the ability of the script to represent new sounds.

2. *Acceptance* The script is very much alive and enjoys increasing usage within the Kurukh-speaking communities in India. It is taught to Kurukh speakers in primary schools, cultural organizations, and through community publications.
3. *Development* Two digitized fonts are available for Tolong Siki. “Singi Dai” was designed by Nemhas Ekka of Lohardaga district, Jharkhand and released in March 2007. The second, “Kelly Tolong” was designed by Kislaya IT Services (KITS) and released in April 2007. The “Kelly Tolong” font is used in this proposal, and is the font used in the *Kurukh Times* and most books.
4. *Publications* Primers for the script have been published both by community organizations and state education boards, and in different languages. Various Kurukh-Hindi primers have been published by Narayan Oraon (see figs. 1-3 for excerpts from a primer from 2021) A Kurukh-Bengali primer was published in 2016 by Mahesh Minz (see figs. 12–13). A newsletter called the *Kurukh Times* is published regularly (see fig 9, 10). The script is used on covers of books (see fig. 14).
5. *Institutional Support* The Tribal Cultural Society, a philanthropic organization of the Tata Steel Foundation, provides support for publications in Tolong Siki (see book covers in fig. 14).
6. *Official Recognition* Tolong Siki has received formal recognition in several states as one of the official scripts for the Kurukh language (see, for example, the reference in fig. 6). On April 3, 2007, it was recognized as the formal script for Kurukh by the government of Jharkhand (*The Telegraph* 2007). In 2009, the state of Jharkhand permitted usage of the script for secondary school examinations (see fig. 7). In February 2016, Jharkhand granted permission for the script to be used for writing university examinations (see fig. 8). In 2017, the government of West Bengal recognized Tolong Siki as an official script for Kurukh (*The Hindu* 2017).

Accordingly, this document is a formal proposal to include Tolong Siki in The Unicode Standard. Given the stability of the script, the core repertoire presented in L2/10-106 remains unchanged. Major changes to the proposed encoding include:

- Inclusion in the repertoire of the auspicious sign **ꠘꠤ** *ūggū*
- Documentation of additional diacritics used for modifying consonant and vowel sounds
- Unification of Tolong Siki diacritics with common combining signs already encoded in Unicode

2 The Script

Tolong Siki (ᱵᱤᱨᱫᱟᱹᱜᱟᱲ ᱵᱤᱨᱫᱟᱹᱜᱟᱲ) is an alphabetic, left-to-right script. It is an invented writing system, but it possesses structural and graphical features similar to other scripts that evolved organically from Brahmi.

There are 6 vowel letters (ᱵᱤᱨᱫᱟᱹᱜᱟᱲ ᱵᱤᱨᱫᱟᱹᱜᱟᱲ *sarah tōr*):

ᱵ	ᱶ	ᱷ	ᱸ	ᱹ	ᱺ
<i>i</i>	<i>e</i>	<i>u</i>	<i>o</i>	<i>a</i>	<i>ā</i>

There are 35 consonants (ᱵᱤᱨᱫᱟᱹᱜᱟᱲ ᱵᱤᱨᱫᱟᱹᱜᱟᱲ *harah tōr*):

ᱰ	ᱱ	ᱲ	ᱳ	ᱴ
<i>p</i>	<i>ph</i>	<i>b</i>	<i>bh</i>	<i>m</i>
ᱵ	ᱶ	ᱷ	ᱸ	ᱹ
<i>t</i>	<i>th</i>	<i>d</i>	<i>dh</i>	<i>n</i>
ᱺ	ᱻ	ᱼ	ᱽ	᱾
<i>ṭ</i>	<i>ṭh</i>	<i>ḍ</i>	<i>ḍh</i>	<i>ṇ</i>
᱿	᱀	᱁	᱂	᱃
<i>c</i>	<i>ch</i>	<i>j</i>	<i>jh</i>	<i>ñ</i>
᱄	᱅	᱆	᱇	᱈
<i>k</i>	<i>kh</i>	<i>g</i>	<i>gh</i>	<i>ṅ</i>
᱉	᱊	᱋	᱌	ᱍ
<i>y</i>	<i>r</i>	<i>l</i>	<i>v</i>	<i>ṛ</i>
ᱎ	ᱏ	᱐	᱑	᱒
<i>s</i>	<i>h</i>	<i>x</i>	<i>ṣ</i>	<i>ṛh</i>

As per the chart in fig. 5, the letter ᱎ *ha* has the variant form ᱏ. But, that is no longer used.

The arrangement of the consonant repertoire into classes based on the points of articulation is based on the Brahmi pattern. However, the ordering of classes for the base consonant letters is reversed in Tolong Siki, such that the first class consists of letters representing labial sounds instead of velars as in the Brahmi model. Tolong Siki consonant letters are alphabetic, so they do not possess the inherent *a*.









There is a full set of decimal digits:

0 1 2 3 4 5 6 7 8 9








There is an auspicious sign, which is a letter-like character similar to om :

𐎶𐎵
niggū

Tolong Siki has several diacritics and special signs. The original set of signs is:

 <i>selā</i> vowel length sign	 <i>mitalā</i> consonant nasal sign	 <i>talā / hecakā</i> glottal stop	 <i>ghetalā</i> syllable separator	 <i>sulā</i> foreign sound sign	
					
 <i>dugā</i> consonant modifier					 <i>pugā</i> consonant modifier

Additional diacritics were added in 2015:

 <i>evā</i> vowel nasalization	 <i>revā</i> cluster-final /r/	 <i>ilā</i> consonant modifier	 <i>uphā</i> consonant modifier	 <i>phavā</i> vowel modifier
 <i>sēva</i> vowel nasalization length mark		 <i>sengā</i> consonant nasalization length mark		

There are no script-specific signs of punctuation; conventional Latin signs are commonly used. In some cases, Devanagari danda-s are used as well.

·	,	;	!	?	-
<i>gahala ṭurā</i>	<i>ucari ṭurā</i>	<i>micari ṭurā</i>			<i>pacha</i>
full stop	comma	semi-colon	exclamation mark	question mark	hyphen

2.1 Orthography

Vowel lengthening is indicated using **ː** *selā* (alternately depicted as ː, see fig. 11), which is a letter-like mark placed after a vowel letter:

Ṗː	Ṡː	Ṣː	Ṥː	Ṧː	Ṣː
/i:/	/e:/	/u:/	/o:/	/a:/	/ā:/

The a /ʔ/ glottal stop is indicated using the sign **l**, a vertical bar which was formerly called *talā*, but is now called *hecakā*. It is a letter-like mark placed after a vowel letter:

Ṗl	Ṡl	Ṣl	Ṥl	Ṧl	Ṣl
/iʔ/	/eʔ/	/uʔ/	/oʔ/	/eʔ/	/ā:ʔ/

Nasalization is indicated using **˙** *mitalā* and **˜** *evā*. The *mitalā* is used for indicating consonant nasals. This sign is synonymous with the class nasal letter of the following consonant. It is placed above the base letter. Below are examples of the same Kurukh word written with *mitalā* and with the respective class nasal:

ṖṢṢ	ṢṢṢ	ṢṢṢ	ṢṢṢ
ṖṢṢ	ṢṢṢ	ṢṢṢ	ṢṢṢ
/kaṇk/	/cent ^h /	/paṇc/	/ment ^h /

Vowel nasalization was also marked using **˙** *mitalā*. But, in 2015, **˜** *evā* was introduced specifically for this purpose. The *evā* is placed above a vowel letter. If there is another accompanying sign that is a spacing mark, such as **ː** *selā*, the *evā* is positioned above the base.

Ṗ˜	Ṗ˜ː	Ṡ˜	Ṡ˜ː	Ṣ˜	Ṣ˜ː	Ṥ˜	Ṥ˜ː	Ṧ˜	Ṧ˜ː	Ṣ˜	Ṣ˜ː
/ĩ/	/ĩ:/	/ẽ/	/ẽ:/	/ũ/	/ũ:/	/õ/	/õ:/	/ã/	/ã:/	/ã/	/ã:/

The **ʼ** *ghetalā* indicates a syllable boundary within a word:

ṢṢṢṢṢ	ṢṢṢʼṢṢṢ
/bannā/	/banʼnā/

The **ˆ** *phavā* is used for representing English vowel sounds, such as /ɔ/, semantically similar to the **ˆ** U+0945 DEVANAGARI SIGN CANDRA E and **ˆ** U+0949 DEVANAGARI SIGN CANDRA O:

ṢṢṢṢṢ
/ɔfis/

Sounds not native to Kurukh are indicated using a variety of signs. There are two sets of these *nukta*-like signs, which are reflective of orthographic conventions of different phases in the development of Tolong Siki. Until 2015, the signs **ṣ** *sulā*, **ṣ** *dugā*, and **ṣ** *pugā* were used. The *sulā* is the common *nukta*, written beneath a letter to represent sounds in Sanskrit and Urdu words. The *dugā* and *pugā* are used for more specialized transcriptions. In 2015, a committee overseeing the development of Tolong Siki switched from **ṣ** *dugā* and **ṣ** *pugā* to the signs **ṣ** *ilā*, and **ṣ** *uphā*, respectively.

Pre-2015	𑌓	𑌔	𑌕	𑌖	𑌗	𑌘	𑌙	𑌚
Post-2015	𑌓	𑌔	𑌕	𑌖	𑌗	𑌘	𑌙	𑌚
Value (Devanagari)	क्	ग्	क्ष	ज्	फ्	व्	श्	ष्

Although 𑌓 *r* intrinsically represents /r/, a syllable-final /r/ is also written using the below-base sign 𑌚 *revā*. The *revā* is also another innovation introduced in 2015.

𑌓	𑌔	𑌕	𑌖	𑌗
/kr/	/tr/	/dr/	/ʃr/	/kri/
क्	त्र	द्र	स्	कृ / क्रि

3 Approach to Encoding Letter-Like Signs and Combining Signs

The fourteen letter-like and combining signs used throughout the different phases of the development of in Tolong Siki resemble characters that are already encoded in Unicode.

The original set of signs are as follows, with their analogues in Unicode:

	Name	Function in Tolong Siki	Analogous sign in Unicode
:	<i>selā</i>	vowel-length mark	: U+003A COLON
◌̣	<i>mitalā</i>	consonant nasal sign	◌̣ U+0307 COMBINING DOT ABOVE
	<i>hecakā</i>	glottal stop	U+007C VERTICAL BAR
'	<i>ghetalā</i>	syllable separator	' U+02BC MODIFIER LETTER APOSTROPHE
◌̤	<i>sulā</i>	modifier	◌̤ U+0323 COMBINING DOT BELOW
◌̄	<i>dugā</i>	consonant modifier	◌̄ U+0304 COMBINING MACRON
◌̅	<i>pugā</i>	consonant modifier	◌̅ U+0331 COMBINING MACRON BELOW

Additional and replacement signs used from 2015 onwards, with their analogues in Unicode, are as follows:

	Name	Function in Tolong Siki	Analogous sign in Unicode
◌̃	<i>evā</i>	nasalization	◌̃ U+0303 COMBINING TILDE
◌̣̣	<i>revā</i>	cluster-final /r/	◌̣̣ U+0330 COMBINING TILDE BELOW
◌̥̥	<i>ilā</i>	consonant modifier	◌̥̥ U+0308 COMBINING DIARESIS
◌̥̥̥	<i>uphā</i>	consonant modifier	◌̥̥̥ U+0324 COMBINING DIARESIS BELOW
◌̥̥̥̥	<i>phavā</i>	vowel modifier	◌̥̥̥̥ U+0306 COMBINING BREVE
◌̥̥̥̥̥	<i>sēva</i>	vowel modifier	◌̥̥̥̥̥ U+0304 COMBINING MACRON + ◌̃ U+0303 COMBINING TILDE
◌̥̥̥̥̥̥	<i>sengā</i>	vowel modifier	◌̥̥̥̥̥̥ U+0304 COMBINING MACRON + ◌̣̣ U+0307 COMBINING DOT ABOVE

Given the fact that all Tolong Siki diacritics are analogous with common combining signs that already exist in Unicode, it is practical to unify them with the existing Unicode characters. This means that instead of encoding, for example, *mitalā* as a unique character for Tolong Siki, the ◌ U+0307 COMBINING DOT ABOVE would be used instead. Similarly, the ◌ U+0303 COMBINING TILDE would be used for *evā*. Designers of Tolong Siki fonts would then be responsible for ensuring that the glyphs for COMBINING DOT ABOVE or COMBINING TILDE would align with the overall aesthetics of the font.

Two characters are exceptions to the above; these should be encoding as distinct characters for Tolong Siki:

- The letter-like *selā*: basic vowel lengthening, also has variant form *ḥ*; usage of COLON would and *selā* within the same text would hinder such distinctions
- The letter-like *hecaḥ*: although visually similar to VERTICAL BAR, it represents a glottal stop and is, therefore, semantic distinct

The recommendation to use common combining signs for Tolong Siki is not due to any specific aspects of the script; rather, it is based on similar practices in Cyrillic, Greek, Osage, and a whole host of others scripts, which regularly employ common combining signs for various purposes. The usage of these signs may have distinctive meanings in each of these individual scripts. But, the nature of the common combining signs offers users the ability and freedom to use diacritics freely.

In order to mix Latin text with diacritics and Tolong Siki — for example, for purposes of transliteration — a separate font would be used for each script. Accordingly, in an input method for Tolong Siki, the keys for the diacritics would be mapped to the respective code point for the combining signs.

The recommendation to unify combining signs is based on a similar approach taken for the encoding for the Sunuwar script in Unicode (see Pandey 2021: L2/21-157R). For Tolong Siki, the available materials strongly suggest that users and font designers are agreeable to using Latin diacritics. For example, in fig. 3 it is apparent that the nasalization sign *evā* is semantically and graphically equivalent to ◌ U+0303 COMBINING TILDE.

Additional notes:

- As the ◌ *pugā* does not join when there is sequential usage of the sign, it is appropriate to use ◌ U+0331 COMBINING MACRON BELOW instead of ◌ U+0331 COMBINING LOW LINE.
- The *hecaḥ* is nearly, if not, completely identical to the digit *one*. The two are distinguishable by context.

4 Proposed Character Repertoire

At present, the proposed repertoire for Tolong Siki consists of 54 characters: 41 letters, 2 letter-like marks, 1 auspicious sign, 10 digits:

Letters	
Character	Proposed Unicode Character Name
ᱱ	TOLONG SIKI LETTER I
ᱲ	TOLONG SIKI LETTER E
ᱳ	TOLONG SIKI LETTER U
ᱴ	TOLONG SIKI LETTER O
ᱵ	TOLONG SIKI LETTER A
ᱶ	TOLONG SIKI LETTER AA
ᱷ	TOLONG SIKI LETTER P
ᱸ	TOLONG SIKI LETTER PH
ᱹ	TOLONG SIKI LETTER B
ᱺ	TOLONG SIKI LETTER BH
ᱻ	TOLONG SIKI LETTER M
ᱼ	TOLONG SIKI LETTER T
ᱽ	TOLONG SIKI LETTER TH
᱾	TOLONG SIKI LETTER D
᱿	TOLONG SIKI LETTER DH
ᱠ	TOLONG SIKI LETTER N
ᱡ	TOLONG SIKI LETTER TT
ᱢ	TOLONG SIKI LETTER TTH
ᱣ	TOLONG SIKI LETTER DD
ᱤ	TOLONG SIKI LETTER DDH
ᱥ	TOLONG SIKI LETTER NN
ᱦ	TOLONG SIKI LETTER C
ᱧ	TOLONG SIKI LETTER CH
ᱨ	TOLONG SIKI LETTER J
ᱩ	TOLONG SIKI LETTER JH
ᱪ	TOLONG SIKI LETTER NY
ᱫ	TOLONG SIKI LETTER K
ᱬ	TOLONG SIKI LETTER KH
ᱭ	TOLONG SIKI LETTER G
ᱮ	TOLONG SIKI LETTER GH
ᱯ	TOLONG SIKI LETTER NG

Ɔ	TOLONG SIKI LETTER Y
ᄒ	TOLONG SIKI LETTER R
ᄓ	TOLONG SIKI LETTER L
ᄔ	TOLONG SIKI LETTER V
ᄕ	TOLONG SIKI LETTER NNY
ᄖ	TOLONG SIKI LETTER S
ᄗ	TOLONG SIKI LETTER H
ᄘ	TOLONG SIKI LETTER X
ᄙ	TOLONG SIKI LETTER RR
ᄚ	TOLONG SIKI LETTER RRH

Modifier Signs

Character	Proposed Unicode Character Name
:	TOLONG SIKI SIGN SELA
	TOLONG SIKI SIGN HECACA

Auspicious Sign

Character	Proposed Unicode Character Name
ᄛ	TOLONG SIKI UNGGA

Digits

Character	Proposed Unicode Character Name
ᄜ	TOLONG SIKI DIGIT ZERO
ᄝ	TOLONG SIKI DIGIT ONE
ᄞ	TOLONG SIKI DIGIT TWO
ᄟ	TOLONG SIKI DIGIT THREE
ᄠ	TOLONG SIKI DIGIT FOUR
ᄡ	TOLONG SIKI DIGIT FIVE
ᄢ	TOLONG SIKI DIGIT SIX
ᄣ	TOLONG SIKI DIGIT SEVEN
ᄤ	TOLONG SIKI DIGIT EIGHT
ᄥ	TOLONG SIKI DIGIT NINE

4.1 Collation

The collating order for Tolong Siki is as follows:

ꞑ I < Ꞓ E < ꞓ U < ꞔ O < ꞕ A < Ꞗ AA < ꞗ P < Ꞙ PH < ꞙ B < Ꞛ BH < ꞛ M <
 Ꞝ T < ꞝ TH < Ꞟ D < ꞟ DH < Ꞡ N < ꞡ TT < Ꞣ TTH < ꞣ DD < Ꞥ DDH <
 ꞥ NN < Ꞧ C < ꞧ CH < Ꞩ J < ꞩ JH < Ɦ NY < Ɜ K < Ɡ KH < Ɬ G < Ɪ GH <
 ꞯ NG < Ʞ Y < Ʇ R < Ʝ L < Ꭓ V < Ꞵ NNY < ꞵ S < Ꞷ H < ꞷ X < Ꞹ RR <
 ꞹ RHA < Ꞻ UNGGA < ꞻ SELA < Ꞽ HECAKA

5 Character Data

Character Properties: UnicodeData.txt

```

11DB0;TOLONG SIKI LETTER I;Lo;0;L;;;;;N;;;;;
11DB1;TOLONG SIKI LETTER E;Lo;0;L;;;;;N;;;;;
11DB2;TOLONG SIKI LETTER U;Lo;0;L;;;;;N;;;;;
11DB3;TOLONG SIKI LETTER O;Lo;0;L;;;;;N;;;;;
11DB4;TOLONG SIKI LETTER A;Lo;0;L;;;;;N;;;;;
11DB5;TOLONG SIKI LETTER AA;Lo;0;L;;;;;N;;;;;
11DB6;TOLONG SIKI LETTER P;Lo;0;L;;;;;N;;;;;
11DB7;TOLONG SIKI LETTER PH;Lo;0;L;;;;;N;;;;;
11DB8;TOLONG SIKI LETTER B;Lo;0;L;;;;;N;;;;;
11DB9;TOLONG SIKI LETTER BH;Lo;0;L;;;;;N;;;;;
11DBA;TOLONG SIKI LETTER M;Lo;0;L;;;;;N;;;;;
11DBB;TOLONG SIKI LETTER T;Lo;0;L;;;;;N;;;;;
11DBC;TOLONG SIKI LETTER TH;Lo;0;L;;;;;N;;;;;
11DBD;TOLONG SIKI LETTER D;Lo;0;L;;;;;N;;;;;
11DBE;TOLONG SIKI LETTER DH;Lo;0;L;;;;;N;;;;;
11DBF;TOLONG SIKI LETTER N;Lo;0;L;;;;;N;;;;;
11DC0;TOLONG SIKI LETTER TT;Lo;0;L;;;;;N;;;;;
11DC1;TOLONG SIKI LETTER TTH;Lo;0;L;;;;;N;;;;;
11DC2;TOLONG SIKI LETTER DD;Lo;0;L;;;;;N;;;;;
11DC3;TOLONG SIKI LETTER DDH;Lo;0;L;;;;;N;;;;;
11DC4;TOLONG SIKI LETTER NN;Lo;0;L;;;;;N;;;;;
11DC5;TOLONG SIKI LETTER CA;Lo;0;L;;;;;N;;;;;
11DC6;TOLONG SIKI LETTER CH;Lo;0;L;;;;;N;;;;;
11DC7;TOLONG SIKI LETTER J;Lo;0;L;;;;;N;;;;;
11DC8;TOLONG SIKI LETTER JH;Lo;0;L;;;;;N;;;;;
11DC9;TOLONG SIKI LETTER NY;Lo;0;L;;;;;N;;;;;
11DCA;TOLONG SIKI LETTER K;Lo;0;L;;;;;N;;;;;
11DCB;TOLONG SIKI LETTER KH;Lo;0;L;;;;;N;;;;;
11DCC;TOLONG SIKI LETTER G;Lo;0;L;;;;;N;;;;;
11DCD;TOLONG SIKI LETTER GH;Lo;0;L;;;;;N;;;;;
11DCE;TOLONG SIKI LETTER NG;Lo;0;L;;;;;N;;;;;
11DCF;TOLONG SIKI LETTER Y;Lo;0;L;;;;;N;;;;;

```

```

11DD0;TOLONG SIKI LETTER R;Lo;0;L;;;;N;;;;;
11DD1;TOLONG SIKI LETTER L;Lo;0;L;;;;N;;;;;
11DD2;TOLONG SIKI LETTER V;Lo;0;L;;;;N;;;;;
11DD3;TOLONG SIKI LETTER NNY;Lo;0;L;;;;N;;;;;
11DD4;TOLONG SIKI LETTER S;Lo;0;L;;;;N;;;;;
11DD5;TOLONG SIKI LETTER H;Lo;0;L;;;;N;;;;;
11DD6;TOLONG SIKI LETTER X;Lo;0;L;;;;N;;;;;
11DD7;TOLONG SIKI LETTER RR;Lo;0;L;;;;N;;;;;
11DD8;TOLONG SIKI LETTER RRH;Lo;0;L;;;;N;;;;;
11DD9;TOLONG SIKI SIGN SELA;Lo;0;L;;;;N;;;;;
11DDA;TOLONG SIKI SIGN HECACA;Lo;0;L;;;;N;;;;;
11ddb;TOLONG SIKI UNGGA;Lo;0;L;;;;N;;;;;
11DE0;TOLONG SIKI DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
11DE1;TOLONG SIKI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
11DE2;TOLONG SIKI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
11DE3;TOLONG SIKI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
11DE4;TOLONG SIKI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
11DE5;TOLONG SIKI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
11DE6;TOLONG SIKI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
11DE7;TOLONG SIKI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
11DE8;TOLONG SIKI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
11DE9;TOLONG SIKI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;

```

Linebreaking Properties: LineBreak.txt

```

11DB0..11ddb;AL    # Lo  [44] TOLONG SIKI LETTER I..TOLONG SIKI UNGGA
11DE0..11DE9;NU    # Nd  [10] TOLONG SIKI DIGIT ZERO..TOLONG SIKI DIGIT NINE

```

Internal ‘Confusable’ Characters

```

11DDA;TOLONG SIKI SIGN HECACA ; 11DE1;TOLONG SIKI DIGIT ONE

```

6 References

- Ekka, Nemhas. 2007. “Tolong Siki: Basic rules of letters”. <http://www.kurukhworld.bravehost.com/>
- Krishnamurti, Bhadriraju. 2003. *The Dravidian Languages*. Cambridge: Cambridge University Press.
- Kurukhworld*. “Cultural Elements of Kurukh Community: Language & Literature”. 2008. http://www.kurukhworld.bravehost.com/english/lang_lit.htm
- Minz, Mahesh. 2016 *Tolong Siki Parichay*. Paschim Bang Sarkar: Janjati Vikash Vibhag.
- Singh, Anjana 2018. “Linguistic Politics and Kurukh Language Movement of the Oraons in Jharkhand”, *Journal of Adivasi and Indigenous Studies*, vol. VIII, no. 2 (August 2018), pp. 37–50. <http://joais.in/Journal/JAIS,%20August%202018,%20%20Vo1.%20VIII,%20No.%202.pdf>

- Singh, Shiv Sahay. 2017. “Revitalising a language”. *The Hindu*, March 2, 2017, Kolkata ed.
<https://www.thehindu.com/news/cities/kolkata/revitalising-a-language/article17395995.ece>
- Pandey, Anshuman. 2010. “Preliminary Proposal to Encode the Tolong Siki Script in the UCS” (WG2 N3811 L2/10-106). <https://www.unicode.org/L2/L2010/10106-n3811-tolongsiki.pdf>
- . 2021. “Proposal to encode the Sunuwar script in Unicode” L2/21-157R.
<https://www.unicode.org/L2/L2021/21157r-sunuwar.pdf>
- The Telegraph*. “Kurukh font launched”. Wednesday, April 04, 2007. Calcutta.
http://www.telegraphindia.com/1070404/asp/ranchi/story_7602864.asp
- TolongSiki.com*. <http://tolongsiki.com/>.























































7 Acknowledgments

I would like to express my gratitude to the following individuals who provided materials and information about Tolong Siki:


- Dr. Narayan Oraon
- Ashwin Kumar Kispotta
- Biswajit Mandal

The project to encode Tolong Siki has been made possible in part by funding from the Adopt-A-Character program of the Unicode Consortium and the effort is supervised by Deborah Anderson and Andrew Glass.




































It was also made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project PR-268710-20 (part of the Script Encoding Initiative at UC Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

	11DB	11DC	11DD	11DE
0	 11DB0	 11DC0	 11DD0	 11DE0
1	 11DB1	 11DC1	 11DD1	 11DE1
2	 11DB2	 11DC2	 11DD2	 11DE2
3	 11DB3	 11DC3	 11DD3	 11DE3
4	 11DB4	 11DC4	 11DD4	 11DE4
5	 11DB5	 11DC5	 11DD5	 11DE5
6	 11DB6	 11DC6	 11DD6	 11DE6
7	 11DB7	 11DC7	 11DD7	 11DE7
8	 11DB8	 11DC8	 11DD8	 11DE8
9	 11DB9	 11DC9	 11DD9	 11DE9
A	 11DBA	 11DCA	 11DDA	
B	 11DBB	 11DCB	 11ddb	
C	 11DBC	 11DCC		
D	 11DBD	 11DCD		
E	 11DBE	 11DCE		
F	 11DBF	 11DCF		

Vowel letters

11DB0		TOLONG SIKI LETTER I
11DB1		TOLONG SIKI LETTER E
11DB2		TOLONG SIKI LETTER U
11DB3		TOLONG SIKI LETTER O
11DB4		TOLONG SIKI LETTER A
11DB5		TOLONG SIKI LETTER AA

Consonant letters

11DB6		TOLONG SIKI LETTER PA
11DB7		TOLONG SIKI LETTER PHA
11DB8		TOLONG SIKI LETTER BA
11DB9		TOLONG SIKI LETTER BHA
11DBA		TOLONG SIKI LETTER MA
11DBB		TOLONG SIKI LETTER TA
11DBC		TOLONG SIKI LETTER THA
11DBD		TOLONG SIKI LETTER DA
11DBE		TOLONG SIKI LETTER DHA
11DBF		TOLONG SIKI LETTER NA
11DC0		TOLONG SIKI LETTER TTA
11DC1		TOLONG SIKI LETTER TTHA
11DC2		TOLONG SIKI LETTER DDA
11DC3		TOLONG SIKI LETTER DDHA
11DC4		TOLONG SIKI LETTER NNA
11DC5		TOLONG SIKI LETTER CA
11DC6		TOLONG SIKI LETTER CHA
11DC7		TOLONG SIKI LETTER JA
11DC8		TOLONG SIKI LETTER JHA
11DC9		TOLONG SIKI LETTER NYA
11DCA		TOLONG SIKI LETTER KA
11DCB		TOLONG SIKI LETTER KHA
11DCC		TOLONG SIKI LETTER GA
11DCD		TOLONG SIKI LETTER GHA
11DCE		TOLONG SIKI LETTER YA
11DCF		TOLONG SIKI LETTER RA
11DD0		TOLONG SIKI LETTER LA
11DD1		TOLONG SIKI LETTER VA
11DD2		TOLONG SIKI LETTER NYA
11DD3		TOLONG SIKI LETTER SA
11DD4		TOLONG SIKI LETTER HA
11DD5		TOLONG SIKI LETTER XA
11DD6		TOLONG SIKI LETTER RRA
11DD7		TOLONG SIKI LETTER RRHA
11DD8		TOLONG SIKI LETTER RRHA











Special marks

11DD9	:	TOLONG SIKI SIGN SELA = vowel length mark
11DDA		TOLONG SIKI SIGN HECAKA • also called tala = glottal stop

Auspicious sign

11ddb		TOLONG SIKI UNGGA
-------	---	-------------------

Digits

11DE0		TOLONG SIKI DIGIT ZERO
11DE1		TOLONG SIKI DIGIT ONE
11DE2		TOLONG SIKI DIGIT TWO
11DE3		TOLONG SIKI DIGIT THREE
11DE4		TOLONG SIKI DIGIT FOUR
11DE5		TOLONG SIKI DIGIT FIVE
11DE6		TOLONG SIKI DIGIT SIX
11DE7		TOLONG SIKI DIGIT SEVEN
11DE8		TOLONG SIKI DIGIT EIGHT
11DE9		TOLONG SIKI DIGIT NINE

ᱵᱚᱠᱟᱨ ᱵᱷᱚᱨ ᱵᱤᱨᱫᱟ ᱵᱚᱠᱟᱨ (कुँडुख़ तोलोड़ सिकि तोःड़पाब)

Tolong Siki Alphabet / तोलोड़ सिकि वर्णमाला

ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ (सरह तोड़) = स्वर वर्ण (Vowels)

ᱵ ᱵᱚ i ᱵᱚ ᱵᱚ e ᱵᱚ ᱵᱚ u ᱵᱚ ᱵᱚ o ᱵᱚ ᱵᱚ a ᱵᱚ ᱵᱚ ā

: (सेला) = लम्बी ध्वनि, ' (मितला) = नासिक्य व्यंजन सूचक, ' (घेतला) = शब्दखण्ड सूचक,

˜ (एवाँ) = नासिक्य स्वर सूचक, ˘ (रेवाँ) = लुप्ताकार र, । (हेचका) = अर्द्धस्वर अ/व्यंजन अ

ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ (हरह तोड़) = व्यंजन वर्ण (Consonants)

ᱵ ᱵᱚ p	ᱵ ᱵᱚ ph	ᱵ ᱵᱚ b	ᱵ ᱵᱚ bh	ᱵ ᱵᱚ m
ᱵ ᱵᱚ t	ᱵ ᱵᱚ th	ᱵ ᱵᱚ d	ᱵ ᱵᱚ dh	ᱵ ᱵᱚ n
ᱵ ᱵᱚ ᱵᱚ	ᱵ ᱵᱚ ᱵᱚ	ᱵ ᱵᱚ ᱵᱚ	ᱵ ᱵᱚ ᱵᱚ	ᱵ ᱵᱚ ᱵᱚ
ᱵ ᱵᱚ ch	ᱵ ᱵᱚ chh	ᱵ ᱵᱚ j	ᱵ ᱵᱚ jh	ᱵ ᱵᱚ ᱵᱚ
ᱵ ᱵᱚ k	ᱵ ᱵᱚ kh	ᱵ ᱵᱚ g	ᱵ ᱵᱚ gh	ᱵ ᱵᱚ ᱵᱚ
ᱵ ᱵᱚ y	ᱵ ᱵᱚ r	ᱵ ᱵᱚ l	ᱵ ᱵᱚ w	ᱵ ᱵᱚ ᱵᱚ
ᱵ ᱵᱚ s	ᱵ ᱵᱚ h	ᱵ ᱵᱚ x	ᱵ ᱵᱚ ᱵᱚ	ᱵ ᱵᱚ ᱵᱚ

ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ (लेख्या) = संख्या, Numerals

ᱵ	ᱵ	ᱵ	ᱵ	ᱵ	ᱵ	ᱵ	ᱵ	ᱵ	ᱵ	ᱵ
0	1	2	3	4	5	6	7	8	9	10

ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ

ख़ददर गही कत्था अरा कत्थ बेयॉख़ो = बच्चे की बोली और भाषा विज्ञान

ᱵᱚᱠᱟᱨ (पपा) = रोटी, ᱵᱚᱠᱟᱨ (बबा) = पिता, ᱵᱚᱠᱟᱨ (ममा) = मात ।

ᱵᱚᱠᱟᱨ (पल्ल) = दाँत, ᱵᱚᱠᱟᱨ (बई) = मुँह, ᱵᱚᱠᱟᱨ (मेलखा) = कण्ठ ।

ᱵᱚᱠᱟᱨ (ततखा) = जीभ, ᱵᱚᱠᱟᱨ (दुदु) = दूध, ᱵᱚᱠᱟᱨ (नरटी) = भोजन नली ।

ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ, ᱵᱚᱠᱟᱨ, ᱵᱚᱠᱟᱨ, ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ
ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ ᱵᱚᱠᱟᱨ

Figure 1: Chart of Tolong Siki (provided by Narayan Oraon, March 2021).

Tolong Siki							
3. कुँडुख हहस/कुँडुख भाषा की ध्वनियाँ (Kūṛux phonemes)							
1. कुँडुख कथा ही मूली सरह (कुँडुख भाषा की मूल स्वर स्वर ध्वनियाँ / Primary Vowel) :-							
ᱵᱤᱨᱫᱟ ᱵᱤᱨᱫᱟ / बड़ता सरह (मौखिक स्वर, Oral vowel)				ᱵᱤᱨᱫᱟ ᱵᱤᱨᱫᱟ / मुँडता सरह (नासिक्य स्वर, Nasal vowel)			
ᱵᱤᱨᱫᱟ ᱵᱤᱨᱫᱟ सन्नी सरह (हृस्व स्वर, short vowel)		ᱵᱤᱨᱫᱟ ᱵᱤᱨᱫᱟ दिगहा सरह (दीर्घ स्वर, long vowel)		ᱵᱤᱨᱫᱟ ᱵᱤᱨᱫᱟ सन्नी सरह (हृस्व स्वर, short vowel)		ᱵᱤᱨᱫᱟ ᱵᱤᱨᱫᱟ दिगहा सरह (दीर्घ स्वर, long vowel)	
ᱱ - इ, i		ᱱ: - इ: i:		ᱱ̃ - ई̃ i		ᱱ̃: - ई̃: i:	
ᱵ - ए, e		ᱵ: - ए: e:		ᱵ̃ - ऐ̃ e		ᱵ̃: - ऐ̃: e:	
ᱥ - उ, u		ᱥ: - उ: u:		ᱥ̃ - औ̃ u		ᱥ̃: - औ̃: u:	
ᱦ - ओ, o		ᱦ: - ओ: o:		ᱦ̃ - औ̃ o		ᱦ̃: - औ̃: o:	
ᱠ - अ, a		ᱠ: - अ: a:		ᱠ̃ - अँ̃ a		ᱠ̃: - अँ̃: a:	
ᱡ - आ, ā		ᱡ: - आ: ā:		ᱡ̃ - आँ̃ ā		ᱡ̃: - आँ̃: ā:	
2. कुँडुख कथा ही जोट्टा सरह (कुँडुख भाषा का संयुक्त स्वर / Diphthong) :-							
जोट्टा सरह – ᱱᱱ, ᱱᱥ, ᱱᱦ, ᱱᱦ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ, ᱱᱱ.							
अइ, अउ, ऐ, औ इउ, उइ, अय, अव, अँइ, अँउ, ऐँ, औँ, ईउ, उँइ, अँय, अँव.							
3. कुँडुख कथा ही मूली हरह (कुँडुख भाषा की मूल व्यंजन ध्वनियाँ / Primary Consonant) -							
अघोष / NonVoiced		सघोष / voiced		सघोष / voiced		अघोष/NonVoiced	
मौखिक अल्पप्राण Non aspirated	मौखिक महाप्राण Aspirated	मौखिक अल्पप्राण Non aspirated	मौखिक महाप्राण Aspirated	नासिक्य अल्पप्राण Non aspirated	मौखिक अल्पप्राण Non aspirated	मौखिक अल्पप्राण Non aspirated	
ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	- -	- -	
ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	- -	
ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	
ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	
ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	
ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	
ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	ᱱ ᱱ	

Figure 2: Chart of Tolong Siki (provided by Narayan Oraon, March 2021).

ଓଲଙ୍ଗ ସିକି ବାବଦ / vowel sound / स्वर ध्वनि	ଓଲଙ୍ଗ ସିକି ବାବଦ / consonant sound / व्यंजन ध्वनि
<p> p = i = इ - ପଇନା - इज्जो, इज्जो v = e = ए - वउता - एड़पा, एपटा ʒ = u = उ - ʒωता - उगता, उड़दु ɔ = o = ओ - ʒωता - ओसगा, ओरोख ɪ = a = अ - ɪʒता - अड़खा, अखड़ा ʈ = ā = आ - ʈɪ - आल, आरु p: = i: = ई - p:ɪता - इमा, कीड़ा v: = e: = ऐ - v:ʒता - ऐड़ा, नेला ʒ: = u: = ऊ - ʒ:ता - उरा, कूबी ɔ: = o: = औ - ɔ:ʒता - ओड़ा, ओड़ा ɪ: = a: = आ - ɪ:ʒ - अव, दव, कव ʈ: = ā: = आ - ʈ:ɪʒ - आली, आलो ṽ = ī = ई - ṽ - चिंचो, बिड़ना ṽ = ē = ऐ - ṽ - चेड़ा, गेंड़ा ṽ = ū = ऊ - ṽ - गुड़ना, कुँडुख ṽ = ō = ओ - ṽ - कौहड़ा, गौहड़ा ṽ = ā = अ - ṽ - अँटा, खँटी ṽ = ā = आ - ṽ - चाँड़, काँड़ p: = i: = ई - p: - बीड़ी, चींचो v: = e: = ऐ - v: - पेंड़े, हेंड़े ʒ: = u: = ऊ - ʒ: - कूँड़, पूँप ɔ: = o: = औ - ɔ: - कौड़ा, तौड़ा ɪ: = a: = आ - ɪ: - अँत, अँख, सँह ʈ: = ā: = आ - ʈ: - आँड़को, भाँड़ो ɪp = ai = अइ - ɪpɪता - कइला, चइला ɪʒ = au = अउ - ɪʒʒता - अउड़का, कउड़ी ɪv = ae = ऐ/अय - ɪvω - नैग, कैर ɪʒ = ao = औ/अव - ɪʒω - बौग, नौर pʒ = iu = इउ - pʒता - इउन्दा, चिउरा ʒp = ui = उइ - ʒpता - उइरा, कुइला ṽp = ai = अँइ - ṽp - कँइत, गँइता ṽʒ = au = अँउ - ṽʒ - कँउड़, झँउड़ ṽv = ae = ऐ/अँय - ṽvता - चँदना, पैड़ा ṽʒ = ao = औ/अँव - ṽʒ - चौर, भौर pʒ = iu = इँउ - pʒता - इँउरा, चिँउरा ʒp = ūi = उँइ - ʒp - पुँइदा, सुँइतु </p>	<p> ɪ = p = प - ɪpɪता - पपला, पचरी ʈ = ph = फ - ʈʈता - फँसरी, फटका ɪ = b = ब - ɪbɪʒ - बल्लु, बली ʈ = bh = भ - ʈbʒ - भदड़ी, भँड़की ɪ = m = म - ɪmω - मरग, मचा ɪ = t = त - ɪtता - ततखा, तड़री ɪ = th = थ - ɪth - थड़ा, थइला ɪ = d = द - ɪdʒ - दउली, दुदही ɪ = dh = ध - ɪdh - धनु, धरती ɪ = n = न - ɪnω - नड़गा, नड़हा ɪ = t̪ = ट - ɪt̪ता - टटखा, टमरस ɪ = t̪h = ठ - ɪt̪h - ठरकी, ठररा ɪ = d̪ = ड - ɪd̪ʒ - डउड़ा, डिगची ɪ = d̪h = ढ - ɪd̪h - ढरा, ढेरा ɪ = ɳ = ण - ɪɳता - भण्डा, डुण्डी ɪ = ch = च - ɪch - चलकी, चपटा ɪ = chh = छ - ɪchh - छटका, छलका ɪ = j = ज - ɪjʒ - जल्ली, जंगला, जट्टा ɪ = jh = झ - ɪjh - झमड़ा, झण्डा ɪ = ñ = ञ - ɪñ - चिजा, सिजा, कौयछा ɪ = k = क - ɪk - ककड़ो, कड़मा ɪ = kh = ख - ɪkh - खपरा, खरपा ɪ = g = ग - ɪg - गमछा, गगरा ɪ = gh = घ - ɪgh - घुघरी, घघरा ɪ = ŋ = ङ - ɪŋ - झिलडी, झोलडो, डोडा ɪ = y = य - ɪy - किया, जिया, पया, मया ɪ = r = र - ɪr - रसड़ी, रहटा, रोगे, रट्टा ɪ = l = ल - ɪl - लट्टु, लड़ड़, लवटो ɪ = v = व - ɪv - जवा, लवा, तवा, मड़वा ɪ = z = ज - ɪz - कोरजो, तूज, कुजी ɪ = s = स - ɪs - सगड़, सिकड़ी, सुरती ɪ = h = ह - ɪh - हड़का, हरिन, हँसली ɪ = x = ख - ɪx - खन्न, खज्ज, खँस ɪ = r̥ = ङ - ɪr̥ - गेड़े, घोड़ो, गड़ना ɪ = r̥h = ङ - ɪr̥h - गढ़े, बाढ़ी, उढ़ीयारना ɪ = ɹ = अ - ɪɹ - चिअना, नेअना, बअना </p>

Figure 3: Chart of Tolong Siki (provided by Narayan Oraon, March 2021).

ଟୋଲଂସିକି ଉପଲପ ଟୋଂସିକି

ତୋଲଂସିକି ବର୍ଣ୍ଣମାଳା

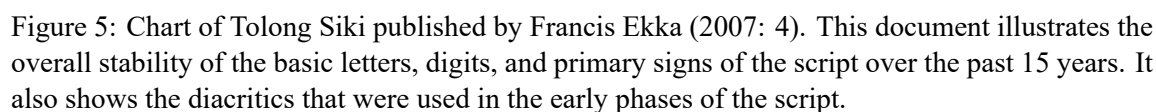
ଓଂଟିଂଟିଂ ଟୋଂସିକି = ସ୍ଵର ବର୍ଣ୍ଣ

p	v	ɹ	ɹ	ɹ	ɹ
i ଇ	e ଏ	u ଓ	o ଓ	a ଅ	a ଆ

ଟିଂଟିଂଟିଂ ଟୋଂସିକି = ବ୍ୟଞ୍ଜନ ବର୍ଣ୍ଣ

						ଲେଖା
୦	୦	୦	୦	୦	୦	୦
p ପ୍	ph ଫ୍	b ବ୍	bh ଖ୍	m ମ୍	୦	- 0
୧	୧	୧	୧	୧	୧	୧
t ଡ୍	th ଡ୍	d ଡ୍	dh ଡ୍	n ନ୍	୧	- 1
୨	୨	୨	୨	୨	୨	୨
t ଡ୍	th ଡ୍	d ଡ୍	dh ଡ୍	n ନ୍	୨	- 2
୩	୩	୩	୩	୩	୩	୩
t ଡ୍	th ଡ୍	d ଡ୍	dh ଡ୍	n ନ୍	୩	- 3
୪	୪	୪	୪	୪	୪	୪
ch ଚ୍	chh ଛ୍	j ଞ୍	jh ଞ୍	ɳ ଣ୍	୪	- 4
୫	୫	୫	୫	୫	୫	୫
k ଝ୍	kh ଞ୍	g ଞ୍	gh ଞ୍	n ଞ୍	୫	- 5
୬	୬	୬	୬	୬	୬	୬
k ଝ୍	kh ଞ୍	g ଞ୍	gh ଞ୍	n ଞ୍	୬	- 6
୭	୭	୭	୭	୭	୭	୭
y ଯ୍	r ଋ	l ଌ	v ଵ୍	ñ ଞ୍	୭	- 7
୮	୮	୮	୮	୮	୮	୮
s ଣ୍	h ଣ୍	x ଣ୍	r ଣ୍	rh ଣ୍	୮	- 8
୯	୯	୯	୯	୯	୯	୯
s ଣ୍	h ଣ୍	x ଣ୍	r ଣ୍	rh ଣ୍	୯	- 9
୧୦	୧୦	୧୦	୧୦	୧୦	୧୦	୧୦
s ଣ୍	h ଣ୍	x ଣ୍	r ଣ୍	rh ଣ୍	୧୦	- 10

Figure 4: Chart of Tolong Siki (from TolongSiki.com).



मुख्त्यार सिंह, मा० प्र० से०
Mukhtiar Singh, I.A.S.
आयुक्त एवं सचिव
Commissioner and Secretary

कार्मिक, प्रशासनिक सुधार तथा राजभाषा
विभाग, झारखण्ड, राँची ।
Department of Personnel, Administrative
Reforms & Rajbhasha, Government
Jharkhand, Ranchi.
0651 - 2403221 (Off.)
0651 - 2480048 (Res.)
0651 - 2403253 (Fax)

पत्रांक :- 8/रा०-8/2001का०-129, दिनांक : 18-9-2003

सेवा में,
सचिव,
गृह मंत्रालय,
भारत सरकार, नई दिल्ली ।

विषय :- भारतीय संविधान की आठवीं अनुसूची में संथाली, मुण्डारी, हो एवं उरांव/कुरुख भाषा को शामिल करने के संबंध में ।

महोदय,
झारखण्ड राज्य के अन्तर्गत संथाली, मुण्डारी, हो, उरांव/ कुरुख भाषा का एक महत्वपूर्ण स्थान है । संथाली भाषा की लिपि "ओल चिकी", मुण्डारी भाषा की लिपि "देवनागरी", हो भाषा की लिपि "वारंगक्षिति" तथा उरांव/ कुरुख भाषा की लिपि "तोलोंग सिकी" है । वर्ष 1991 की जनगणना के अनुसार संथाली, मुण्डारी, हो तथा उरांव/ कुरुख भाषा का प्रयोग झारखण्ड सहित देश के 29 राज्यों एवं दो केन्द्र शासित प्रदेशों में करने वाली जनसंख्या क्रमशः 52,16,325, 8,61,378, 9,49,216 तथा 14,26,618 है ।

झारखण्ड सहित अन्य कई राज्यों में इन भाषाओं की पढ़ाई विद्यालयों, महाविद्यालयों तथा विश्वविद्यालयों स्तर पर होती है ।

जनजातीय भाषाओं के उत्थान के दृष्टिकोण से झारखण्ड सरकार का यह सुविचारित मतव्य है कि उक्त चारों भाषाओं को संविधान की आठवीं अनुसूची में शामिल किया जाय । अतएव यह अनुरोध है कि संविधान के अनुच्छेद 344(1) एवं अनुच्छेद 351 के प्रावधानों के अन्तर्गत उपाबद्ध आठवीं अनुसूची में संथाली, मुण्डारी, हो, एवं उरांव/ कुरुख भाषा को सम्मिलित किया जाय ।


निश्वासभाजन,
(मुख्त्यार सिंह)
आयुक्त एवं सचिव

03

Figure 6: A letter submitted by the Government of Jharkhand to the central government in 2003, requesting the inclusion of Santhali, Mundari, Ho, and Kurukh in the Eighth Schedule of the Constitution of India. The underline occurs in a sentence about the scripts associated with the aforementioned languages: "The script for the Santhali language is 'Ol Chiki', the script for Mundari is 'Devanagari', the script for Ho is 'Varang Kshiti', the script for Oraon / Kurukh is 'Tolong Siki'." Document courtesy of Ashwin Kumar Kispotta.

प्रभात खबर

शुक्रवार, 20 फरवरी, 2009



झारखण्ड अधिविद्य परिषद्, राँची

JHARKHAND ACADEMIC COUNCIL, RANCHI

वार्षिक माध्यमिक परीक्षा, 2009 के कुडुख भाषा पत्र के संबंध में आवश्यक सूचना

विज्ञप्ति संख्या - 17/2009

एतद् द्वारा सभी छात्रों, उनके अभिभावकों संबंधित विद्यालय के प्राचार्य/पदाधिकारियों को सूचित किया जाता है कि सचिव, मानव संसाधन विकास विभाग, झारखण्ड सरकार, के पत्रांक 6/नि-1-12/2002-565 दिनांक 18.02.2009 द्वारा स्थापना अनुमति प्राप्त उच्च विद्यालय “कुडुख कथा खोइहा लुर एइया भागी टोली” डुमरी गुमला के 39 (उन्नचालीस) छात्रों को वार्षिक माध्यमिक परीक्षा 2009 के कुडुख (उराँव) भाषा पत्र को अपनी लिपि “तोलोंग सिकी” से परीक्षा लिखने की अनुमति प्रदान की गई है।

अध्यक्ष के आदेशानुसार
ह./-
(पोलिकार्प तिकी)
सचिव,
झारखण्ड अधिविद्य परिषद्, राँची।

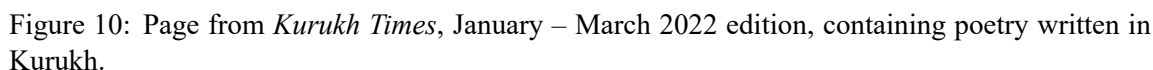
JAC/SECY/078/09 DI. 19.02.09
D.O.P : 20.02.09

Figure 7: Announcement published in 2009 by the Jharkhand Academic Council permitting the usage of Tolong Siki in secondary school examinations. Document courtesy of Ashwin Kumar Kispotta.



Figure 8: Announcement published in 2016 by the Jharkhand Academic Council permitting the usage of Tolong Siki in university examinations. Document courtesy of Ashwin Kumar Kispotta.

Figure 9: Content page from the *Kurukh Times*, January – March 2022 edition, showing text in Tolong Siki, Devanagari, and Latin scripts.



सोने की अण्डे देने वाली मूर्गी

(ඉහත වාර්තා වලට අනුරූපව)

Figure 11: A Kurukh story printed in Tolong Siki with intralinear transliteration in Devanagari (from Ekka 2007: 13). Note the depiction of *ːsela* as the *ː* form with open circles.

তোলোঙ সিকির বর্ণমালা

(তোড়পাব)

‘তোলোঙ সিকি’ তে মোট ৪৬ টা বর্ণ আছে। এখানে ১০ টা স্বর বর্ণ
তথা ৩৫ টা ব্যঞ্জন বর্ণ আছে। আর একটা নাসিক্য ব্যঞ্জনসূচক বর্ণ ‘মিতলা’
আছে। অতএব তোলোঙ সিকির বর্ণমালা কে দুই বর্ণে বিভাজিত করা হয়, সরহ
তোড় তথা হরহ তোড়।

সরহ তোড় = স্বরবর্ণ

প	ব	ঙ	চ	ৱ	ঈ
i ই	e এ	u উ	o ও	a অ	ā আ
: সেলা = লম্বা ধ্বনি,			~ (এবাঁ) এওয়াঁ = নাসিক্য স্বর ধ্বনিসূচক,		
’ ঘেতলা = শব্দখঁড় সূচক,			। হেচকা = বিকারী অ, অর্ধস্বর ‘অ্’		
সূচক ।					

বাস্তবে কুড়ুখ ভাষা তে ছয় টাই মূল স্বর রয়েছে। সেলা তথা এওয়ঁ
সাথে মিশ্রিত হইয়া, এই মূল স্বর গুলো বিভিন্ন রূপ নির্মাণ করে। জেমন:-

প	ব	ঙ	চ	ৱ	ঈ
প্	ব্	ঙ্	চ্	ৱ্	ঈ্
প:	ব:	ঙ:	চ:	ৱ:	ঈ:
প্:	ব্:	ঙ্:	চ্:	ৱ্:	ঈ্:

এ ছাড়া উপরোক্ত স্বর গুলো নিজেদের সাথে মিশে তথা য়, ওয়া
সাথে মিশে বিভিন্ন রূপ নির্মিত করে।

ঘেতলা তথা হেচকাও স্বর মতই ব্যবহৃত হয়।

Figure 12: A Tolong Siki primer in Bengali, showing the basic vowel letters and representations of vowel length (Minz 2016: 7).

সংযুক্ত ব্যঞ্জন বর্ণ (জোড়ঠা হরহ তোড়)

দরতলা নানান রস

যখন দুই খানা ব্যঞ্জন বর্ণ এক সাথে মিলে যায় বা জুড়ে যায়, তখন তাকে সংযুক্ত ব্যঞ্জন বলা হয়। সংযুক্ত হওয়ার নিয়ম কুড়ুখ তথা বাংলা তে এক রকমই আছে।

যমালাগপ িঃঃ দরতাল নানান রস রথল তঁঝখল
ঝখতখগ'প মাগপ পখার দঃঃগ'প মাগপ, ালাগপ াখ দরতলা
নানান ালাগপ. দরতলা ঝালালা ঝালা পথ মঃঃগে ালা ালালা
মঃ রথলখ (রথল ঙালালাখ) ঙাপ.

নিচে সংযুক্ত ব্যঞ্জনের উদাহরন পড়ুন আর লিখুন।

মঃঃঃ দরতলা নানান ঝালা ঝালাঃঃ ালা ালা ঝঃঃ'া.

প্+প া+া = প্প াা, প্পড়া াাাা = এক রকমের ফল।

প্+প া+া = প্প াা, কপ্পনা াাাা = ধরা / স্পর্শ করা।

চ্+চ া+া = চ্চ াা, ঝচ্চনা াাাা = ভাঙ্গা, বাচ্চা াাাা=লুট

দ্+দ া+া = দ্দ াা, ঝদ্দ াাা = বেদা াাাা = খোজ করো।

চ্+ছ া+া = চ্ছ াা, মেচ্ছা াাা =উঁচু, গেচ্ছা াাা=দূর।

জ্+জ া+া = জ্জ াা, অজ্জা াাা =দাদু অজ্জী াাা= দাদী।

ট্+ট া+া = ট্টি াা, চুটী াাা = চুল, সুটী াাা = ছড়ি।

ট্+ঠ া+া = ট্ঠ াা, মুট্ঠী াাা=মুষ্টি, গুট্ঠী াাা=ইত্যাদি।

ড্+ড া+া = ড্ড াা, গড্ঠী াাা =গর্ত, অড্ঠা াাা = জায়গা।

ল্+ল া+া = ল্ল াা, পেল্লা াাা=যুবতী, গল্লা াাা = গাল।

Figure 13: Excerpt from a Tolong Siki primer in Bengali, showing representation of consonant clusters (Minz 2016: 30).



Figure 14: Tolong Siki used as primary title on covers of books, along with other scripts. Images courtesy of Aswin Kumar Kispotta.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

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

Please read Principles and Procedures Document (P & P) from <http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

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

A. Administrative

1. Title:	Proposal to encode Tolong Siki in Unicode	
2. Requester's name:	Anshuman Pandey <pandey@umich.edu>	
3. Requester type (Member body/Liaison/Individual contribution):	Expert contribution	
4. Submission date:	2023-01-05	
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:		

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):		Yes
Proposed name of script:	Tolong Siki	
b. The proposal is for addition of character(s) to an existing block:		
Name of the existing block:		
2. Number of characters in proposal:		54
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. 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
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Anshuman Pandey	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	Anshuman Pandey	
6. 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
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

8. 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 Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

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?	
Narayan Oraon <oraon.narayan@gmail.com> Ashwin Kumar Kispotta <ashlumen37@gmail.com>	
If YES, available relevant documents:	
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 text of proposal	
4. The context of use for the proposed characters (type of use; common or rare)	Common
Reference: See text of proposal	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference: Tolong Siki is used by Kurukh speakers in India and has official recognition in the states of Jharkhand and West Bengal	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	No
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
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, or could be confused with, 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?	
No	
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 characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	