Proposal to encode the Sunuwar script in Unicode

Anshuman Pandey

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

September 10, 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 existing 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

A request has been submitted to register the code "sunu" 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:

𑴈	𑴊	𑴋	𑴌	𑴍	𑴎	𑴏	𑴐	𑴑	𑴒	𑴓
devi	*tasla*	*eko*	*imar*	*reu*	*utthi*	*kik*	*ma*	*appho*	*pip*	*gil*
d, ɖ	t, ʈ	e	i	r, ɽ	u	k	m	a	p	g
𑴔	𑴕	𑴖	𑴗	𑴘	𑴙	𑴚	𑴛	𑴜	𑴝	𑴞
hamso	*carmi*	*nah*	*bur*	*jyah*	*loacha*	*otthi*	*shyele*	*varca*	*yat*	*ava*
h	tʃ	n	b	ʤ	l	o	s, ʃ	v	y	ɓ

𑴠	𑴡	𑴢	𑴣	𑴤	𑴥	𑴦	𑴧	𑴨	𑴩
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: 𑴓 /a/, 𑴖 /i/, 𑴋 /e/, 𑴎 /u/, 𑴚 /o/
- Vowel length was not distinguished orthographically
- A letter could also represent a retroflex consonant, eg. 𑴈 for /d/, /ɖ/; 𑴊 for /t/, /ʈ/; 𑴍 for /r/, /ɽ/
- Aspirated consonants were written as digraphs using 𑴔 /h/ for /ʰ/, eg. 𑴏𑴔 /kʰ/, 𑴕𑴔 /tʃʰ/
- The letter 𑴘 /n/ was used for both vowel nasalization and consonantal nasals, eg. 𑴓𑴘 /an/, 𑴘𑴊 /nt/
- The velar nasal /ŋ/ was written as the digraph 𑴘𑴞 /ng/
- The letter 𑴝 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:

aal	*donga*	*thari*	*phar*	*ngar*	*kha*	*shyer*	*chelap*	*tentu*	*thele*	*kloko*	*laissi*
a	ɖ	tʰ	pʰ	ŋ	kʰ	ʃ	tʃʰ	ʈ	ʈʰ	ʔ	ː

These extensions enabled more precise representation of the Sunuwar language:

- The letter [aal], borrowed from Limbu, was introduced for writing long /a/
- Writing of the /ʔ/ glottal stop using the letter [kloko]
- Separate letters for retroflex consonants [tentu] /ʈ/, [thele] /ʈʰ/, [donga] /ɖ/
- Distinctive letters for aspirated consonants, eg. [kha] /kʰ/, [chelap] /tʃʰ/, [thari] /tʰ/, [phar] /pʰ/
- The letter [shyer] for the palatal sibilant /ʃ/
- Usage of [ngar] for the velar nasal /ŋ/
- Indication of vowel length using the sign [laissi], eg. [Sunuwar glyph] /i/, [Sunuwar glyph][laissi] /iː/

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. [Sunuwar glyph] and [tentu] /ʈ/
- Aspirated consonants may be expressed using digraphs or a dedicated letter, eg. [Sunuwar glyphs] or [chelap] for /tʃʰ/
- The velar nasal may be written as [Sunuwar glyphs] /ng/ or as [ngar] /ŋ/
- Vowel nasalization may be written using [Sunuwar glyph] /n/ or the nasalization sign, eg. [Sunuwar glyphs] or [Sunuwar glyph] for /ĩ/

2.3 Experimentation

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

sangmilu *sangkirs* *sangrums* *taslathenk*

The experimental orthography had the following features:

- The alphabetic consonant letters were interpreted as possessing an inherent vowel, eg, [Sunuwar glyph] /d/ = /da/. The sign [Sunuwar glyph] *sangmilu* was created to indicate suppression of the inherent vowel; eg. [Sunuwar glyph] /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. [Sunuwar glyphs] /daː/

- The consonant [Sunuwar glyph] /r/ was given characteristics similar to Devanagari र *ra*. The [Sunuwar glyph] *sangkirs* represents cluster-initial /r/ and functions like Devanagari *repha*: it was written as part of the following orthographic syllable, eg. [Sunuwar glyph] /rda/. If a 'dependent' vowel is part of the following syllable, the *sangkirs* is written above it: [Sunuwar glyphs] /rdaː/

- 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. ⴷ̭ /dra/
- Nasalization was indicated using ◌̈̄ *taslathenk*, eg. ੭̈ /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. < for cluster-initial < *c* and ⊓ for ⫪ *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:

◌̃	◌̍	◌́	◌̀
nesante	*sotte'si*	*ta tenmaresi*	*di renmaresi*

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' /koĩtʃ/ (ᵼ0+ϡ< or ᵼ0ī<, Devanagari कोइँच *koĩc*). 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 जैँतिच ब्रेसे *jẽtic 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 *Kirãtī-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 Ṅāvacā-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
𑿄	SUNUWAR LETTER DEVI	*devi*	/d/, /ɖ/	द, ड
𑿅	SUNUWAR LETTER TASLA	*tasla*	/t/, /ʈ/	त, ट
𑿆	SUNUWAR LETTER EKO	*eko*	/e/	ए
𑿇	SUNUWAR LETTER IMAR	*imar*	/i/	इ
𑿈	SUNUWAR LETTER REU	*reu*	/r/, /ɽ/, /ɭ/	र, ड़
𑿉	SUNUWAR LETTER UTTHI	*utthi*	/u/	उ
𑿊	SUNUWAR LETTER KIK	*kik*	/k/	क
𑿋	SUNUWAR LETTER MAMA	*mama*	/m/	म
𑿌	SUNUWAR LETTER APPHO	*appho*	/ə/	अ
𑿍	SUNUWAR LETTER PIP	*pip*	/p/	प

𑿉	SUNUWAR LETTER GIL	*gil*	/g/	ग
𑿃	SUNUWAR LETTER HAMSO	*hamso*	/h/	ह
𑿇	SUNUWAR LETTER CARMI	*carmi*	/tʃ/	च
𑿎	SUNUWAR LETTER NAH	*nah*	/n/	न, ण
𑿄	SUNUWAR LETTER BUR	*bur*	/b/	ब
𑿈	SUNUWAR LETTER JYAH	*jyah*	/ʤ/	ज
𑿏	SUNUWAR LETTER LOACHA	*loacha*	/l/, /ɭ/	ल
𑿐	SUNUWAR LETTER OTTHI	*otthi*	/o/	ओ
𑿅	SUNUWAR LETTER SHYELE	*shyele*	/s/	स
𑿋	SUNUWAR LETTER VARCA	*varca*	/v/	व
𑿌	SUNUWAR LETTER YAT	*yat*	/y/	य
𑿆	SUNUWAR LETTER AVA	*ava*	/ɓ/	ब
𑿁	SUNUWAR LETTER AAL	*aal*	/aː/	आ
𑿓	SUNUWAR LETTER DONGA	*donga*	/ɖ/	ड
𑿔	SUNUWAR LETTER THARI	*thari*	/tʰ/	थ
𑿖	SUNUWAR LETTER PHAR	*phar*	/pʰ/	फ
𑿍	SUNUWAR LETTER NGAR	*ngar*	/ŋ/	ङ
𑿗	SUNUWAR LETTER KHA	*kha*	/kʰ/	ख
𑿑	SUNUWAR LETTER SHYER	*shyer*	/ʃ/	श, ष
𑿘	SUNUWAR LETTER CHELAP	*chelap*	/tʃʰ/	छ
𑿒	SUNUWAR LETTER TENTU	*tentu*	/ʈ/	ट
𑿕	SUNUWAR LETTER THELE	*thele*	/ʈʰ/	ठ
𑿙	SUNUWAR LETTER KLOKO	*kloko*	/ʔ/	अ◌

Auspicious Sign

	Unicode Character Name	Name	Value	Devanagari
𑿫	SUNUWAR SIGN PVO	*pvo*	/ɓ̥/ (/ɸ/)	—

Digits

	Unicode Character Name	Name	Value	Devanagari
𑿀	SUNUWAR DIGIT ZERO	*sum*	0	०
𑿁	SUNUWAR DIGIT ONE	*ka*	1	१

ꣴ	SUNUWAR DIGIT TWO	*ni'ashi*	2	ꣲ
ꣵ	SUNUWAR DIGIT THREE	*san*	3	ꣳ
ꣶ	SUNUWAR DIGIT FOUR	*le*	4	ꣴ
ꣷ	SUNUWAR DIGIT FIVE	*nga*	5	ꣵ
꣸	SUNUWAR DIGIT SIX	*raku*	6	ꣶ
꣹	SUNUWAR DIGIT SEVEN	*cani*	7	ꣷ
꣺	SUNUWAR DIGIT EIGHT	*sasi*	8	꣸
ꣻ	SUNUWAR DIGIT NINE	*yan*	9	꣹

5.1 Notes on proposed characters

The 𑼏 SUNUWAR SIGN PVO represents an auspicious syllable, which is articulated as the unvoiced bilabial implosive /ɓ̥/ (formerly /ɸ/) 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:

𑼀	𑼁	𑼂	𑼃	𑼄	𑼅	𑼆	𑼇	𑼈	𑼉	𑼊
devi	*tasla*	*eko*	*imar*	*reu*	*utthi*	*kik*	*ma*	*appho*	*pip*	*gil*
𑼋	𑼌	𑼍	𑼎	𑼏	𑼐	𑼑	𑼒	𑼓	𑼔	𑼕
hamso	*carmi*	*nah*	*bur*	*jyah*	*loacha*	*otthi*	*shyele*	*varca*	*yat*	*ava*
𑼖	𑼗	𑼘	𑼙	𑼚	𑼛	𑼜	𑼝	𑼞	𑼟	𑼠
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

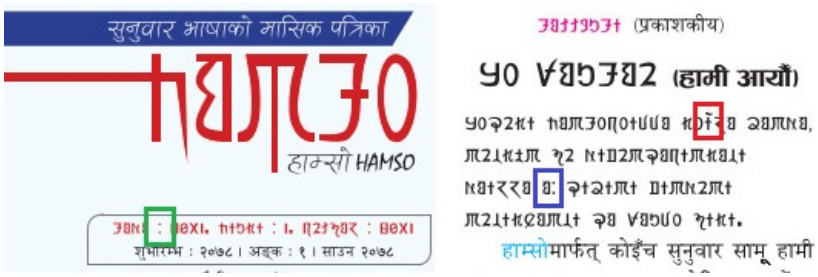
There are several diacritic signs used in the Sunuwar script. Some are currently used, while others are no longer used. There is no requirement to encode these diacritics as Sunuwar-specific characters. Rather, all of the current and legacy diacritics can be represented using combining signs already present in Unicode. However, the design of the diacritics in a Sunuwar font should match the ductus of the Sunuwar characters. The diacritics and their recommended representations are enumerated below. The general combining signs have been specified as script extensions for Sunuwar.

7.1 Conventional diacritics

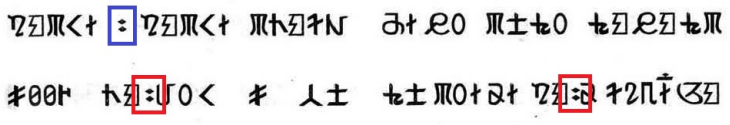
The following diacritic signs are used conventionally:

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

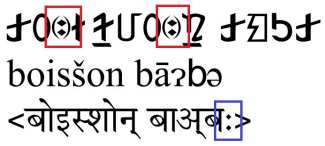
Usage of *nesante* (red) and *laissi* (blue) are shown below; also shown is usage of the colon (green), which is identical in this Sunuwar font to *laissi*.



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.



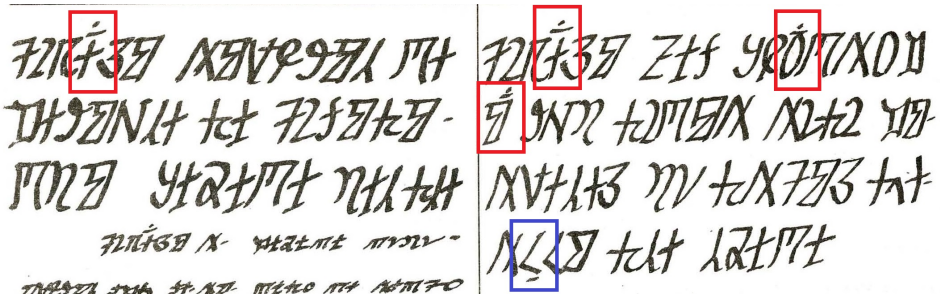
7.2 Legacy diacritics

The following characters were intoduced by Rapaca as part of an orthographic experiment. They were not widely adopted. The Sunuwar experts consulted during the development of the proposal have stated that these are not part of the conventional script. However, they 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.

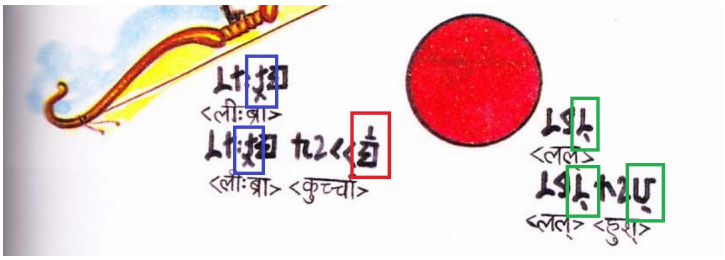
	Name	Function	Unicode representation
◌̲	*sangmilu*	*halanta* / vowel silencer	◌̲ U+0331 COMBINING MACRON BELOW
◌̐	*taslathenk*	nasal sign	◌̐ U+0310 COMBINING CANDRABINDU
◌̪	*sangkirs*	*repha*	◌̪ U+1DF5 COMBINING UP TACK ABOVE
◌̭	*sangrums*	*ra-vattu* / trill mark	◌̭ U+032D COMBINING CIRCUMFLEX ACCENT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

It is recommended to use U+0310 for representing *taslathenk* even though the shapes of the signs are different: U+0310 has a curved element while that element in *taslathenk* is horizontal. The U+0310 carries the same semantics as *taslathenk* 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 ◌् 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	◌́ 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 HAMSO;Lo;0;L;;;;;N;;;;;
11BCC;SUNUWAR LETTER CARMI;Lo;0;L;;;;;N;;;;;
11BCD;SUNUWAR LETTER NAH;Lo;0;L;;;;;N;;;;;
11BCE;SUNUWAR LETTER BUR;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;o;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
11BE2l;AL         # So         SUNUWAR SIGN PVO
11BF0..11BF9;NU   # Nd    [10] SUNUWAR DIGIT ZERO..SUNUWAR DIGIT NINE
```

Script Extensions: `ScriptExtensions.txt`

```
# Script_Extensions=Sunu

02D0          ; Sunu # Lm       MODIFIER LETTER TRIANGULAR COLON
0300          ; Sunu # Mn       COMBINING GRAVE ACCENT
0301          ; Sunu # Mn       COMBINING ACUTE ACCENT
0303          ; Sunu # Mn       COMBINING TILDE
030D          ; Sunu # Mn       COMBINING VERTICAL LINE ABOVE
0310          ; Sunu # Mn       COMBINING CANDRABINDU
032D          ; Sunu # Mn       COMBINING CIRCUMFLEX ACCENT BELOW
0331          ; Sunu # Mn       COMBINING MACRON BELOW
1DF5          ; Sunu # Mn       COMBINING UP TACK ABOVE
```

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. [illegible] [= *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, माझ्र र पल्लो किराँत. काठमांडौ : किरांतविज्ञान अध्ययन संस्थान.

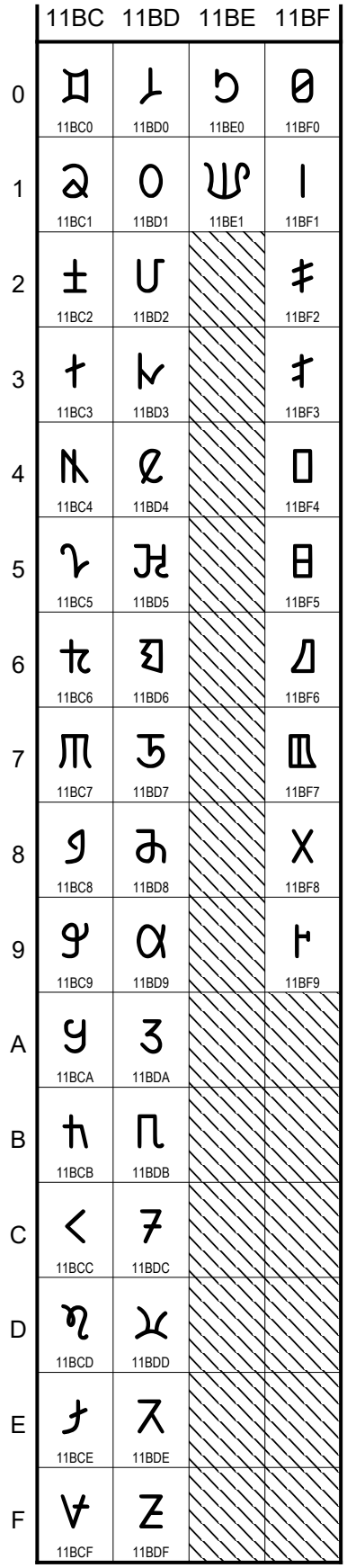
Subba, Jash Raj. 2008. *The Evolution of Man and the Modern Society in Mountainous Sikkim*. Delhi: Gyan Publishing House.

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.

The project to encode the Sunuwar script 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.



The script is known indigenously as 'Koinch brese'

Letters

11BC0 SUNUWAR LETTER DEVI
11BC1 SUNUWAR LETTER TASLA
11BC2 SUNUWAR LETTER EKO
11BC3 SUNUWAR LETTER IMAR
11BC4 SUNUWAR LETTER REU
11BC5 SUNUWAR LETTER UTTHI
11BC6 SUNUWAR LETTER KIK
11BC7 SUNUWAR LETTER MA
11BC8 SUNUWAR LETTER APPHO
11BC9 SUNUWAR LETTER PIP
11BCA SUNUWAR LETTER GIL
11BCB SUNUWAR LETTER HAMSO
11BCC SUNUWAR LETTER CARMI
11BCD SUNUWAR LETTER NAH
11BCE SUNUWAR LETTER BUR
11BCF SUNUWAR LETTER JYAH
11BD0 SUNUWAR LETTER LOACHA
11BD1 SUNUWAR LETTER OTTHI
11BD2 SUNUWAR LETTER SHYELE
11BD3 SUNUWAR LETTER VARCA
11BD4 SUNUWAR LETTER YAT
11BD5 SUNUWAR LETTER AVA
11BD6 SUNUWAR LETTER AAL
11BD7 SUNUWAR LETTER DONGA
11BD8 SUNUWAR LETTER THARI
11BD9 SUNUWAR LETTER PIIAR
11BDA SUNUWAR LETTER NGAR
11BDB SUNUWAR LETTER KHA
11BDC SUNUWAR LETTER SHYER
11BDD SUNUWAR LETTER CHELAP
11BDE SUNUWAR LETTER TENTU
11BDF SUNUWAR LETTER THELE
11BE0 SUNUWAR LETTER KLOKO

Auspicious sign

11BE1 SUNUWAR SIGN PVO

Digits

11BF0 SUNUWAR DIGIT ZERO
11BF1 SUNUWAR DIGIT ONE
11BF2 SUNUWAR DIGIT TWO
11BF3 SUNUWAR DIGIT THREE
11BF4 SUNUWAR DIGIT FOUR
11BF5 SUNUWAR DIGIT FIVE
11BF6 SUNUWAR DIGIT SIX
11BF7 SUNUWAR DIGIT SEVEN
11BF8 SUNUWAR DIGIT EIGHT
11BF9 SUNUWAR DIGIT NINE

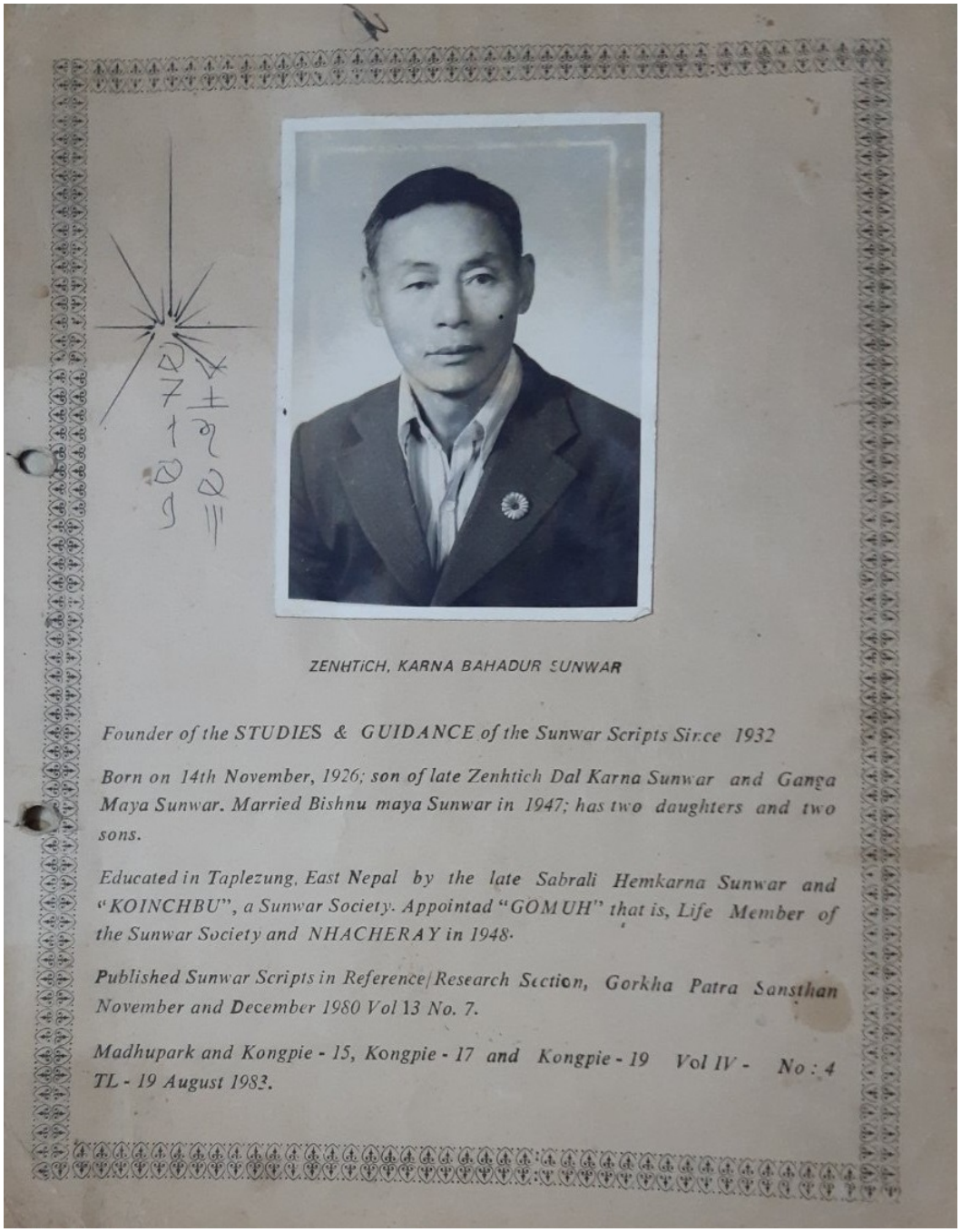


Figure 1: Biographical note about K. B. Jentich. Image courtesy of Dev Kumar Sunuwar.

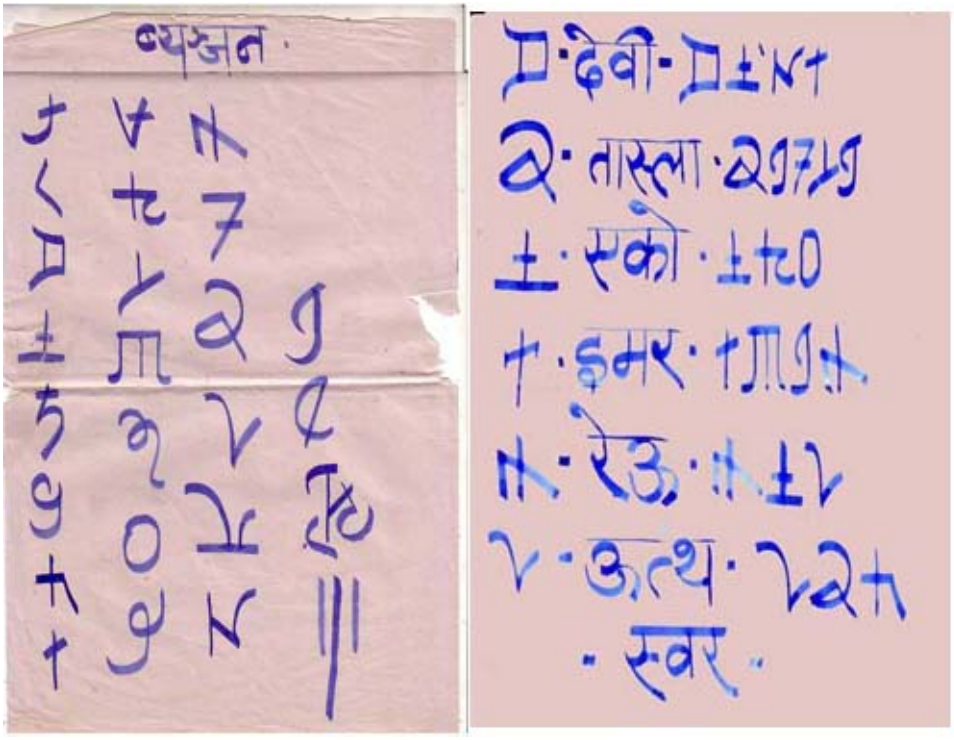


Figure 2: Sunuwar letterforms reported written by K. B. Jentich (from kiratsunuwar.org.np)

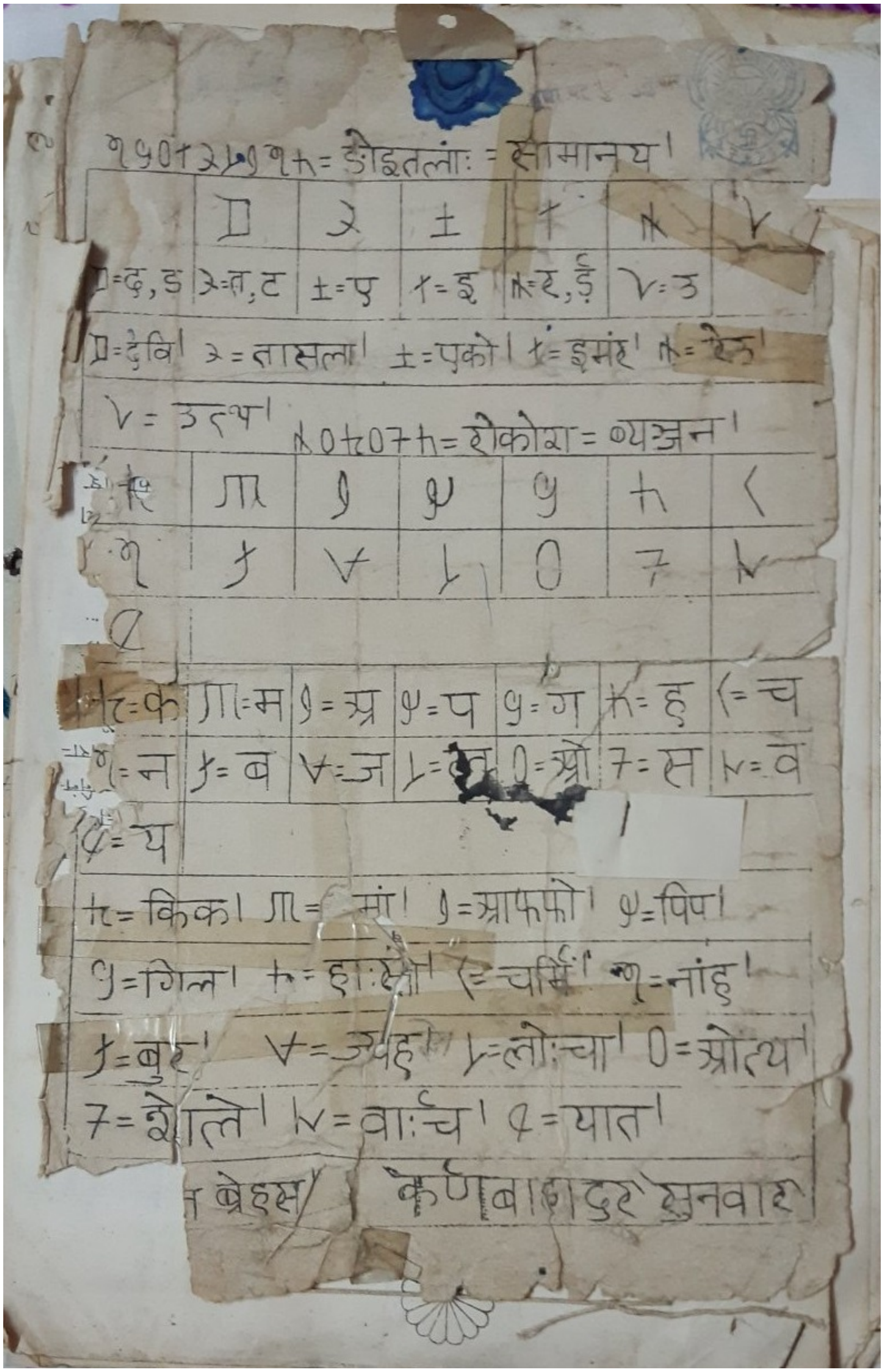


Figure 3: A document written by Karna Bahadur Jentich showing the order and names of Sunuwar letters along with sound values in Devanagari. Note the use of the same letter for dental and retroflex analogues of a given consonant, eg. 𑯀 was used for both /d/ and /ḍ/; 𑯁 for both /t/ and /ṭ/. Image courtesy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

Sunwar	Deva nā-garī	Roman
	अ	a
	ब	b
	च	c
	द	d
	ए `	e, e
	ग	g
	ह	h
	इ ई	i, ī
	ज	j
	क	k
	ल	l
	म	m
	न	n
	ओ	o
	प	p
	र ड	r, ḍ
	स	s
	त ट	t, ṭ

Sunwar	Deva-nāgarī	Roman/ Arabic
	उ ऊ	u, ū
	व	v
	व्हो	vho
	य	y
	०	0
	१	1
	२	2
	३	3
	४	4
	५	5
	६	6
	७	7
	८	8
	९	9
	१०	10

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

किराँती–कौँइच ब्रेःस (जेँःतिचा लिपि) को मोडिफाइड र रिफम्ड भर्सन नमूना ट्रान्स्लिटरेशन सहित

Consonants व्यन्जन वर्ण

/k,क/	/k^{h},ख/	/g,ग/	/ŋ,ङ/	/ʔ,अ्/
/c,च/		/dz,ज/		
/T,ट/		/T^{h},ठ/		
/t,त/	/th,थ/	/d,द/		/n,न/
/p,प/	/p^{h},फ/	/b,ब/		/m,म/
/y,य/	/r,र/	/l,ल/		/w,व/
	/s,स/	/š,श/		/h,ह/

Vowels स्वर वर्ण

/i,इ/		/u,उ/
	/ə,अ/	/o,आ/
/ɛ,ए/		/ā,आ/

*means optional and /ɛ, ā, i, u/ are also subject to have such lengthening phonetically only and all 6 vowels have nasal differences

(स्रोतः रापचा ई.२००१ए २००२डी)

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

किराँती–कौँइच ब्रेःस (जेँःतिचा लिपि) को अङ्क लेखन विधिमा कुनै परिमार्जन नगरिएको सबै भन्दा बायाँ

0	*sum* /sum/ 'zero' शुन्य
1	*kA* /kā/ 'one' एक
2	*niʔsi* /niʔsi/ 'two' दुई
3	*sAN* /sã/ 'three' तीन
4	*le* /lɛ/ 'four' चार
5	*nga* /ŋə/ 'five' पाँच
6	raku /rəku/ 'six' ६
7	*tsani* /cəni/ 'seven' सात
8	*sasi* /səsi/ 'eight' आठ
9	*yAN* /yã/ 'nine' नौ
10	gau /gəu/ 'ten' दश

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

Koĩts Script (कोइँच लिपीको वर्णमाला)

(ways of pronunciation with alphabetical order)

Alphabet	Name	pronouns
A. Segmental Alphabet		
a. Primary		
	Dewi	/d/
	Tasla	/t/
	Eko	/e/
	Imar	/i/
	Reu	/r/
	Utthi	/u/
b. Secondary		
	Kiki	/k/
	Mama	/m/
	Appho	/a/
	Pipi	/p/
	Gil/Gas	/g/
	Hamso	/h/
	Charmi	/ʧ/
	Na	/n/
	Bur	/b/
	Jurmi/juljal	/j/
	loacha	/l/
	Otthi	/o/
	Secha/selel	/s/
	Warchi	/w/
	Yat	/y/
	Awa	/ɓ/
c. Temshyo Apchi		
	Aal	/a/
	Donga	/d/

Alphabet	Name	pronouns
	Thari	/tʰ/
	Phar	/f/
	Ngar	/ŋ/
	Kha	/kʰ/
	shyer	/ʃ/
	Chhelap	/cʰ/
	Tentu	/ʈ/
	Thele	/ʈʰ/
	Kloko	/ʔ/
	Pwopwo	/ƥ/
B. Non-Segmental Alphabet		
˜	Nesante	nasal
ˈ	Sotteʔsi	stress
ˊ	Ta Renmarẽsi	high tone
ˋ	Di Renmarẽsi	low tone
:	Laissi	length
C. Digit		
	Sum	/sum/
	Kaa	/ka/
	Niashi	/niʔsi/
	San	/sã/
	Le	/le/
	Nga	/ŋa/
	Raku	/raku/
	Chani	/ʧani/
	Sasi	/sasi/
	Yan	/yã/

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

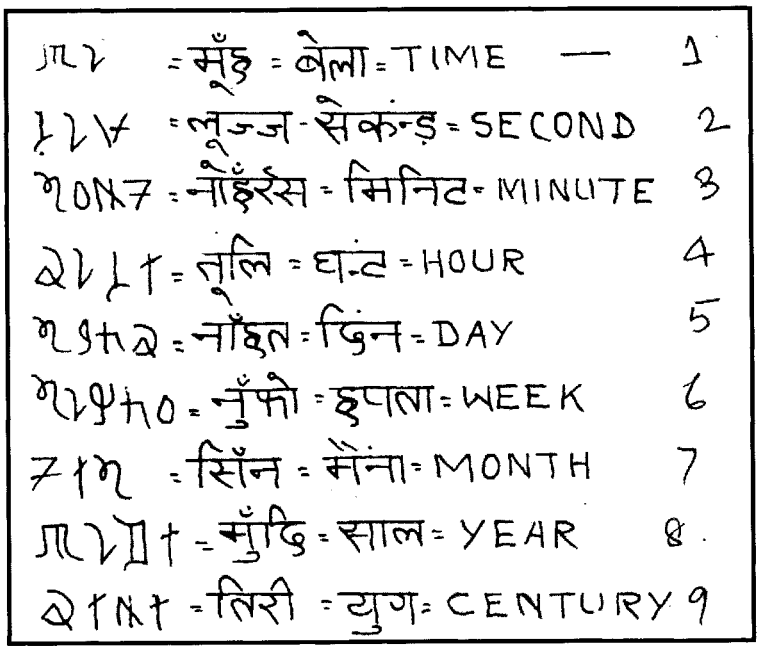
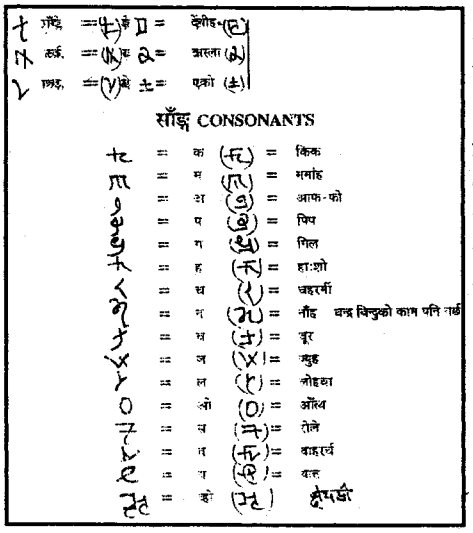


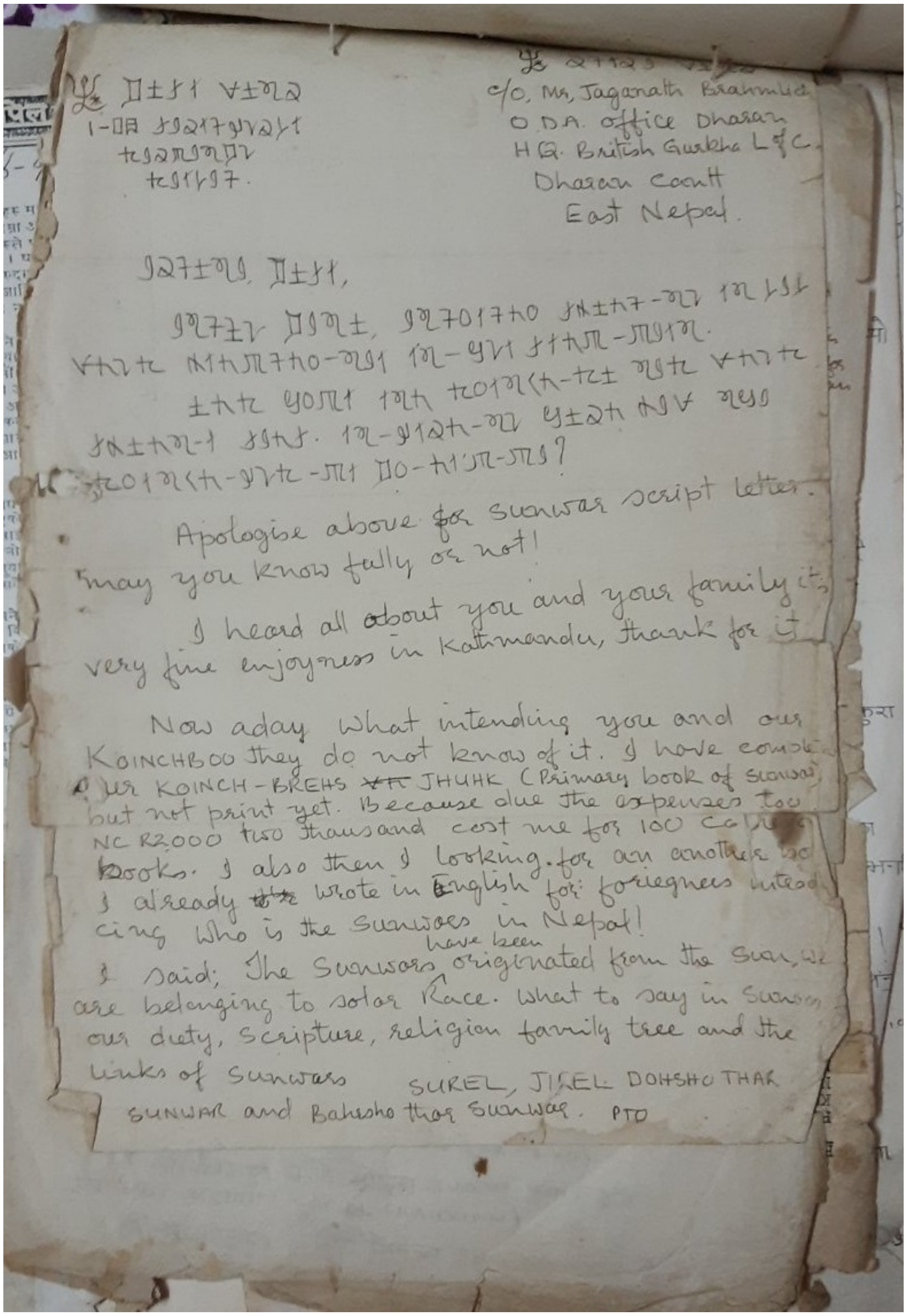
Figure 8: A table showing names for chronological and calendrical terminology in Sunuwar (with Hindi equivalents) written in Sunuwar and Devanagari (from Borchers 2008: 48).

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).



c/o, Mr, Jaganath Brahmu..
O.D.A. office Dharan
H.Q. British Gurkha L of C.
Dharan Contt
East Nepal.

Apologise above for Sunuwar script letter. may you know fully or not!

I heard all about you and your family it, very fine enjoyness in Kathmandu, thank for it.

Now aday what intending you and our KOINCHBOO they do not know of it. I have compl.. our KOINCH-BREHS JHUHK (Primary book of Sunuwar) but not print yet. Because due the expenses too NC R2000 two thousand cost me for 100 co.. books. I also then I looking for an another .. I already wrote in English for foriegners introd.. cing who is the Sunuwars in Nepal!

I said; The Sunuwars have been originated from the Sun, we are belonging to solar Race. What to say in Sunu.. our duty, scripture, religion family tree and the links of Sunuwars SUREL, JIREL DOHSHO THAR SUNUWAR and Bahsho thar Sunuwar. PTO

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 𑯀 before the names of the sender and recipient. Image courtesy of Dev Kumar Sunuwar.

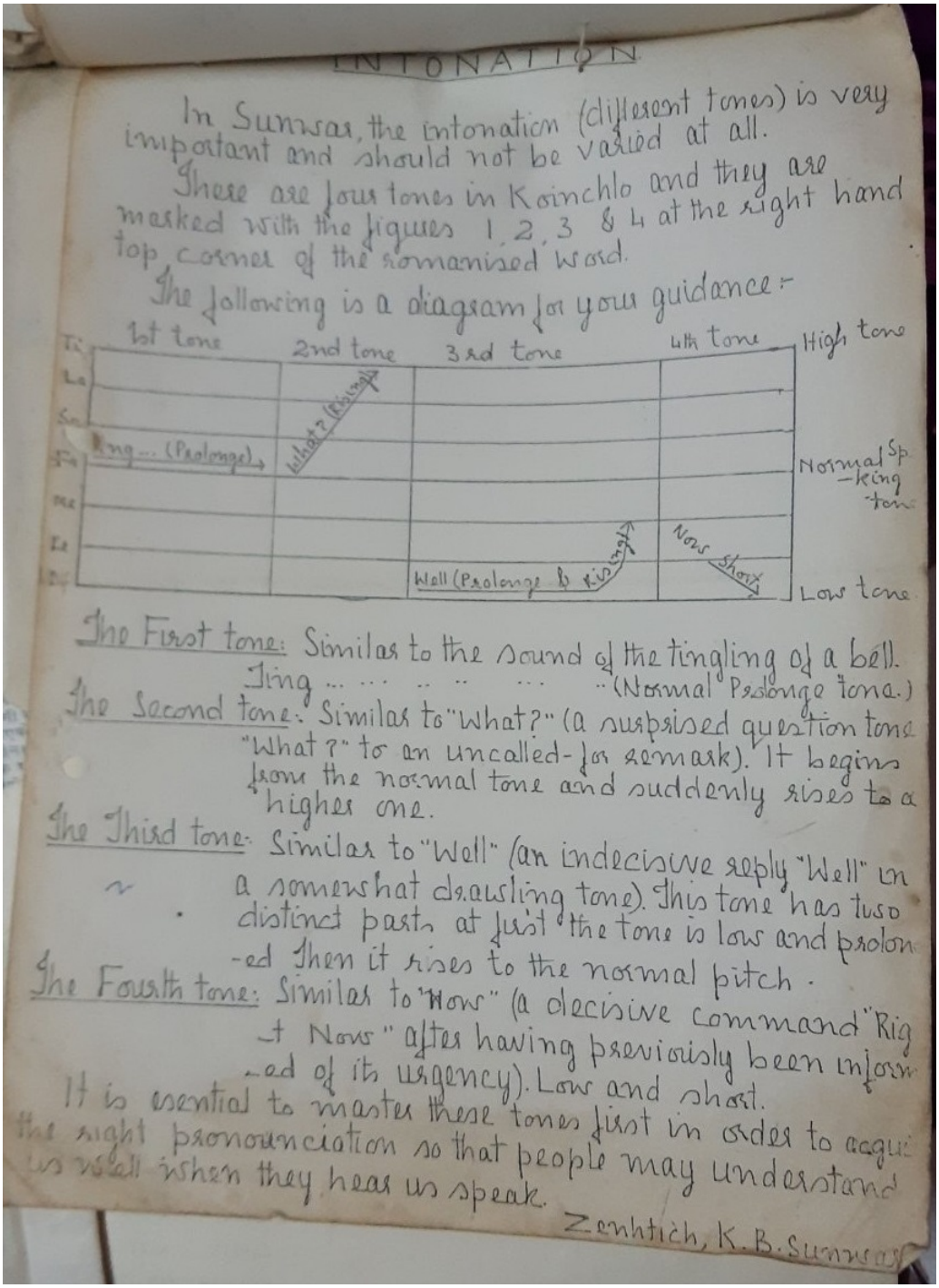
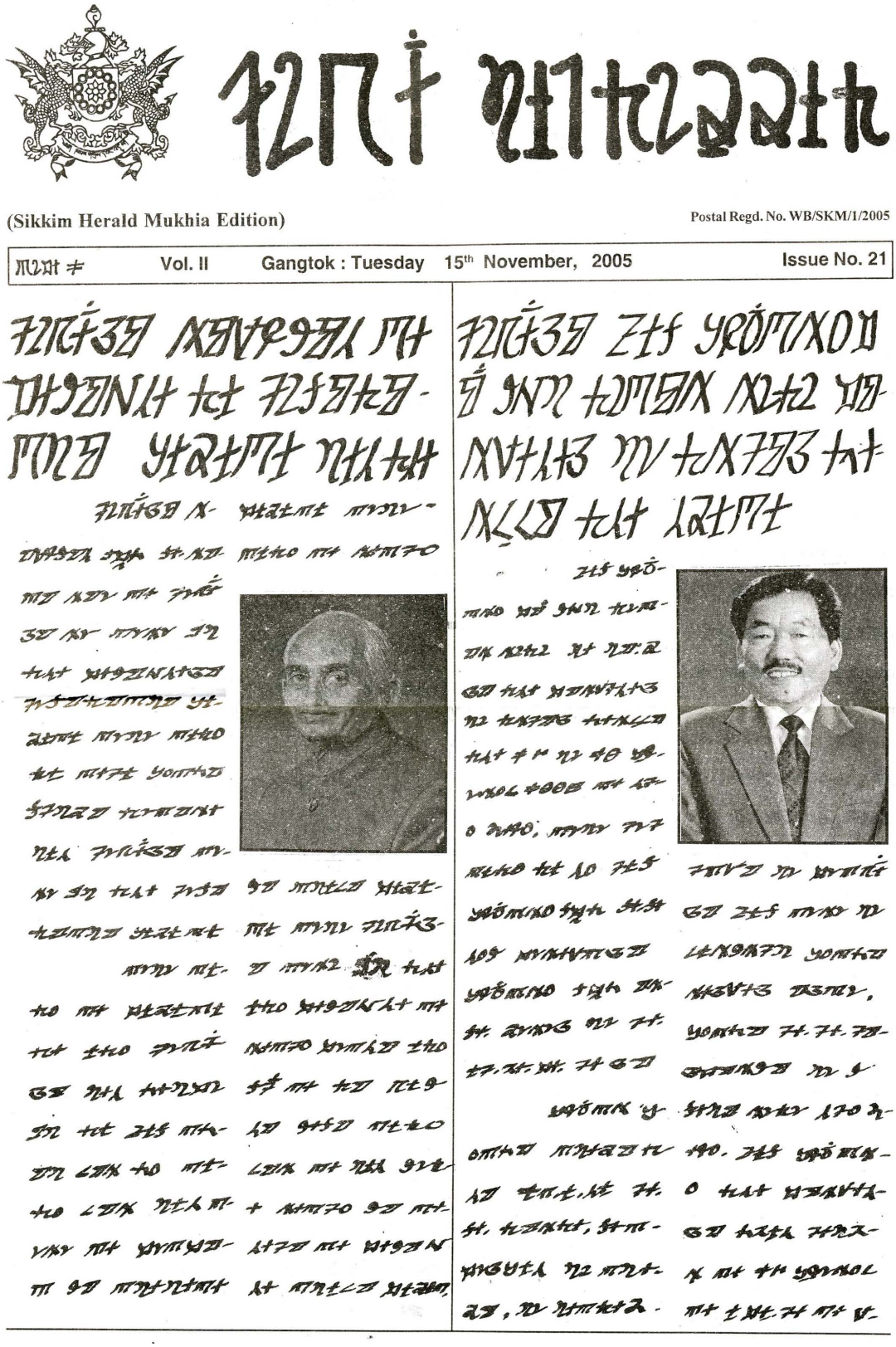


Figure 11: A description of Sunuwar tones written by K. B. Jentich. Image courtesy of Dev Kumar Sunuwar.



(Sikkim Herald Mukhia Edition) Postal Regd. No. WB/SKM/1/2005

Vol. II Gangtok : Tuesday 15th November, 2005 Issue No. 21

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

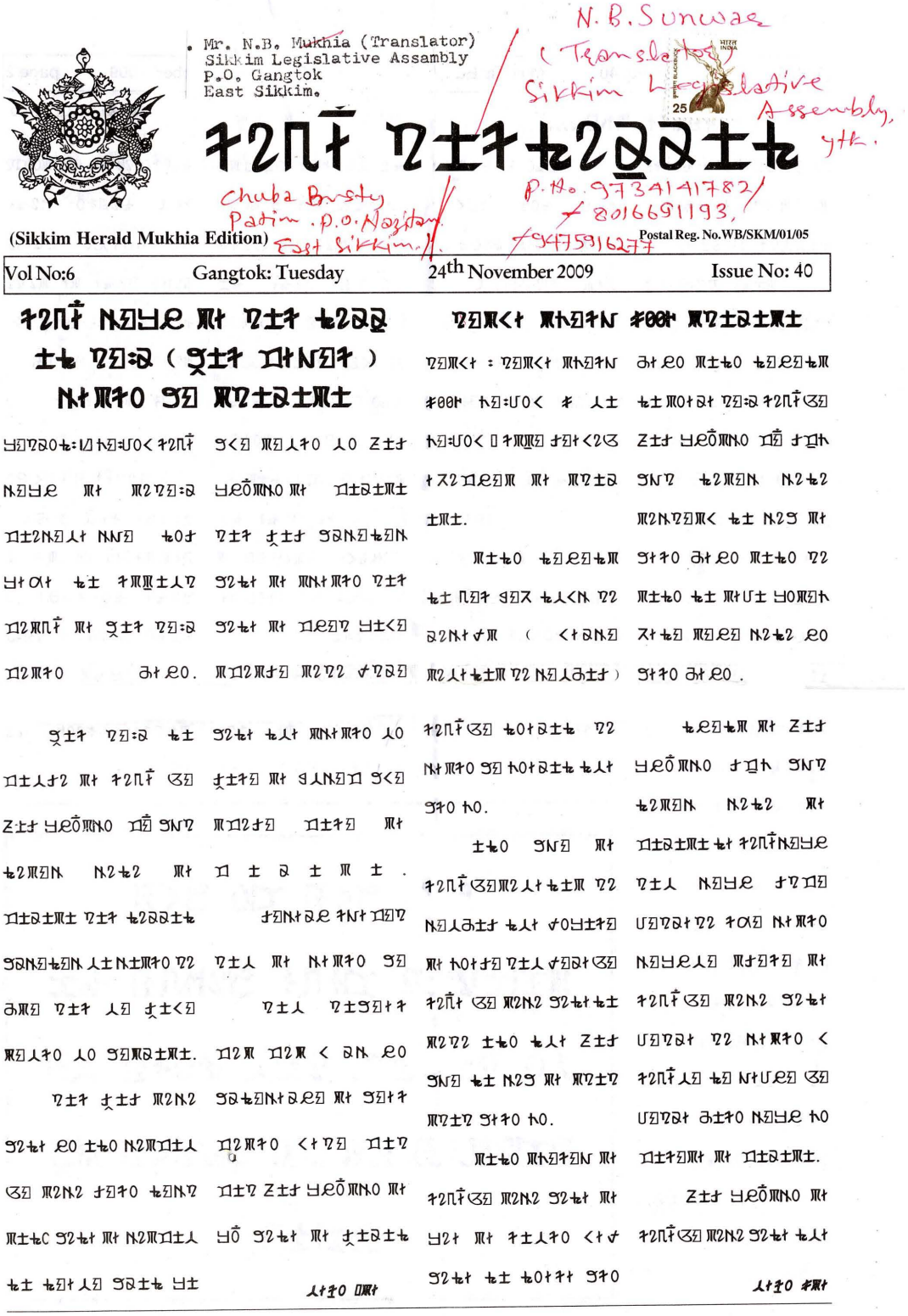


Figure 13: A digitally typeset edition of the *Sikkim Herald*, 2009.

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–कौँइच ब्रे:समा लेखन नमूना ट्रान्स्लिटरेसन सहित

[illegible]
ləlmits məmibə
<ललमिच् ममिबः>

[illegible]
gom tənātən
<गोम् तनातन्>

[illegible]
rāgiā ləlmits rəwākəli
<रागीआ ललमिच् रवाकलि>

[illegible]
rɛum tupšo yo məməi
<रेउम् तुप्शो यो ममइ>

[illegible]
pʰəsim hɛmšoyo məməi
<फश्यीम् हेम्शो यो ममइ>

[illegible]
nām k´išo yo məməi
<नाम की'शो या ममइ>

[illegible]
gom təšonu
<गोम् तशोनु>

[illegible]
mɛko ləlmits rəwā
<मेको ललमिच् रवा>

[illegible]
boisšon bāʔbə
<बोइस्शोन् बाअ्बः>

[illegible]
s´išon bāʔbə
<सी'शोन् बाअ्बः>

[illegible]
kā pāts ā ts´ibpiki
<का पाच्आ ची'बपिकी>

[illegible]
mɛkɛn bɛrnimi
<मेकेन् बेर्नमी>

[illegible]
mɛkɛn nā thɛrnimi
<मेकेन् ना थेर्नमी>

[illegible]
mɛkɛn dz´unimi
<मेकेन् जू'निमी>

[illegible]
minu
<मिनु>

[illegible]
kumšo pəšyā šyil pāinimi
<कुम्शो पश्या श्यील् पाइनिमी>

[illegible]
šyɛŋ mɛko rāgin ā
<श्यङ् मेको रागीन्आ>

[illegible]
wɛk pāts ā
<वेक् पाच्आ>

[illegible]
ts´ib ānkəli
<ची:बआन्कलि>

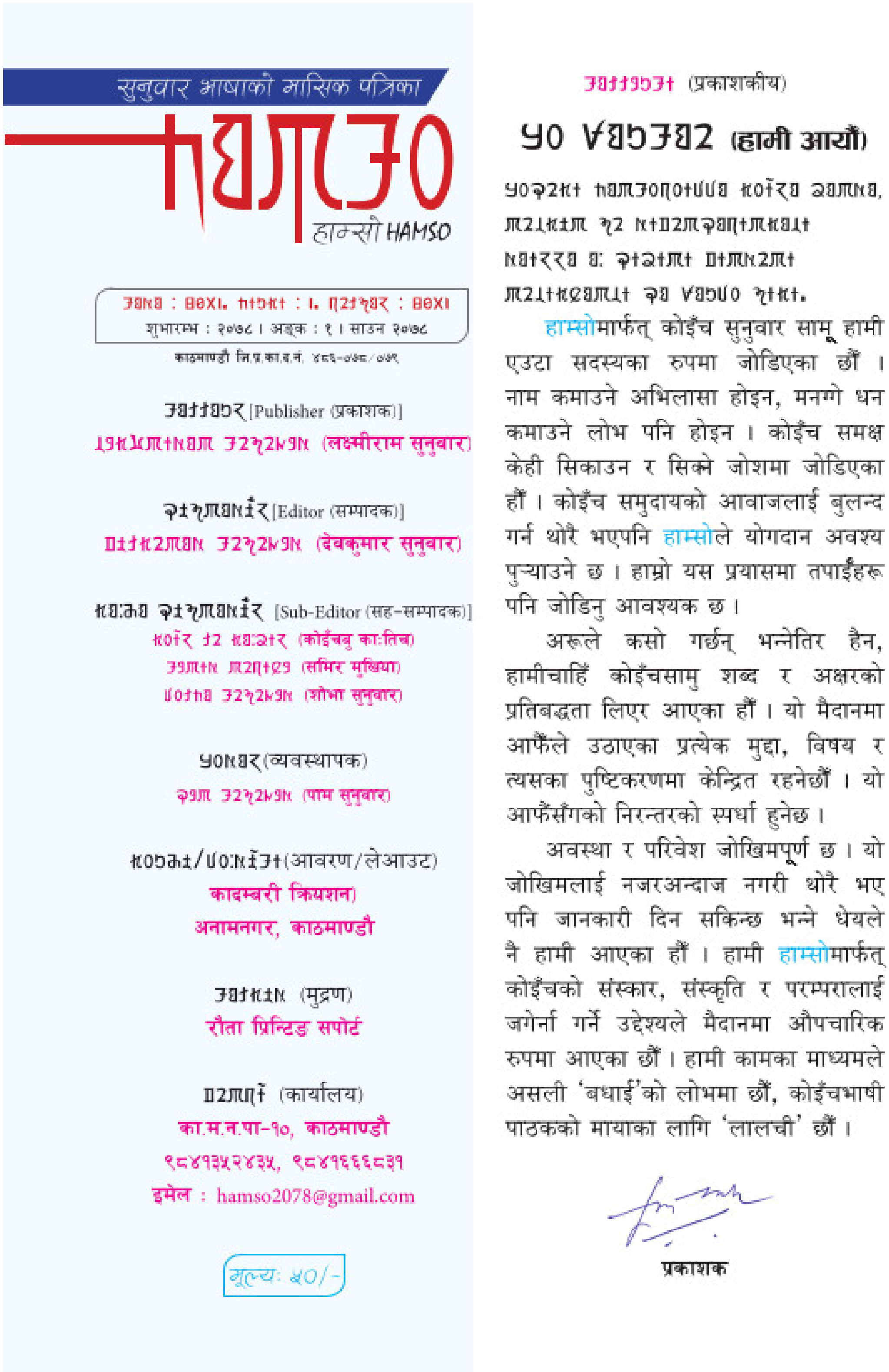
[illegible]
ãkɛ rāgimi gɛnāiyo ləlmits məmibə
<आँके रागीमि गेनाइयो लल्मिच् ममिबः>

(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.



सुनुवार भाषाको मासिक पत्रिका

हाम्सो HAMSO

शुभारम्भ : २०७८ । अङ्क : १ । साउन २०७८

काठमाण्डौ जि.प्र.का.द.नं. ४८६–०७८/०७९

[Publisher (प्रकाशक)]

(लक्ष्मीराम सुनुवार)

[Editor (सम्पादक)]

(देवकुमार सुनुवार)

[Sub-Editor (सह–सम्पादक)]

(कोईंचबु का:तिच)

(समिर मुखिया)

(शोभा सुनुवार)

(व्यवस्थापक)

(पाम सुनुवार)

(आवरण/लेआउट)

कादम्बरी क्रियशन)

अनामनगर, काठमाण्डौ

(मुद्रण)

रौता प्रिन्टिङ सपोर्ट

(कार्यालय)

का.म.न.पा–१०, काठमाण्डौ

९८४१३५२४३५, ९८४१६६६८३१

इमेल : hamso2078@gmail.com

मूल्य: ५०/-

(प्रकाशकीय)

(हामी आयौं)

हाम्सोमार्फत् कोईंच सुनुवार सामु हामी एउटा सदस्यका रुपमा जोडिएका छौं । नाम कमाउने अभिलासा होइन, मनग्गे धन कमाउने लोभ पनि होइन । कोईंच समक्ष केही सिकाउन र सिक्ने जोशमा जोडिएका हौं । कोईंच समुदायको आवाजलाई बुलन्द गर्न थोरै भएपनि हाम्सोले योगदान अवश्य पुऱ्याउने छ । हाम्रो यस प्रयासमा तपाईंहरू पनि जोडिनु आवश्यक छ ।

अरूले कसो गर्छन् भन्नेतिर हैन, हामीचाहिँ कोईंचसामु शब्द र अक्षरको प्रतिबद्धता लिएर आएका हौं । यो मैदानमा आफैंले उठाएका प्रत्येक मुद्दा, विषय र त्यसका पुष्टिकरणमा केन्द्रित रहनेछौं । यो आफैंसँगको निरन्तरको स्पर्धा हुनेछ ।

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

प्रकाशक

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