Proposal to encode the Sunuwar script in Unicode

Anshuman Pandey

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

December 6, 2021

1 Introduction

The last proposal to encode the Sunuwar script in Unicode was submitted over a decade ago (L2/11-218). Since then, I have intermittently monitored the script and have seen evidence of its ongoing usage by Sunuwar communities in both Nepal and Sikkim. However, the encoding effort has been on hold due to inconsistent contact with the user community. A renewal of the project began in June 2020, when Jeannette Stewart of Translation Commons initiated a collboration between Dev Kumar Sunuwar (Indigenous Media Foundation, Nepal), Deborah Anderson (Script Encoding Initiative), Craig Cornelius, Craig Cummings, Gerry Leonidas, Tex Texin, myself, and others. This group has met at least once a month through video calls to discuss the usage and development of the script. Dev Kumar Sunuwar has facilitated discussions and exchange of documents between the group and the Sunuwar Welfare Society. Lal-Shyakarelu Rapaca, who provided me with information on the script in 2010, is involved in these discussions as part of the Society's team. This collaboration has confirmed that the script has been in continuous usage since its creation by K. B. Jentich in the 1940s. It has been enhanced by users to better represent the Sunuwar language and is actively used, taught, and promoted today. Given the stability and usage of the script, it is suitable for inclusion in Unicode.

The current proposal supersedes the following documents:

- L2/10-466R: "Preliminary Proposal to Encode the Jenticha Script"
- L2/11-218: "Proposal to Encode the Jenticha Script"

The major changes between it and L2/11-218 are as follows:

- Change of the Unicode designation for the script from 'Jenticha' to 'Sunuwar'
- Usage of Sunuwar names for the letters instead of Latin transliterations
- · Ordering of letters according to Jentich's arrangement
- Addition of letters to complete the repertoire
- Recommendation to use common combining signs for various diacritics instead of script-specific signs
- Expanded description of nasalization sign, vowel length mark, tone marks, and legacy diacritics
- · Addition of an auspicious sign representing a mystical sound in Sunuwar culture

The code "Sunu" (#274) has been registered for the script in ISO 15924.

2 Description

The proposed script was developed in 1942 by Krishna Bahadur Jentich (1926–1991) for writing Sunuwar (ISO 639-3: suz), a Kiranti language of the Tibeto-Burman family that is spoken in eastern Nepal and Sikkim, India. It is a linear alphabet that is written from left to right, in lines arranged from top to bottom. While it has no genetic relationship to other writing systems, it is apparent that its structure is modeled upon the Latin alphabet, and some of its letterforms appear to be inspired by Latin shapes (see fig. 2, 4, 8).

The repertoire consists of 33 letters, 10 digits, and an auspicious sign. Combining signs and tone marks corresponding to Latin forms are also used. The current repertoire is a combination of Jentich's original script and additions by users (see fig. 7). The orthography established by Jentich has been expanded through the addition of new letters and diacritics. There is no formal standard for the script. However, the Sunuwar Welfare Society of Nepal promotes a repertoire that may be considered a *de facto* standard.

2.1 Creation

The original script consisted of 22 letters and the digits. Jentich assigned to each letter a distinctive name in the Sunuwar language and arranged them in the following order:

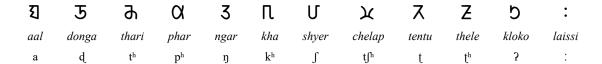
Д	a	±	†	Ν.		7	τ	π	9	ያ	g
devi	tasla	eko	imar	reu		utthi	kik	ma	appho	pip	gil
d, d	t, t	e	i	r, ţ		u	k	m	a	p	g
tγ	<	Ŋ	J	٨		T	0	7	k	Q	거
hamso	carmi	nah	bur	jyah		loacha	otthi	shyele	varca	yat	ava
h	t∫	n	b	dз		1	o	s,∫	v	у	6
	0	1	‡	‡		В	⊿	П	Χ	۲	
	sum	ka	niashi	san	le	nga	raku	chani	sasi	yan	
	0	1	2	3	4	5	6	7	8	9	

The original orthography is described below. Whether intentional or not, the rules resemble the Latin orthography for English and the conventions used in the simplified romanization of South Asian scripts.

- Short and long forms of vowels were expressed using a single letter: 9 /a/, + /i/, ± /e/, \(\gamma \) /u/, 0 /o/
- Vowel length was not distinguished orthographically
- A letter could also represent a retroflex consonant, eg. π for /d/, /d/; π for /t/, /t/; π for /r/, /t/
- Aspirated consonants were written as digraphs using \hbar /h/ for /h/, eg. \hbar \hbar /k/, \hbar /t/h/
- The letter η /n/ was used for both vowel nasalization and consonantal nasals, eg. η /an/, η /nt/
- The velar nasal $/\eta$ was written as the digraph $\eta 9 / ng$
- The letter 7 was used for both /s/ and /ʃ/
- Although Sunuwar is a tonal language, there were no conventions for indicating tone
- The 10 digits represent decimal numbers and correspond to the Hindu-Arabic digits 0–9.

2.2 Expansion

By the turn of the 21st century, Lal-Shyakarelu Rapaca and other users had added the following 11 letters and a vowel-length marker to the repertoire:



These extensions enabled more precise representation of the Sunuwar language:

- The letter 3, borrowed from Limbu, was introduced for writing long /a/
- Writing of the /?/ glottal stop using the letter 5
- Distinctive letters for aspirated consonants, eg. $\Pi / k^h / \chi / t^h / \lambda / t^h / \alpha / p^h /$
- The letter U for the palatal sibilant /ʃ/
- Usage of 3 for the velar nasal $/\eta$ /
- Indication of vowel length using the sign:, eg. + /i/, +: /i:/

The introduction of specific letters for certain phonemes resulted in an orthography that provided more than one method for writing certain sounds:

- Retroflex consonants may be written with a general class letter or the dedicated letter, eg. 2 and 7 /t/
- Aspirated consonants may be expressed using digraphs or a dedicated letter, eg. < h or x for /tʃh/
- The velar nasal may be written as 99 / ng/ or as $3 / \eta/$
- Vowel nasalization may be written using η /n/ or the nasalization sign, eg. $\dagger \eta$ or $\dot{\dagger}$ for /i/

2.3 Experimentation

During the same period, Rapaca experimented with the script by interpreting it as an alphasyllabary, along the lines of Devanagari:



The experimental orthography had the following features:

- The alphabetic consonant letters were interpreted as possessing an inherent vowel, eg, $\mathbb{1}/d/=/da/$. The sign \circ *sangmilu* was created to indicate suppression of the inherent vowel; eg. $\mathbb{1}/d/$
- Vowel letters functioned as both independent vowels and as dependent forms. A sequence of consonant + vowel letter could be treated as an orthographic syllable, eg. JI /da:/
- The consonant N /r/ was given characteristics similar to Devanagari ₹ ra. The sangkirs represents cluster-initial /r/ and functions like Devanagari repha: it was written as part of the following orthographic syllable, eg. ¬ /rda/. If a 'dependent' vowel is part of the following syllable, the sangkirs is written above it: ¬ /rda:/

- Cluster-medial /r/ was written as sangrums. This followed the model of Devanagari vattu. It was written beneath the preceding letter in the cluster, eg. \(\mathbb{I} \) /dra/
- Nasalization was indicated using $\dot{\bar{b}}$ taslathenk, eg. $\dot{\bar{g}}$ /ã/
- In some sources, the : *laissi* was enclosed in parentheses, eg. (:) in order to distinguish it from the colon
- There was an attempt to represent consonant clusters using conjuncts, in particular half-forms of non-final consonants, eg. \langle for cluster-initial \langle c and Π for M m. However, half-forms for all consonants were not produced. Only the aforementioned three were used in published materials.

Rapaca's re-interpretation of the script was not adopted by the user community. It added complexities to a simple orthography. While the *taslathenk* was a useful addition, the other characters are unnecessary. The *sangmilu* is redundant in an alphabet. Its usage for indicating a consonant cluster has potential benefits, but those would be marginal to a seasoned reader of the script. The usage of special signs for /r/ (*sangkirs* and *sangrums*) are cumbersome. The *sangkirs*, specially, reduces the efficiency of writing: a user has to write all the letters of a given syllable before jotting down the /r/. This experiment is similar to various idiosyncratic revisions proposed for other scripts, but Rapaca employed the signs in writing the Sunuwar edition of the *Sikkim Herald* (see the masthead in fig. 14), as well as in the publication of a script primer. Therefore, there is a potential need to accommodate the representation of these signs.

2.4 Growth

More recently, in order to fully express the Sunuwar language, the script committee of the Sunuwar Welfare Society recently recommended a new sign for nasalization and the usage of three signs for indicating tone:



These signs are influenced by diacritics used in the Latin script. The script committee believes that using well-understood signs for nasalization and tones would faciliate learning and retention of the script better than the invention of new signs specific to the script.

The expanded repertoire shows both stability and innovation. It is stable in that the repertoire as created by Jentich forms the core of the alphabet and the letters added by Rapacha and others twenty years ago have become naturalized as part of the script. It is innovative in that users continue to enhance the script as needed for representing the Sunuwar language while adhering to the basic principles of Jentich's script.

3 Designation for the Script in Unicode

The proposed Unicode designation for the script is 'Sunuwar'. It is the name by which K. B. Jentich referred to his script in English (see fig. 10). The name 'Sunuwar' (सुनुवार sunuvāra) is a Nepali exonym for the community. It is the name used officially by the Government of Nepal for the community and the language. The endonym is 'Koinch' /koîtʃ/ (†0†१< or †0 रें<), Devanagari कोईंच koîc). But, members of the community use the term 'Sunuwar' externally, such as the English name "Sunuwar Welfare Society" and its Nepali analogue "Sunuwar Sewa Samaj". The name may be romanized as 'Sunwar', as is the case in ISO 639-3. However the spelling 'Sunuwar' is preferred by the community and the spelling that is used in Nepal.

Other candidates for the Unicode identifier are given below. Their lack of suitability for this purpose is also provided.

- 1. 'Jenticha' Earlier proposals to encode the script used the name 'Jenticha', which is a romanization of the script creator's surname. It was chosen for those documents based upon feedback that the script was known within the Sunuwar community as जेहंतिच ब्रेसे or जैंतिच ब्रेसे jetic brese = 'Jentich's script'. That name is not suitable for use in Unicode as it does not provide any context regarding the script or association to a language or community.
- 2. 'Jentich Sunuwar' A recent convention of the Script Ad Hoc has been to refer to 'neographies' using the nomenclature <script creator> + <language>. Accordingly, the script would be called 'Jentich Sunuwar'. However, as Jentich is the name of a Sunuwar clan, such a designation would inaccurately suggest that the script belongs specifically to this sub-community. Moreover, the Sunuwar experts with whom I have been collaborated are not in favor of this term. For this reason, there is no need to add the creator's name in the script designation.
- 3. 'Koinch' The autonym is कोइँच ब्रेसे kõic brese / kõits brese or 'Koinch script' / 'Kõits script'. While known within the Sunuwar community, this term is too restrictive for an international standard because it does not provide any English context for identifying the script.
- 4. 'Kiranti Koinch' Rapaca used the term किराँती-कोँइच kirãtī-kõica 'Kiranti-Kõits' to refer to the script. This is unsuitable because the term 'Kiranti' refers to a group of languages of eastern Nepal and Sikkim, which includes Limbu, Yakkha, and Rai. The Limbu script of Sirijonga may be considered a 'Kiranti' script.

Usage of the identifier 'Sunuwar' for this script in Unicode would not present any issues for misidentification. There are no other indigenous scripts that are used by the community. The 'Tikamuli' script by Tikaram Mulich that I presented in L2/10-465 is a novelty and was never used.

4 Usage

Sunuwar has been one of the official local languages of Sikkim, where it is known as 'Mukhia', since 1996. Both the language and the script proposed here for encoding were recognized by the state in that year. Government records were maintained and newspapers were published in the script. In January 2021, the Sikkim government mandated that state-run schools offer instruction to students in both the primary languages of the state and in a local language (*Indian Express*, January 29, 2021). The Sunuwar script will likely be the medium of instruction.

The Sunuwar script is currently used in writing and print for various purposes, as described below:

- Educational Materials Primary educational materials for teaching the Sunuwar language and script are published by Kõic Bu in Sikkim. Examples of script primers are Kirātī-Kõica Blesethekā (2003) and Kõicā The:si Khyõpat (2004); see Figure 15. The Department of Education of the Government of Sikkim is planning to prepare textbooks in the script for primary-level education.
- *Literary Works* Anthologies of Sunuwar poetry are printed in the script.
- *Historical Records* Histories of the Sunuwar communities, known as *vaṃśāvalī*, were written in Sunuwar beginning in the 1940s.

- Newspapers A Sunuwar ('Mukhiya') edition of the Sikkim Herald is produced in Sunuwar. The newspaper was first published in April 2002 and maintains an active circulation. The first edition was hand-written by Lāl-Śyāṃkārelu Rāpacā, with subsequent editions written by Balarām Sunuwār (see Figure 12). More recent editions are printed in digitized type (see Figure 13).
- Linguistic Works The script is used in grammars of Sunuwar, ie. Elementary Grammar of Kiranti-Sunuwar by Kamalādevī Nāvacā-Mukhiyā, Raghuvīra Rujicā-Mukhiya, and Lāla Rāpacā (Sikkim: Kõica Bu, 2003).
- *Government Records* Proceedings of the Sikkim legislative assembly were translated into Sunuwar using the script in until recently.
- Academic Works The script is used in scholarly works on the Sunuwar language and it is also a focus of study (cf. Rāpacā 2009).
- *Ephemera* The Sikkim Sunuwar Mukhia Koinchbu (SSMK) produces and distributes calendars using Sunuwar (see Figure 17 and 18).
- *Digital Resources* Sunuwar has been adapted for use in digital media. A digitized font for the script named 'Kirat1' was developed by Shyan Kirat Rai. It is based upon a non-Unicode encoding and is mapped to Latin letters. The font is used for publishing the *Sikkim Herald* (see Figure 13) and by Rāpacā in the charts and examples shown in figures 5–16. The Sunuwar Welfare Society has developed fonts named 'Kaatich' and 'Mukdum'. These fonts are mapped to the Latin encoding and are commonly used in publications.

5 Proposed Unicode Repertoire

The proposed Unicode repertoire for 'Sunuwar' contains 44 characters, which include 33 letters, 1 auspicious sign, and 10 digits:

	Letters					
	Unicode Character Name	Name	Value	Devanagari		
Д	SUNUWAR LETTER DEVI	devi	/d/, /d/	द, ड		
a	SUNUWAR LETTER TASLA	tasla	/t/, /t/	त, ट		
±	SUNUWAR LETTER EKO	eko	/e/	ए		
t	SUNUWAR LETTER IMAR	imar	/i/	इ		
Ν.	SUNUWAR LETTER REU	reu	/r/, /r/, /l/	र, ड़		
7	SUNUWAR LETTER UTTHI	utthi	/u/	उ		
τ	SUNUWAR LETTER KIK	kik	/k/	क		
π	SUNUWAR LETTER MAMA	mama	/m/	म		
9	SUNUWAR LETTER APPHO	appho	/e/	अ		
ያ	SUNUWAR LETTER PIP	pip	/p/	Ч		

g	SUNUWAR LETTER GIL	gil	/g/	ग
ħ	SUNUWAR LETTER HAMSO	hamso	/h/	ह
<	SUNUWAR LETTER CARMI	carmi	/tʃ/	च
Ŋ	SUNUWAR LETTER NAH	nah	/n/	न, ण
ż	SUNUWAR LETTER BUR	bur	/b/	ब
Y	SUNUWAR LETTER JYAH	jyah	/d3/	ज
7	SUNUWAR LETTER LOACHA	loacha	/1/, /[/	ल
0	SUNUWAR LETTER OTTHI	otthi	/o/	ओ
7	SUNUWAR LETTER SHYELE	shyele	/s/	स
k	SUNUWAR LETTER VARCA	varca	/v/	व
Q	SUNUWAR LETTER YAT	yat	/y/	य
H	SUNUWAR LETTER AVA	ava	/6/	ৰ
រា	SUNUWAR LETTER AAL	aal	/a:/	आ
5	SUNUWAR LETTER DONGA	donga	/d/	ड
ъ	SUNUWAR LETTER THARI	thari	$/t^{\rm h}/$	थ
α	SUNUWAR LETTER PHAR	phar	$p^{\rm h}$	फ
3	SUNUWAR LETTER NGAR	ngar	/ŋ/	ङ
Π	SUNUWAR LETTER KHA	kha	$/k^{\rm h}/$	ख
U	SUNUWAR LETTER SHYER	shyer	/ʃ/	श, ष
х	SUNUWAR LETTER CHELAP	chelap	$/t \int^h /$	छ
Z	SUNUWAR LETTER TENTU	tentu	/t/	ਟ
Z	SUNUWAR LETTER THELE	thele	/th/	ਠ
b	SUNUWAR LETTER KLOKO	kloko	\3/	अ्

Auspicious Sign

	Unicode Character Name	Name	Value	Devanagari
Т	SUNUWAR SIGN PVO	pvo	/\bar{\bar{\bar{\bar{\bar{\bar{\bar{	_

Digits

	Unicode Character Name	Name	Value	Devanagari
0	SUNUWAR DIGIT ZERO	sum	0	0
I	SUNUWAR DIGIT ONE	ka	1	१

۲	SUNUWAR DIGIT NINE	yan	9	9	
X	SUNUWAR DIGIT EIGHT	sasi	8	۷	
	SUNUWAR DIGIT SEVEN	cani	7	9	
Д	SUNUWAR DIGIT SIX	raku	6	ξ	
В	SUNUWAR DIGIT FIVE	nga	5	ų	
0	SUNUWAR DIGIT FOUR	le	4	8	
1	SUNUWAR DIGIT THREE	san	3	3	
#	SUNUWAR DIGIT TWO	ni'ashi	2	7	

5.1 Notes on proposed characters

The $\mbox{\em W}$ SUNUWAR SIGN PVO represents an auspicious syllable, which is articulated as the unvoiced bilabial implosive $/\mbox{\em \wp}/$ (formerly $/\mbox{\em \wp}/$) and transcribed as 'pvo'. In spoken language, the syllable is often utter twice, ie. "pvo, pvo..." The sign is written in salutations and benedictory phrases, eg. before names in the header of the letter shown in fig. 10 and at the end of the list of letters of the alphabet (see figs. 2, 7). It's structure is based upon a trident, but it has various representation from simple to ornate. The representative glyph used here is based upon the form shown in fig. 7.







from fig. 10

from fig. 2

from fig. 7

6 Collation

The sort order for Sunuwar is as follows:

Д	a	±	ł	Ν.	7	tc	π	9	ያ	g
devi	tasla	eko	imar	reu	utthi	kik	ma	appho	pip	gil
ħ	<	Ŋ	ł	٨	Ţ	0	7	k	Q	౫
hamso	carmi	nah	bur	jyah	loacha	otthi	shyele	varca	yat	ava
য্	5	ъ	α	3	Π	U	х	Z	Z	b
aal	donga	thari	phar	ngar	kha	shyer	chelap	tentu	thele	kloko

Some sources show letters sorted according to the Latin order of their transliterated values (figs. 2, borchers:a) or according to the Devanagari pattern (fig. 5). These orders are not used at present.

7 Diacritics

There are several diacritic signs used in the Sunuwar script. They resemble signs used in Latin and Devanagari. Some of these signs are part of the modern standard orthography promoted by the Sunuwar Welfare Society; others are not part of the conventional orthography, but may still be used on account of precedence. At present, there is no requirement to encode these diacritics as Sunuwar-specific characters. Instead, if there is a need to use these signs, they may be represented using common combining signs that are already present in Unicode. In a Sunuwar font, it is expected that the glyphs for these common combining signs would align with the design of the Sunuwar characters. The diacritics and possible analogues to common combining signs are enumerated below. These common signs have been specified as script extensions for Sunuwar.

7.1 Conventional diacritics

Conventional Sunuwar diacritic signs may be represented using existing common Unicode characters:

	Name	Function	Unicode representation
ិ៍	nesante	nasalization	Õ U+0303 COMBINING TILDE
:	laissi	vowel-length mark	: U+003A COLON

Usage of *nesante* (red) and *laissi* (blue) are shown below; also shown is usage of the colon (green), which is identical in this Sunuwar font to *laissi*.



The *laissi* is generally not distinguished from the conventional: colon. In the excerpt below (from fig. 12), the shape of *laissi* (red) is identical to the colon (blue):

In one publication Rapaca enclosed *laissi* in parentheses (red), likely to distinguish the sign from the Devanagari : *visarga* (blue) or the colon (excerpt from fig. 16). Here the parentheses have been resized such that their x-height matches the top and bottom of the colon.

廿०० १ १०० १ ४५५ boisšon bā2bə <बोइस्शोन् बाअ्बः≯

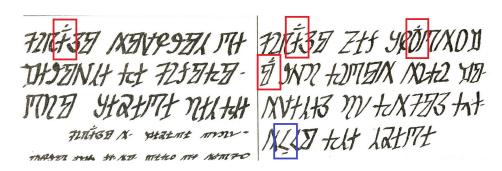
7.2 Legacy diacritics

The following characters were intoduced by Rapaca as part of an orthographic experiment. They were not widely adopted. However, from a strict perspective, to develop an encodinf that supports all texts in the Sunuwar script, it may be necessary to provide support for legacy conventions. The Sunuwar experts consulted during the development of the current proposal have stated that these sign are not included in the formal orthography that is promoted by the Sunuwar Welfare Society. Nonetheless, these signs were used by Rapaca in writing the *Sikkim Herald*, in his personal publications, and in a published script primer (see fig. 15). For this reason, there should be support for these signs in order to fully encode texts in which they appear. They are to be represented using existing common Unicode characters:

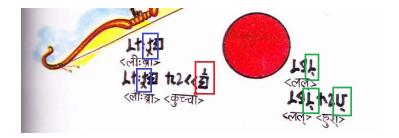
	Name	Function	Unicode representation
ੁ	sangmilu	halanta / vowel silencer	ႍ U+0331 COMBINING MACRON BELOW
ं	taslathenk	nasal sign	ំ U+0310 COMBINING CANDRABINDU
់	sangkirs	repha	់ U+1DF5 COMBINING UP TACK ABOVE
्र	sangrums	ra-vattu / trill mark	Ç U+032D COMBINING CIRCUMFLEX ACCENT BELOW

The recommended Unicode representations for *sangmilu*, *sangkirs*, and *sangrums* use common Unicode characters with forms that are identical to the Sunuwar signs. The recommended Unicode representation for *taslathenk* differs slightly from the Sunuwar form: U+0310 COMBINING CANDRABINDU has a curved base stroke while that stroke in *taslathenk* is horizontal. Despite this slight difference, it is recommended that U+0310 for *taslathenk* because it is sufficient for supporting a legacy sign. The COMBINING CANDRABINDU possesses the same semantics as *taslathenk* – and their names have the same meaning – and the resemblance between the two is sufficient enough to identify the function and meaning of the sign.

Usage of the sangmilu (blue) and taslathenk (red) are shown below:



Usage of the *sangkirs* (red) and *sangrums* (blue) are shown below, as well as a variant of the *sangmilu* (green) that resembles Devanagari Q VIRAMA:



7.3 Diacritics being promoted for usage

The Sunuwar Welfare Society is advocating usage of tone marks for disambiguation. Existing combining signs may be used for representing these tone marks:

	Name	Function	Unicode representation
់	sotte'si	stress mark	் U+030D COMBINING VERTICAL LINE ABOVE
'	ta renmaresi	high tone	Ó U+0301 COMBINING ACUTE ACCENT
े	di renmaresi	low tone	` U+0300 COMBINING GRAVE ACCENT

8 Punctuation

There is no script-specific punctuation. Words are delimited using spaces. Hyphenation is indicated using the hyphen, but there are no formal rules for splitting words.

Latin punctuation marks are used, eg. period, comma, colon, semi-colon, single and double quotes, etc.

9 Character Data

Character Properties: UnicodeData.txt

```
11BC0;SUNUWAR LETTER DEVI;Lo;0;L;;;;N;;;;
11BC1;SUNUWAR LETTER TASLA;Lo;0;L;;;;N;;;;
11BC2;SUNUWAR LETTER EKO;Lo;0;L;;;;N;;;;
11BC3;SUNUWAR LETTER IMAR;Lo;0;L;;;;N;;;;
11BC4;SUNUWAR LETTER REU;Lo;0;L;;;;N;;;;
11BC5;SUNUWAR LETTER UTTHI;Lo;0;L;;;;N;;;;
11BC6;SUNUWAR LETTER KIK;Lo;0;L;;;;N;;;;
11BC7;SUNUWAR LETTER MA;Lo;0;L;;;;N;;;;
11BC8;SUNUWAR LETTER APPHO;Lo;0;L;;;;N;;;;
11BC9;SUNUWAR LETTER PIP;Lo;0;L;;;;N;;;;
11BCA;SUNUWAR LETTER GIL;Lo;0;L;;;;N;;;;
11BCB;SUNUWAR LETTER CARMI;Lo;0;L;;;;N;;;;
11BCC;SUNUWAR LETTER NAH;Lo;0;L;;;;N;;;;
11BCD;SUNUWAR LETTER NAH;Lo;0;L;;;;N;;;;
```

```
11BCF;SUNUWAR LETTER JYAH;Lo;0;L;;;;;N;;;;;
11BD0;SUNUWAR LETTER LOACHA;Lo;0;L;;;;;N;;;;
11BD1;SUNUWAR LETTER OTTHI;Lo;0;L;;;;;N;;;;;
11BD2;SUNUWAR LETTER SHYELE;Lo;0;L;;;;;N;;;;;
11BD3;SUNUWAR LETTER VARCA;Lo;0;L;;;;;N;;;;;
11BD4;SUNUWAR LETTER YAT;Lo;0;L;;;;;N;;;;;
11BD5;SUNUWAR LETTER AVA;Lo;0;L;;;;;N;;;;;
11BD6;SUNUWAR LETTER AAL;Lo;0;L;;;;;N;;;;;
11BD7;SUNUWAR LETTER DONGA;Lo;0;L;;;;;N;;;;;
11BD8;SUNUWAR LETTER THARI;Lo;0;L;;;;;N;;;;;
11BD9;SUNUWAR LETTER PHAR;Lo;0;L;;;;;N;;;;
11BDA; SUNUWAR LETTER NGAR; Lo; 0; L;;;;; N;;;;;
11BDB;SUNUWAR LETTER KHA;Lo;0;L;;;;;N;;;;;
11BDC;SUNUWAR LETTER SHYER;Lo;0;L;;;;;N;;;;;
11BDD;SUNUWAR LETTER CHELAP;Lo;0;L;;;;;N;;;;;
11BDE; SUNUWAR LETTER TENTU; Lo; 0; L;;;;; N;;;;;
11BDF;SUNUWAR LETTER THELE;Lo;0;L;;;;;N;;;;;
11BE0;SUNUWAR LETTER KLOKO;Lo;0;L;;;;;N;;;;;
11BE1;SUNUWAR SIGN PVO;Po;0;L;;;;;N;;;;;
11BF0;SUNUWAR DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
11BF1;SUNUWAR DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
11BF2;SUNUWAR DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
11BF3;SUNUWAR DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
11BF4;SUNUWAR DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
11BF5;SUNUWAR DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
11BF6;SUNUWAR DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
11BF7;SUNUWAR DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
11BF8;SUNUWAR DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
11BF9;SUNUWAR DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
```

Linebreaking Properties: LineBreak.txt

```
11BC0..11BE0;AL  # Lo  [33] SUNUWAR LETTER DEVI..SUNUWAR LETTER KLOKO
11BE1;AL  # Po  SUNUWAR SIGN PVO
11BF0..11BF9;NU  # Nd  [10] SUNUWAR DIGIT ZERO..SUNUWAR DIGIT NINE
```

Script Extensions: ScriptExtensions.txt

```
# Script Extensions=Sunu
0300
                                COMBINING GRAVE ACCENT
              ; Sunu # Mn
0301
              ; Sunu # Mn
                                COMBINING ACUTE ACCENT
0303
              ; Sunu # Mn
                                COMBINING TILDE
                                COMBINING VERTICAL LINE ABOVE
030D
              ; Sunu # Mn
0310
              ; Sunu # Mn
                                COMBINING CANDRABINDU
032D
              ; Sunu # Mn
                                COMBINING CIRCUMFLEX ACCENT BELOW
              ; Sunu # Mn
                                COMBINING MACRON BELOW
0331
                                COMBINING UP TACK ABOVE
1DF5
              ; Sunu # Mn
```

10 References

- Borchers, Dörte. 2008. A Grammar of Sunwar: Descriptive Grammar, Paradigms, Texts and Glossary. Tibetan Studies Library; Languages of the Greater Himalayan Region, vol. 5/7. Leiden: Brill.
- kiratsunuwar.org.np. (n.d.) "Summary of K. B. Sunuwar". Written by Shankar Sunuwar. Source: Purna Bahadur Sunuwar. http://kiratsunuwar.org.np/index.php?option=com_content&view=article& id=8:summary-of-kb-sunuwar&catid=3:articles&Itemid=19. Accessed January 2011.
- Indian Express. "Sikkim asks schools to introduce 11 local languages in curriculum", January 29, 2021.
- Mukhia, Kamaladevi, and Raghubir Mukhia, and Lal Rapaca. 2004. ช่วี่+<ปี ฉ±:7† กินอั๋ฯฉ [= Elementary Reader of Sunuwar]. 1st ed. Sikkim: Kõica Bu.
- Pandey, Anshuman. 2011a. "Preliminary Proposal to Encode the Tikamuli Script in ISO/IEC 10646" (N3963 L2/10-465R). https://www.unicode.org/L2/L2010/10465r-tikamuli.pdf
- ——. 2010a. "Preliminary Proposal to Encode the Jenticha Script in ISO/IEC 10646" (N3962 L2/10-466R). https://www.unicode.org/L2/L2010/10466r-jenticha.pdf
- ——. 2010b. "Preliminary Proposal to Encode the Jenticha Script in ISO/IEC 10646" (N3962 L2/10-466R). https://www.unicode.org/L2/L2010/10466r-jenticha.pdf
- ——. 2011. "Proposal to Encode the Jenticha Script" (N4028 L2/11-218). https://www.unicode.org/L2/L2011/11218-n4028-jenticha.pdf
- रापचा, लाल-श्याँकारेलु [Rāpacā, Lāl-Śyāṃkārelu]. 2009. नेपालको आदिवासी किराँती जाति सेरोफेरो : ओल्लो, माझ र पल्लो किराँत. काठमांडौ : किरांतविज्ञान अध्ययन संस्थान.
- Subba, Jash Raj. 2008. *The Evolution of Man and the Modern Society in Mountainous Sikkim*. Delhi: Gyan Publishing House.

11 Acknowledgments

This latest proposal for encoding the Sunuwar script has been made possible by the collaboration initiated in June 2020 by Jeannette Stewart between Dev Kumar Sunuwar and me. Dev has provided materials that show evidence of the script's stability and continuity. He has also served as a conduit between me and the Sunuwar Welfare Society.

My initial research on the script in 2010 was assisted greatly by Lal-Shyakarelu Rapaca, who provided details about the history and orthography of Jentich's script. At that time, Krishna B. Sunuwar of Kathmandu also provided information about the scripts used for writing Sunuwar in Sikkim and Nepal.

This effort 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 through the material support of the Script Encoding Initiative (SEI) and its director, Deborah Anderson.

	11BC	11BD	11BE	11BF
0	Д	1	5	0
1	11BC0 2	11BD0	11BE0	11BF0
	11BC1	11BD1	11BE1	11BF1
2	11BC2	11BD2		# 11BF2
3	11BC3	11BD3		1 11BF3
4	1 1BC4	Q		11BF4
5	7	H		В
6	11BC5	11BD5		11BF5
	11BC6	11BD6		11BF6
7	11BC7	11BD7		11BF7
8	11BC8	11BD8		X 11BF8
9	9 11BC9	X		11BF9
Α	ઝ	3		
В	11BCA	11BDA		
	11BCB	11BDB		
С	11BCC	7 11BDC		
D	1 1BCD	11BDD		
Е	5	7		
F	A	Z		
	11BCF	11BDF	/////	(III)

The script is known indigenously as 'Koinch brese'

Letters

11BC0 🕱 SUNUWAR LETTER DEVI 11BC1 SUNUWAR LETTER TASLA a 11BC2 ± SUNUWAR LETTER EKO 11BC3 + SUNUWAR LETTER IMAR 11BC4 N SUNUWAR LETTER REU 11BC5 γ SUNUWAR LETTER UTTHI 11BC6 tc SUNUWAR LETTER KIK 11BC7 π SUNUWAR LETTER MA 11BC8 9 SUNUWAR LETTER APPHO 11BC9 SUNUWAR LETTER PIP 11BCA 9 SUNUWAR LETTER GIL 11BCB th SUNUWAR LETTER HAMSO 11BCC < SUNUWAR LETTER CARMI 11BCD የ SUNUWAR LETTER NAH 11BCE Ł SUNUWAR LETTER BUR 11BCF ∀ SUNUWAR LETTER JYAH 11BD0 人 SUNUWAR LETTER LOACHA 11BD1 0 SUNUWAR LETTER OTTHI 11BD2 U SUNUWAR LETTER SHYELE 11BD3 ▶ SUNUWAR LETTER VARCA 11BD4 Q SUNUWAR LETTER YAT 11BD5 귓 SUNUWAR LETTER AVA SUNUWAR LETTER AAL 11BD7 **5** SUNUWAR LETTER DONGA 11BD9 Q SUNUWAR LETTER PHAR 11BDA 3 SUNUWAR LETTER NGAR 11BDB Π SUNUWAR LETTER KHA 11BDC **7** SUNUWAR LETTER SHYER 11BDD 🗶 SUNUWAR LETTER CHELAP 11BDE Z SUNUWAR LETTER TENTU 11BDF **Z** SUNUWAR LETTER THELE 11BE0 5 SUNUWAR LETTER KLOKO

Auspicious sign

11BE1 Ψ SUNUWAR SIGN PVO

Digits

11BF0	0	SUNUWAR DIGIT ZERO
11BF1	-	SUNUWAR DIGIT ONE
11BF2	#	SUNUWAR DIGIT TWO
11BF3	#	SUNUWAR DIGIT THREE
11BF4		SUNUWAR DIGIT FOUR
11BF5	В	SUNUWAR DIGIT FIVE
11BF6	Δ	SUNUWAR DIGIT SIX
11BF7	\blacksquare	SUNUWAR DIGIT SEVEN
11BF8	Χ	SUNUWAR DIGIT EIGHT
11BF9	۲	SUNUWAR DIGIT NINE

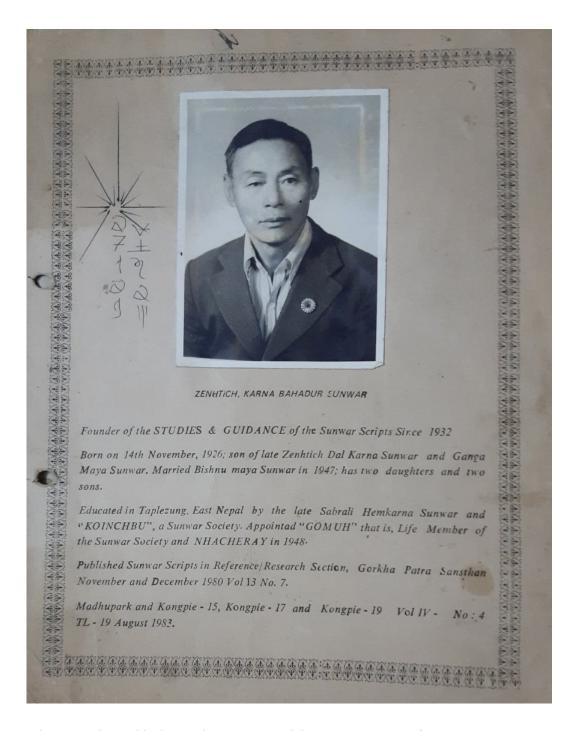


Figure 1: Biographical note about K. B. Jentich. Image courtesy of Dev Kumar Sunuwar.

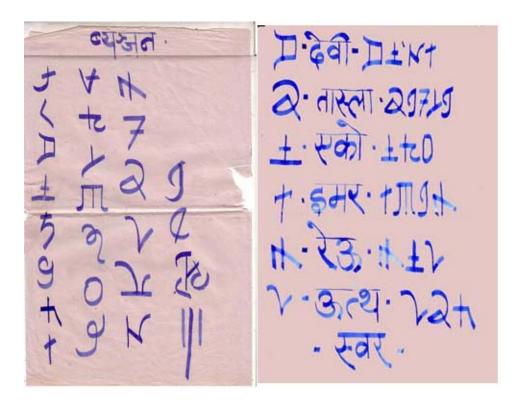


Figure 2: Sunuwar letterforms reported written by K. B. Jentich (from kiratsunuwar.org.np)

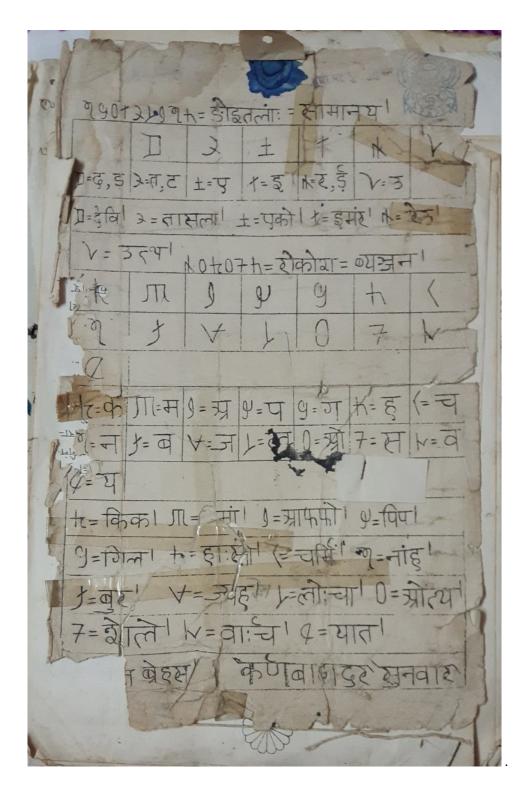


Figure 3: A document written by Karna Bahadur Jentich showing the order and names of Sunuwar letters along with sound values in Devanagari. Note the use of the same letter for dental and retroflex analogues of a given consonant, eg. \square was used for both /d/ and /d/; \square for both /t/ and /t/. Image courtesy of Dev Kumar Sunuwar.

Sunwar	Deva nā- garī	Roman	Sunwar	Deva- nāgarī	Roman/ Arabic
<u> </u>	<u>अ</u>	a	V	उ ऊ	u, ū
J	ब	b	\triangleright	व	V
<	च	c	H	व्हो	vho
I	द	d	Z	य	y
<u> </u>	ए `	e, e	0	0	0
y	ग	g		٩	1
+	ह	h	+	२	2
1	इ ई	i, ī	*	Ŗ	3
\checkmark	ज	j		8	4
t	क	k	B	ሂ	5
1	ल	1	H	६	6
Т	म	m	V	૭	7
D	न	n	X	5	8
Ò	ओ	o	+	९	9
بو	प	p	10	90	10
K	र ड	r, ḍ			
7	स	S			
_2	त ट	t, ţ			

Figure 4: A chart from a Western grammatical treatise on the Sunuwar language showing Sunuwar letters and digits (from Borchers 2008: 47).

किराँती-कोँइच ब्रेःस (जैँ:तिचा लिपि) को मोडिफाइड र रिफम्ड भर्सन नमूना ट्रान्स्लिटरेशन सहित

Consonants व्यन्जन वर्ण			त्र्यन्जन	वर्ण	Vowels स्वर वर्ण		
た /k,क/ く /c,च/ ス /T,ヹ/ み /t,त/	几 /k ^h ,평/ み /th,थ/ の	★ /dz,ज/ Z /T ^h ,ö/ 1 /d,द/ ★	॔ ८ /ŋ,ङ/	5 /2,3{/ /7 /n,-t/ M	** ** ** ** ** ** ** ** ** **		
/p,प/ & /y,य/	1 (/b,ब/ 人 I,ल/ ऽ,श/	ħ	/m,म/ b f v,व/ ।,ह/	phonetically only and all 6 vowels have nasal differences (स्रोत: रापचा ई.२००१ए २००२डी)		

Figure 5: Description of the consonant and vowel letters by Rāpacā (2009: 143).

किराँती-कोँइच ब्रेःस (जैँ:तिचा लिपि) को अङ्क लेखन विधिमा कुनै परिमार्जन नगरिएको सबै भन्दा बायाँ

```
0
       12  sum /sum/ 'zero' शुन्य
ı
       फॻ kA /kā/ 'one' एक
       पि ७२१ nizsi /nizsi/ 'two' दुई
1
       १ इंडिं sAN /sã/ 'three' तीन
人士 le /le/ 'four' चार
В
       ব্ৰ nga /ŋə/ 'five' पाँच
Δ
       Nቴ2 raku/rəku/ 'six' ६
Ш
       <गि tsani /cəni/ 'seven' सात
Χ
       11 sasi /səsi/ 'eight' आठ
       Æं yAN /yã/ 'nine' नौ
       出র2 gau /gəu/ 'ten' বश
```

Figure 6: Description of the digits by Rāpacā (2009: 144).

Koîts Script (कोइँच लिपीको वर्णमाला)

(ways of pronunciation with alphabetical order)

Alphabet	Name	pronouns			
A. Segmental Alphabet					
	a. Primary				
П	Dewi	/d/			
a	Tasla	/t/			
±	Eko	/e/			
t	Imar	/i/			
И	Reu	/r/			
2	Utthi	/u/			
	b. Secondary				
Ħ	Kiki	/k/			
π	Mama	/m/			
9	Appho	/a/			
ą	Pipi	/p/			
9	Gil/Gas	/g/			
ħ	Hamso	/h/			
7	Charmi	/ʧ/			
જ	Na	/n/			
ţ	Bur	/b/			
¥	Jurmi/juljal	/j/			
Ţ	loacha	/١/			
0	Otthi	/o/			
3	Secha/selel	/s/			
k	Warchi	/w/			
Q	Yat	/y/			
Ą	Awa	/b/			
c. Temshyo Apchi					
B	Aal	/a/			
5	Donga	/d/			

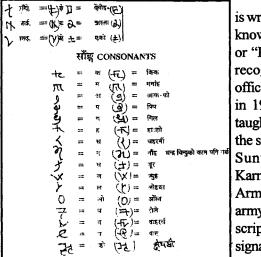
Alphabet	Name	pronouns			
ъ	Thari	/tʰ/			
α	Phar	/f/			
उ	Ngar	/ŋ/			
П	Kha	/k ^h /			
ľ	shyer	/ʃ/			
π	Chhelap	/c ^h /			
Z	Tentu	/t/			
Z	Thele	/tʰ/			
b	Kloko	/?/			
w.	Pwopwo	/þ/			
B. Non-Segmental Alphabet					
~	Nesante	nasal			
•	Sotte?si	stress			
•	Ta Renmaresi	high tone			
•	Di Renmaresi	low tone			
:	Laissi	length			
	C. Digit				
0	Sum	/sum/			
ı	Kaa	/ka/			
#	Niashi	/ni?si/			
*	San	/sa~/			
	Le	/le/			
В	Nga	/ŋa/			
Д	Raku	/raku/			
Ш	Chani	/ʧani/			
Χ	Sasi	/sasi/			
н	Yan	/yaı̃/			



Figure 7: Chart from the magazine *Hamso Sawan*, published August 9, 2021, showing the characters of the Sunuwar script as used today.

Figure 8: A table showing names for chronological and calendrical terminology in Sunuwar (with Hindi equivalents) written in Sunuwar and Devanagari (from Borchers 2008: 48).

7.6 Sunuwar Script (Koinch Bress):



The Sunuwar language is written in Sunuwar script known as "Koinch Bress" or "Koinch Bnu" since its recognition as one of the official languages of Sikkim in 1997. Now, it is being taught upto primary level in the schools of Sikkim. The Sunuwars believe that Karna Bahadur Jehntich, an Army Officer engaged in army signals developed the script combining wireless signals and Brahmi script in the 19th century. The script

has three vowels and 16 consonants. A considerable effort is yet to be put in for further development of writing systems in this script (personal interview with many Sunuwars; Mukhia, R.B. in Lipi Sangalo (ed) by Limboo, R.B. 1998: 113-139)

Figure 9: A description of the Sunuwar script in a book about Sikkim (from Subba 2008: 95).

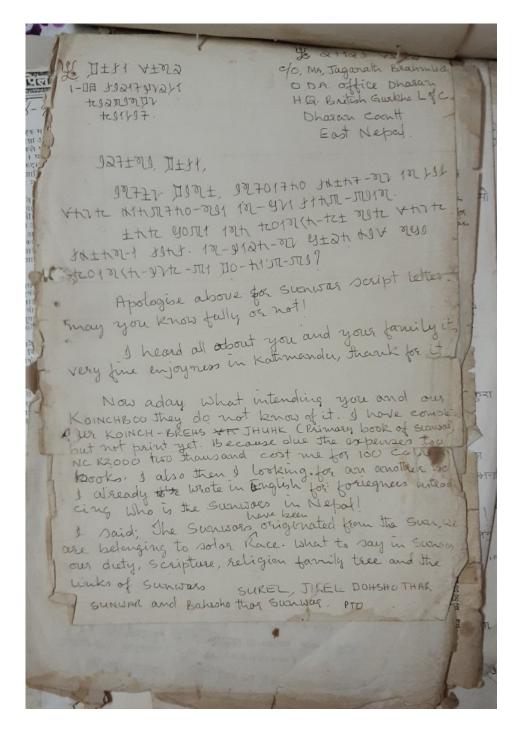


Figure 10: A letter written by K. B. Jentich with portions in the Sunuwar and Latin scripts. Note his reference to "Sunuwar script". Also, note usage of a glyphic variant of the auspicious sign W before the names of the sender and recipient. Image courtesy of Dev Kumar Sunuwar.

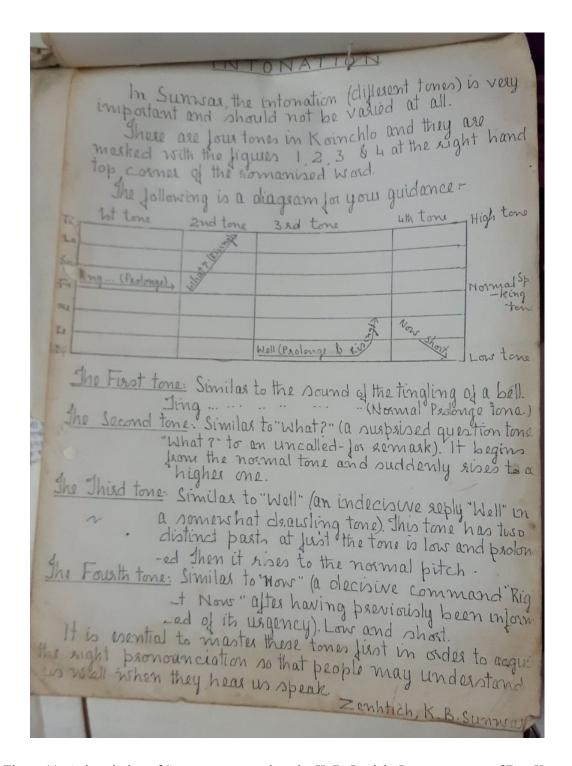


Figure 11: A description of Sunuwar tones written by K. B. Jentich. Image courtesy of Dev Kumar Sunuwar.



(Sikkim Herald Mukhia Edition)

Postal Regd. No. WB/SKM/1/2005

M2时 +

Vol. II

Gangtok: Tuesday

15th November, 2005

Issue No. 21

721GF321 | XXVP9371 | MF THIZINH KE FLIZIKI

THASTA STAM ST. NO. METERO ATT MEM FO

MI NIV MA FREE 32 NY MYKY 12 teat WHOZINATEZ 2524211192 4t-ZINT MYNN MIENO the mitte yourse タタカスク テンパカバナ nex Fraiss mr.



AV IN TILL FYST 90 MINELT HERE-भगतान अध्यक्त तार लगा गार्थि :-ATTO ME T MY 12 SR text MH HERENIE THO HISOKAT MIT

7216337 ZH YROTTXOX 9 MN HIMBIX 10242 YB NYLHI MY TO X7973 TAT NLLD telt latint

लाया अर्थ असार स्थाय REND tot 10 743



SOMINO TYN HA 109 MYNHVATEZZ

THE TO SEVENTE 67 245 mm 10 LENGATOR YOUTHOU BUNNERS IN S MADER LACK STAFF OF NIMOPER

Figure 12: A hand-written edition of the Sikkim Herald, 2005.

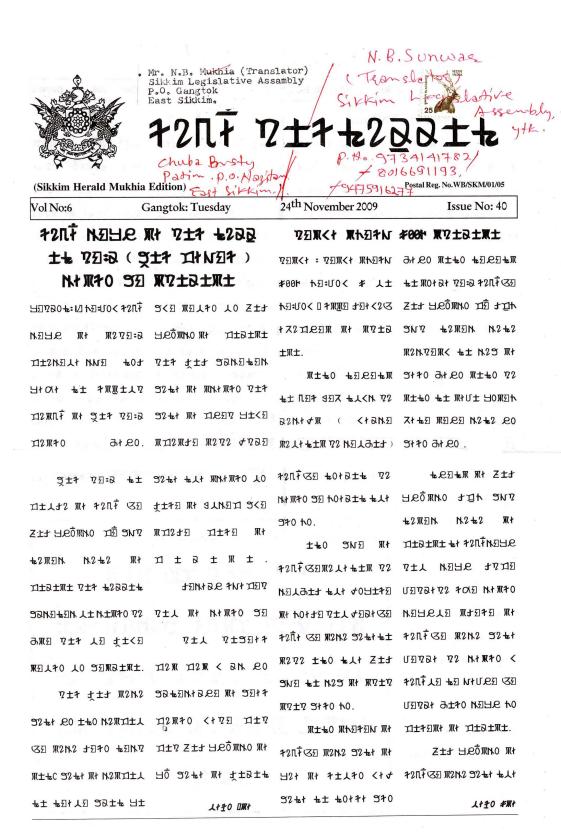
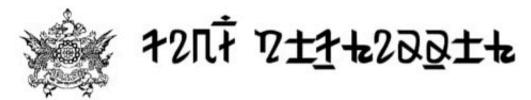


Figure 13: A digitally typeset edition of the Sikkim Herald, 2009.



(Sikkim Herald-Sunuwar(Mukhia) Edition)

Postal Reg. No.WB/SKM/01/2017-19

東京工計:-IX

MB±6-#

TARE FLOW ONOS IN HELDY IN MYTHER



OS THE HER HOLE ±日本0 日2+7日 日主第日主第士 第272 3く4第 91公上町人 もま OMROSE LIS BS 15104 172 開中十 32 開中0 17 11 75 TITA MITTE OFTH ATTACK 827 7E 出土服品土服土.

0## 1M ARCSHEM VE:0 七土 □土人士2 Rt

編 * 0291人

MI RONO 75:0 FOFF 41 34 RO. 17 # # # BE VARE IN BIRV CHESTA ALTA ARESHER WASH # # # PLOT VITVY I TE

172里年日2.

#日口 Wt ±40 Wt 20NO W272 ±40 口面 ±5NB 4±

FER IN EXE VAGE IN LEN GESTEN 1924 HEE ONOS IM SULLTS TERE BEST ATT LEES EFFERER HILL IN SER MIT AUT ONNOSE TES THE THIS EST IN CHROSELLET THE THE 924t 東2爪+足費キくtい Z土ナ 出土B土も 172票キの東t 9日人で

Figure 14: A digitally typeset edition of the *Sikkim Herald*, 2018.



Figure 15: Excerpt from the *Elementary Reader of Sunuwar* showing usage of Rapaca's experimental orthography (from Mukhia, Mukhia, and Rapaca 2004: 40).

तालिका घ किराँती-काँइच ब्रे:समा लेखन नमूना ट्रान्स्लिटरेसन सहित LLMt \summath momiba | Low and and an applied the second and applied to the secon

大子区がり とがです はんにいい rāgiā ləlmits rəwākəli <रागीआ ललिमच् रवाकलि> TERM COS COURCES MS+1 reum tupšo yo məməi <रेउम् तुप्शो यो ममइ> THIM COS COUNTY MYUN p^həsim hemšoyo məməi <फश्यीम् हेम्शो यो ममइ> TEMM COS COULT MET nām k'išo yo məməi <नाम की'शो या ममइ> 20°0 ₽U0°72 gom təšonu <गोम् तशोनु> mεko ləlmits rəwā <मेको ललमिच् खा> ተመተፈመመ ተፈፀታ boisšon bā?bə <बोइस्शोन् बाअ्बः> **1**€25£ <u>7</u>©00091£ s'išon bā?bə <सी'शोन् बाअ्बः> ቱ ነርተውተል ይይር ይቀ kā pāts ā ts'ibpiki <का पाच्आ ची'बपिकी>

M主も士命文 ナ土の八ワイ Mt meken bernimi <मेकेन् बेर्नमी>

州土七土⊕文 7豆 み土<u>八</u>ヤイ Mł meken nā thernimi

<मेकेन् ना थेर्नमी> ጠ±ቈ±፫ ታ2፡ንፒ∤ጠ∤ meken dz'unimi <मेकेन् जु'निमी> Mt 72 minu <मिन्> #2M NO SUE Uto A SET TH MT kumšo pəšyā šyil pāinimi <कुम्शो पश्या श्यील पाइनिमी> U**Ç**U 14EBM ©0**3+±**M ∑\9\0 šyen meko rāgin ā <श्यङ् मेको रागीन्आ> NTF SZI wek pāts ā <वेक् पाच्आ> くとより立を了り ts'ib ānkəli <चीःबआन्कलि> ãke rāgimi genāiyo ləlmits məmibə <आँके रागीमि गेनाइयो लल्मिच् ममिबः> (Source: Rapacha (2001a: 24)

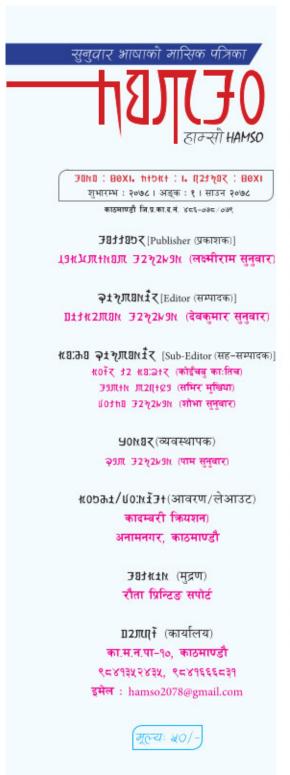
Figure 16: Sunuwar phrases given in Sunuwar, Latin, and Devanagari transliteration, showing Rapaca's experimental orthography (adapted from Rāpacā 2009: 145–146).



Figure 17: An excerpt from a 2009 wall calendar in Sunuwar published by the Sikkim Sunuwar Mukhia Koinchbu.



Figure 18: An excerpt from a 2010 wall calendar in Sunuwar published by the Sikkim Sunuwar Mukhia Koinchbu.



∃ध1195∃+ (प्रकाशकीय)

90 ₹85₹82 (हामी आयाँ)

हाम्सोमार्फत् कोइँच सुनुवार सामू हामी एउटा सदस्यका रुपमा जोडिएका छौँ। नाम कमाउने अभिलासा होइन, मनग्गे धन कमाउने लोभ पनि होइन। कोइँच समक्ष केही सिकाउन र सिक्ने जोशमा जोडिएका हौँ। कोइँच समुदायको आवाजलाई बुलन्द गर्न थोरै भएपनि हाम्सोले योगदान अवश्य पुऱ्याउने छ। हाम्रो यस प्रयासमा तपाईँहरू पनि जोडिनु आवश्यक छ।

अरूले कसो गर्छन् भन्नेतिर हैन, हामीचाहिँ कोइँचसामु शब्द र अक्षरको प्रतिबद्धता लिएर आएका हौँ। यो मैदानमा आफैँले उठाएका प्रत्येक मुद्दा, विषय र त्यसका पुष्टिकरणमा केन्द्रित रहनेछौँ। यो आफैँसँगको निरन्तरको स्पर्धा हुनेछ।

अवस्था र परिवेश जोखिमपूर्ण छ । यो जोखिमलाई नजरअन्दाज नगरी थोरै भए पनि जानकारी दिन सिकन्छ भन्ने धेयले नै हामी आएका हौँ । हामी हाम्सोमार्फत् कोइँचको संस्कार, संस्कृति र परम्परालाई जगेर्ना गर्ने उद्देश्यले मैदानमा औपचारिक रुपमा आएका छौँ । हामी कामका माध्यमले असली 'बधाई'को लोभमा छौँ, कोइँचभाषी पाठकको मायाका लागि 'लालची' छौँ ।

प्रकाशक

Figure 19: Excerpt from the magazine *Hamso Sawan*, published August 9, 2021, showing current usage of the Sunuwar script

ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646.1

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 the Sunuwar script in Unico	ode
2. Requester's name: Anshuman Pandey (pandey@umich.edu) / Script Encoding	Initiative (SEI)
3. Requester type (Member body/Liaison/Individual contribution): Expert con	tribution
4. Submission date: 2021-1.	2-06
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: Sunuwar	
b. The proposal is for addition of character(s) to an existing block:	
Name of the existing block:	
2. Number of characters in proposal:	44
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	e collection)
C-Major extinct D-Attested extinct E-Minor extinct	,
F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable u	isage symbols
4. Is a repertoire including character names provided?	Yes
a. If YES, are the names in accordance with the "character naming guidelines"	700
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 p	aublishing the
standard?	Jubilishing the
Anshuman Pandey	
b. Identify the party granting a license for use of the font by the editors (include address,	e-mail ftp-site etc.)
Anabuman Banday (nanday@ymich adu)	
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	Vas
b. Are published examples of use (such as samples from newspapers, magazines, or ot	
of proposed characters attached?	nor sources;
7. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) su	ch ac input
presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose infor	
presentation, sorting, searoning, indexing, transiteration etc. (if yes piease encose into	mation): 700
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 ch	

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.

[.] Form number: N3902-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)

C. Technical - Justification

1. Has this proposal for addition of cha	No			
If YES explain				
	s of the user community (for example: National Body,	Yes		
user groups of the script or chara	Sunuwar Welfare Society; Dev Kumar Sunuwar (Indigenous I			
If YES, with whom?	Foundation); Dr. Lal-Shyakarelu Rapacha	vieuia		
If VES, available relevant				
If YES, available relevant	for the proposed characters (for example:			
	technology use, or publishing use) is included?	Yes		
	for print and digital publications; see proposal for details			
		Common		
	characters (type of use; common or rare) See proposal for details	Common		
Reference:				
5. Are the proposed characters in curr		Yes		
If YES, where? Reference:	Slkkim (India) and Nepal. See proposal for details.			
	e principles in the P&P document must the proposed characters b			
in the BMP?		No		
If YES, is a rationale pr	ovided?			
If YES, reference:				
	kept together in a contiguous range (rather than being scattered)?	Yes		
	be considered a presentation form of an existing	A.I		
character or character sequence		No		
If YES, is a rationale for	rits inclusion provided?			
If YES, reference:				
	be encoded using a composed character sequence of either			
existing characters or other prop		No		
If YES, is a rationale for	its inclusion provided?			
If YES, reference:				
	r(s) be considered to be similar (in appearance or function)			
to an existing character?		Yes		
If YES, is a rationale fo		Yes		
If YES, reference:	See proposal for details.			
	ombining characters and/or use of composite sequences?	Yes		
If YES, is a rationale for such us		Yes		
If YES, reference:	Common combining signs			
Is a list of composite sequences	and their corresponding glyph images (graphic symbols) provided	?		
If YES, reference:				
12. Does the proposal contain charact	ers with any special properties such as			
control function or similar seman		No		
If YES, describe in deta	il (include attachment if necessary)			
13. Does the proposal contain any Ide	ographic compatibility characters?	No		
If YES, are the equivalent corresponding unified ideographic characters identified?				
If YES, reference:				