

Proposal to encode Sirmauri in Unicode

Anshuman Pandey

Biswajit Mandal

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

biswajitmandal.bm90@gmail.com

March 28, 2025

1 Introduction

This proposal to encode the Sirmauri script in the Unicode standard supersedes the following documents:

- L2/18-085: “Preliminary proposal to encode Sirmauri in Unicode” (Pandey)
- L2/25-038: “Proposal to encode Sirmauri in Unicode” (Pandey & Mandal, October 2024)

The following revisions were made after the Script Ad Hoc review of L2/25-038 in November 2024:

- The encoding model no longer supports conjunct formation; *virāma* is rendered visibly
- The joiner named SIRMAURI SIGN VIRAMA, a conjunct control character, has been removed
- The SIRMAURI SIGN HALANTA has been renamed as SIRMAURI SIGN VIRAMA
- Atomic vowel letters retained, instead of using sequences of vowel carriers + dependent vowel signs
- The sign SIRMAURI SIGN RA-KARA has been renamed as SIRMAURI CONSONANT SIGN MEDIAL RA
- Indic positional category details for SIRMAURI CONSONANT SIGN MEDIAL RA corrected to ‘bottom’
- Additional details on current usage and the user community’s desire for using the script digitally

Sirmauri is currently allocated to U+11850..1188F on the Roadmap to the Supplementary Multilingual Plane (SMP). If the proposed block is approved (see code chart following p. 18), it would need another column. The adjacent column at U+11890..1189F is unallocated. If extended, the new allocation would be U+11850..1189F.

2 Proposal History

Sirmauri was first introduced to the Unicode Technical Committee (UTC) in L2/18-085. The aim was to establish the suitability of encoding Sirmauri, with initial evidence drawn from charts and specimens published in *The Linguistic Survey of India* by George A. Grierson (1916b) and *Die Schrift* by Hans Jensen (1969). However, Pandey did not want to pursue the encoding using only secondary sources; and at that time, he was unable to locate experts of the script or primary materials. For this reason, the proposal had been on hold. A few years ago, Mandal independently established contact with Sirmauri users in Himachal Pradesh, who provided manuscripts and feedback on the script. In 2024, Pandey and Mandal decided to merge their efforts in order to present a single, comprehensive proposal for encoding for the script in Unicode.

This proposal presents the form of Sirmauri as attested in actual primary sources, such as manuscripts containing accounting (*huṇḍī*), astrological (*sāñcā*), calendrical (*pañcāṅgam*), and other ritual texts, as well as personal diaries and administrative records. These materials are more representative of conventional Sirmauri than the script shown by Grierson. Moreover, there are prominent differences between the two, as well as inaccuracies in Grierson’s form (see § 5). The representative digital letterforms used in this proposal are based upon normalization of characters used in these primary sources, and were drawn by Pandey and vetted by Mandal. Most importantly, this proposal and the representative glyphs have been reviewed and approved by several current users of Sirmauri, who are acknowledged in § 10.

3 Background

Sirmauri (𑖆𑖇𑖉𑖊𑖋𑖌) is a contemporary Brahmi-based script that is written from left to right. Until the middle of the 20th century, it was a commonly used script in Sirmaur (also spelled ‘Sirmour’ and ‘Sirmur’), a princely state in pre-independence India, which is today split across the present district of Sirmaur in Himachal Pradesh and the Jaunsar-Bawar region of the Garwahl district of Uttarakhand. It was used for writing Sirmauri (ISO 639-3: srx), a Western Pahari language, and a sub-regional variant was used for Jaunsari (ISO 639-3: jns).

The script of Sirmaur is a distinctive ‘Takri’ form related to the Takri (U+11680) of Chamba; the official Dogra (U+11800) of Jammu; and the ancestral Sharada (U+11180) of Kashmir. It also possesses characteristics of mercantile scripts such as Mahajani (U+11150). An origin story for the script relates that when the royal families of Sirmaur formed marital alliances with those of Rajasthan, the latter brought their accountants, clerks, and administrators, who, in turn, brought their native Mahajani script. When this Mahajani was combined with the local form of Takri, a distinctive Sirmauri script emerged.

Sirmauri was used as an official script from at least the 18th century for producing documents such as land grants (see fig. 6–12), revenue records, accounting books, etc., which were published in government gazetteers. It was also used in formal education. According to the *Sirmur State Gazetteer, 1904*, there was a school in Kufarmand of Pachchad, where the Sirmauri and Hindi languages were taught in the Sirmauri and Devanagari scripts, respectively. While usage of Sirmauri for official and educational purposes gave way to Devanagari in the 20th century, it continues to be used cultural and religious practices. The script is intimately connected in Himachal Pradesh with *sāñcā vidyā* — ritual, calendrical, and astrological practices — which is maintained in manuscripts known as *sāñcā grantha* (see fig. 13–17).

Western scholars have known about the Sirmauri script since the 18th century. However, there is very little coverage of Sirmauri in the secondary literature. George A. Grierson described the script and published specimens and charts of it in *The Linguistic Survey of India*, vol. IX (1916b); shown here in figs. 1–4. The letters of the Jaunsari variety appears in a chart in *Die Schrift* (1969) by Hans Jensen (fig. 5), but, those forms appear to be derived from those in Grierson.

Today, Sirmauri is an endangered, but living script. It is read and written by members of the community who have maintained knowledge of it through their cultural practices, personal endeavor, and the traditions of their families. While users continue to write the script, they also want to be able to type it on digital devices. They have begun to use digital fonts to create and exchange documents in their script (see fig. 39 for a poem by Sherjung Chauhan). They also express a strong desire to digitize their manuscript collections and to be able to represent the contents of these documents in digital plain text. Encoding Sirmauri in Unicode will enable its users to preserve, interchange, and create content in the script, and to expand its usage on digital platforms on par with other long-lived and currently used minority Indic scripts.

4 The Script

4.1 Vowel letters

There are 4 basic vowel letters in Sirmauri:

३)	६	६	ॡ
<i>a</i>	<i>i, ī</i>	<i>u, ū</i>	<i>e</i>

Generally, the letter ६ represents both *i* and *ī*; and ६ represents both *u* and *ū*. But, vowel length may also be distinctively indicated, as described below, by combining dependent vowel signs with these base letters. The letters *a*, *i*, and *u* also function as vowel carriers for representing independent vowels.

4.2 Vowel signs

The basic letters are accompanied by 7 dependent vowel signs:

◌̄)	◌̄)	◌̄)	◌̄)	◌̄)	◌̄)	◌̄)
<i>-ā</i>	<i>-ī, -ī</i>	<i>-u, -ū</i>	<i>-e</i>	<i>-ai</i>	<i>-o</i>	<i>-au</i>

Like other Indic scripts, there is no dependent sign for *-a*, the vowel inherent in each consonant letter. Length distinction between non-initial *-ī, -ī* and *-u, -ū* is not maintained in the orthography.

4.3 Representation of independent long vowels and diphthongs

4.3.1 *ā, ai, o, au*; and alternate *e*

Independent forms of *ā, ai, o, au*, and an alternate *e* are represented by combining the respective vowel sign with the letter *a*.

३)	३)	३)	३)	३)
३), ◌̄)	३), ◌̄)	३), ◌̄)	३), ◌̄)	३), ◌̄)
<i>a, -ā</i>	<i>a, -ai</i>	<i>a, -o</i>	<i>a, -au</i>	<i>a, -e</i>
<i>ā</i>	<i>ai</i>	<i>o</i>	<i>au</i>	<i>e</i>

Both the independent letter ॡ and the vowel-carrier form ३) are commonly used for writing *e* (see fig. 19, 20).

4.3.2 \bar{i} , \bar{u}

Independent forms of \bar{i} and \bar{u} are written using vowel letters and combining vowel signs (see fig. 19, 20).

$i, -u$	$i, -\bar{i}$	$u, -u$
\bar{i}	\bar{i}	\bar{u}

The variant form of \bar{i} is attested in one document present in the available materials (see fig. 35). The usage of two dots to distinguish i and \bar{i} is a common pattern in related scripts. It is also quite possible that the conventional sign may have evolved from writing the two dots as a single swash.

	i	\bar{i}
Sharada		
Dogra		
Takri		

The usage of the sign - i for writing is also an orthographic convention in related scripts, eg. Khojki \bar{i} , which is palaeographically a combination of i and the sign - \bar{i} .

4.4 Consonant letters

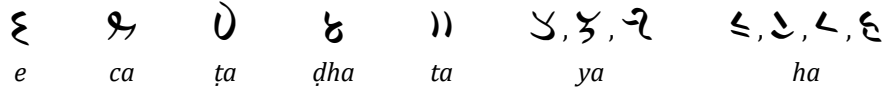
There are 28 consonant letters, arranged according to the typical Brahmi paradigm. The script lacks distinctive letters for representing the nasals $\bar{n}a$ and $\bar{\tilde{n}}a$, the sibilants $\acute{s}a$ and $\acute{s}a$, and the semi-vowel va .

				—					—
ka	kha	ga	gha	$\bar{n}a$	ca	cha	ja	jha	$\bar{\tilde{n}}a$
$\acute{t}a$	$\acute{t}ha$	$\acute{d}a$	$\acute{d}ha$	$\bar{n}a$	ta	tha	da	dha	na
								—	
pa	pha	ba	bha	ma	ya	ra	la	va	
—	—								
$\acute{s}a$	$\acute{s}a$	sa	ha						

Other consonant sounds are represented by combining a distinctive letter with the sign *nukta* (see § 4.6).

4.5 Variants

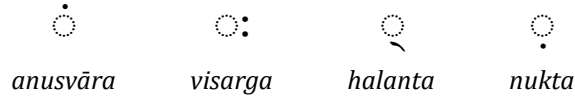
Naturally, there is variation in shapes of letters in hand-written documents, but distinctive variants are also attested (see fig. 32–33, 38):



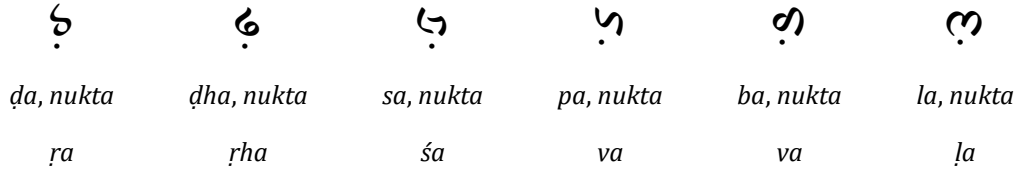
Additionally, some documents show Devanagari-like forms for letters, such as *tha*, *pa*, *pha* (see fig. 34–35). Usage of these is likely due to contact with Devanagari, and the fluency of the writer with both scripts.

4.6 Modifier signs

The following modifier signs are used:



- *anusvāra* indicates vowel nasalization or is a proxy for a nasal consonant, eg. $\dot{\text{ṣ}}$ *kaṁ*
- *visarga* indicates post-vocalic aspiration and occurs in Sanskrit contexts, eg. ṇṇ : *namaḥ* (see fig. 37)
- *halanta* silences the inherent vowel *a* of a consonant letter, eg. ṣ *k* (see fig. 21, 22), ṣṣ *kka*
- *nukta* is used for extending the script to represent sounds for which distinctive letters do not exist:



4.7 Consonant clusters

Consonant clusters are conventionally represented by writing the sign *halanta* (*virāma*) under the consonant whose inherent vowel is silenced, eg. ṣ *k*. There are no atomic conjuncts. Although rare, usage of half or contextual forms of letters is attested:



- The ṣ half *ma* occurs in the cluster ṣṣ *mma* in a single document (see fig. 34). This conjunct representation of a cluster should be considered an idiosyncrasy, and not a typical convention of Sirmauri.
- The *ra-kāra* occurs in a few documents (see fig. 38). It is a contextual form of ṛ *ra* used in a cluster-medial position, eg. ṣ *kra*, ṛ *pra*. The usage of this form indicates a strong orthographic influence from Devanagari. Other contextual forms of *ra*, such as the cluster-initial *repha* are not attested.

4.8 Digits

There is a full set of decimal digits (see fig. 24, 25):

୦	୧	୨	୩	୪	୫	୬	୭	୮	୯
0	1	2	3	4	5	6	7	8	9

4.9 Auspicious sign

An auspicious sign known as *ekam* is used in Sirmauri (see fig. 26, 36), which is related to signs known as *añjī* in other northern and eastern Indic scripts:

୧
ekam

4.10 Punctuation

Several signs of punctuation are used (see examples in figs. 27–30, 36):

))	:	::
<i>daṇḍa</i>	double <i>daṇḍa</i>	section mark	double section mark	two dots	four dots

One document shows rare usage of a Latin sign of punctuation, the question mark ‘?’ (see fig. 31).

Sirmauri documents often show a ‘headline’. This is not a feature of the script, but of the writing surface. Ruled paper was often used, or lines were written on paper before starting with the text. Additional lines would be drawn as needed as the page is filled (see fig. 6 onwards).

5 Comparison with Grierson's form of Sirmauri

The following is a comparison of forms of Sirmauri and Jaunsari letters shown by Grierson (top) and digitized forms of letters used in the primary materials examined by Pandey and Mandal (below):

3	3), 3))	6, ५	—	3	—	3, ५, ५	3̄, 6	3̄)	3̄̄)
3)	3))	6, ५,	6, 6, 62	3	3̄	५, 3̄)	3̄̄)	3̄̄)	3̄̄̄)
<i>a</i>	<i>ā</i>	<i>i, ī</i>	<i>ī</i>	<i>u, ū</i>	<i>ū</i>	<i>e</i>	<i>ai</i>	<i>o</i>	<i>au</i>
x	५	3)	५	—	५	५	५	५	—
५	५	3)	५	—	3	५	५	५	—
<i>ka</i>	<i>kha</i>	<i>ga</i>	<i>gha</i>	<i>ṇa</i>	<i>ca</i>	<i>cha</i>	<i>ja</i>	<i>jha</i>	<i>ṇa</i>
५	6	5	५)))	५	५	ε	ε	५
५	५	५	6)))	५	५	ε	५	५
<i>ṭa</i>	<i>ṭha</i>	<i>ḍa</i>	<i>ḍha</i>	<i>ṇa</i>	<i>ta</i>	<i>tha</i>	<i>da</i>	<i>dha</i>	<i>na</i>
५	५	५	५	५	५	५	५	५	५
५	५	५	५	५	५	५	५	(५)	५
<i>pa</i>	<i>pha</i>	<i>ba</i>	<i>bha</i>	<i>ma</i>	<i>ya</i>	<i>ra</i>	<i>la</i>	<i>va</i>	
५	५	५	ε						
(५)	(५)	५	५						
<i>śa</i>	<i>ṣa</i>	<i>sa</i>	<i>ha</i>						

The letterforms in the two sets of sources have several similarities, but also some prominent differences, which are described below:

- *a* Grierson's 3 *a* is mysterious. It does not occur in any of the source materials reviewed by the proposal authors, nor is it known to experts of the script, who report that they have never observed it in any Sirmauri materials, especially as a form for *a*. In Grierson, this 3 form is the basis for 3) *ā*, 3̄ *e*, 3̄̄ *ai*, 3̄̄̄ *o*, 3̄̄̄̄ *au*. But, despite the consistency of this form across the independent vowel letters, the source for 3 is unknown. It could have been an idiosyncrasy of the scribe of Grierson's text. Experts confirm that 3) is the correct form for *a*. Accordingly, Grierson's forms would be analyzed as follows: 3) *ā* = *a*; 3̄ *e* and 3̄̄ *ai* are invalid forms; 3̄̄̄ *o* = *e*; 3̄̄̄̄ *au* = *ai*. Nonetheless, his recording of the Jaunsari form 3)) as *ā* is correct, as this form has the appropriate 3) *a* as the base combined with the conventional vowel sign 3̄) for *ā*.

- *ā* The forms ३) and ३)) are given for Jaunsari *ā*. Of these, ३)) is the correct form for *ā*. It is possible that ३) *a* was used for *ā* in some contexts, just as ६ is used for both *i* and *ī*, and ॢ for *u* and *ū*.
- *i, ī* The ६ is the paleographical form for *i*, also used for *ī*. The ॡ is the letter *e*, which along with ॡya, is used as a proxy for *i*.
- *e* The source for ॢ is unknown. It is *a*, functioning as a vowel carrier, with the vowel sign *e*. Similarly, ॡ is ॡya bearing the vowel sign *e*. The ॡ is the letter for *e*. The correct vowel carrier form of *e* is ॢ̣.
- *ai* The Jaunsari ६ *ai* is the letter ६ *i* written with the vowel sign *e*. Apart from Grierson, this form is not attested in other materials.
- *o, au* The source for the forms ॢ̣ *o* and ॢ̣̣ *au* is unknown. Apart from Grierson, these forms are not attested in other materials.
- *ṭa* The ॡ is a form heavily influenced by Devanagari *ṭa*. It is more commonly used in Sirmauri for *ṭha*, while the representative form for *ṭa* is ॡ̣.
- *bha* The ॡ̣ is a reversed form of conventional ॡ̣ *bha*.
- *va* The ॡ̣̣ is not a distinct letter, but the letter *ba* written with a *nukta*.
- *śa* The ॡ̣̣̣ is not a distinct letter, but the letter *sa* written with a *nukta*.
- *ṣa* The ॡ̣̣̣̣ shown for *ṣa* is simply a duplication of *kha*, which is commonly used for writing the sibilant.

6 Principles of the Encoding

The Unicode encoding for Sirmauri is based upon the following principles:

6.1 Scope

The ‘Sirmauri’ block encodes the script known as ‘Sirmauri’. The variety known as ‘Jaunsari’ is to be unified with Sirmauri.

6.2 Encoding model

Although Sirmauri is a Brahmi script, it does not use consonant consonants. Therefore, the typical Indic *virāma* model is not proposed for it.

- *virāma* The sign 𑂏 is written under a consonant to indicate that the inherent vowel is silenced. It is called *halanta* in Sirmauri. In L2/25-038, the 𑂏 was proposed for encoding as SIRMAURI SIGN HALANTA along with the joiner SIRMAURI SIGN VIRAMA, a control character to be used for conjunct formation. The names were chosen to distinguish between vowel-silencer (HALANTA) and joiner (VIRAMA) for encoding purposes. But, as it was determined that a *virāma* model is not necessary for Sirmauri, the joiner has been removed. As a result, the former SIRMAURI SIGN HALANTA has been renamed to SIRMAURI SIGN VIRAMA, following the model of other Indic scripts. The SIRMAURI SIGN VIRAMA does not have any control properties for conjunct formation, and is always rendered visibly.

- *Representation of consonant clusters* The following methods are proposed for representing clusters:

- Usage of *virāma*: 𑂏𑂐 *mma* would be represented as:

<𑂐 SIRMAURI LETTER MA , 𑂏 SIRMAURI SIGN VIRAMA , 𑂐 SIRMAURI LETTER MA>

- Normalization of conjuncts: Half-forms that may appear in Sirmauri texts should be normalized using *virāma*; for example, the singular occurrence of 𑂏𑂐 *mma* would be normalized as 𑂏𑂐, using the visible *virāma* as shown above.
- The contextual form 𑂏 *ra-kāra* of 𑂐 *ra* is proposed for encoding as a separate character: SIRMAURI CONSONANT SIGN MEDIAL RA. This reduces the need for a *virāma* for representing clusters. The cluster *kra* would be represented as:

<𑂏 SIRMAURI LETTER KA , 𑂏 SIRMAURI CONSONANT SIGN MEDIAL RA>

6.3 Consonant letters

All distinctively attested consonant letters have been proposed for encoding as independent characters.

6.4 Representation of vowel letters

There are two approaches for supporting the variety of independent vowel letters used in the script. The second approach using atomic characters has been selected.

1. *Sequence of base vowel letter + dependent vowel sign* Encode only the distinctive vowel letters and define them as vowel carriers:

<i>a</i>	३)	SIRMAURI LETTER A
<i>i</i>	७	SIRMAURI LETTER I
<i>u</i>	७	SIRMAURI LETTER U
<i>e</i>	५	SIRMAURI LETTER E

These base vowel letters would be used in combination with dependent vowel signs to represent the full set of independent vowel letters:

<i>ā</i>	३))	३) SIRMAURI LETTER A , ०) SIRMAURI VOWEL SIGN AA
<i>e</i>	३)	३) SIRMAURI LETTER A , ०̣ SIRMAURI VOWEL SIGN E
<i>ai</i>	३̣)	३) SIRMAURI LETTER A , ०̣ SIRMAURI VOWEL SIGN AI
<i>o</i>	३))	३) SIRMAURI LETTER A , ०) SIRMAURI VOWEL SIGN O
<i>au</i>	३̣))	३) SIRMAURI LETTER A , ०̣ SIRMAURI VOWEL SIGN AU

This approach would facilitate the representation of other independent vowel letters without encoding them as separate characters:

<i>ī</i>	७	७ SIRMAURI LETTER I , ० SIRMAURI VOWEL SIGN U
<i>ī</i>	७̣	७ SIRMAURI LETTER I , ०̣ SIRMAURI VOWEL SIGN II
<i>ū</i>	७	७ SIRMAURI LETTER U , ० SIRMAURI VOWEL SIGN U

2. *Atomic characters* Encode all conventional independent vowel letters as atomic characters, following the pattern of other northern Indic scripts:

<i>a</i>	३)	SIRMAURI LETTER A
<i>ā</i>	३))	SIRMAURI LETTER AA
<i>i</i>	७	SIRMAURI LETTER I
<i>ī</i>	७	SIRMAURI LETTER II
<i>u</i>	७	SIRMAURI LETTER U
<i>ū</i>	७	SIRMAURI LETTER UU
<i>e</i>	५	SIRMAURI LETTER E
<i>e</i>	३)	SIRMAURI LETTER ALTERNATE E
<i>ai</i>	३̣)	SIRMAURI LETTER AI
<i>o</i>	३))	SIRMAURI LETTER O
<i>au</i>	३̣))	SIRMAURI LETTER AU

Accordingly, the following would be treated as glyphic variants, whose representation would be controlled using fonts. If evidence of their concurrent usage with the normative form is found, then they may be encoded as alternate characters.

\bar{i} ७ᳵ glyph variant of ७ SIRMAURI LETTER II
 \bar{i} ७ glyph variant of ७ SIRMAURI LETTER II

6.5 Character name for vowel sign for -i, -ī

The ७ᳵ is used for both -i and -ī, but it is palaeographically- \bar{i} . For this reason, it has been assigned the name SIRMAURI VOWEL SIGN II.

6.6 Reservation of codepoints in the block

Space has been reserved for the following characters, in the event that distinctive forms are identified in the future:

- *Vowel signs* The reservation of space for *SIRMAURI VOWEL SIGN I and *SIRMAURI VOWEL SIGN UU is driven by a recent update for Khojki, which also has a merger of *i*, \bar{i} and *u*, \bar{u} . The Khojki ३ᳵ is palaeographically \bar{i} , but it was encoded as U+11202 KHOJKI LETTER I instead of *KHOJKI LETTER II (see L2/11-021). A few years ago, a distinctive independent form ३ for *i* was identified. But as ३ᳵ was already named KHOJKI LETTER I, the ३ was encoded as U+11240 KHOJKI LETTER SHORT I (see L2/21-104). Using character names for Sirmauri based on palaeography avoids such complications.
- *Consonants* The following letters have reserved codepoints in their usual positions in the consonant order: *SIRMAURI LETTER NGA, *SIRMAURI LETTER NYA, *SIRMAURI LETTER SHA, *SIRMAURI LETTER SSA.

6.7 Handling of variant forms

Distinctive variants of consonant letters, such as those shown in fig. 32–33 are to be considered glyphic variants and their display is to be managed using fonts.

6.8 Digits

A set of script-specific digits for Sirmauri is proposed for encoding.

6.9 Auspicious sign

The auspicious sign *ekam* is proposed for encoding as an independent character in the Sirmauri block

6.10 Punctuation

Support for punctuation signs used in Sirmauri documents is as follows:

- *daṇḍā-s* The | *daṇḍā* and || double *daṇḍā* are to be unified with the corresponding characters in the Devanagari block. Script extensions have been specified.

- I U+0964 DEVANAGARI DANDA
- II U+0964 DEVANAGARI DOUBLE DANDA

- *Section marks* The distinctive section marks are proposed for encoding as the following characters:

- 𑂏 SIRMAURI SECTION MARK
- 𑂐 SIRMAURI DOUBLE SECTION MARK

- *Dotted marks* The commonly used : and :: punctuation signs are to be represented using one or two instances of the regular : U+003A COLON.

7 Proposed Character Repertoire

The proposed Unicode repertoire for Sirmauri consists of 64 characters:

Category	Character	Character Name
Vowel letters (11)	𑂑	SIRMAURI LETTER A
	𑂒	SIRMAURI LETTER AA
	𑂓	SIRMAURI LETTER I
	𑂔	SIRMAURI LETTER II
	𑂕	SIRMAURI LETTER U
	𑂖	SIRMAURI LETTER UU
	𑂗	SIRMAURI LETTER E
	𑂘	SIRMAURI LETTER ALTERNATE E
	𑂙	SIRMAURI LETTER AI
	𑂚	SIRMAURI LETTER O
	𑂛	SIRMAURI LETTER AU
Dependent vowel signs (7)	𑂜	SIRMAURI VOWEL SIGN AA
	𑂝	SIRMAURI VOWEL SIGN II
	𑂞	SIRMAURI VOWEL SIGN U
	𑂟	SIRMAURI VOWEL SIGN E
	𑂠	SIRMAURI VOWEL SIGN AI
	𑂡	SIRMAURI VOWEL SIGN O
	𑂢	SIRMAURI VOWEL SIGN AU
Consonant letters (28)	𑂣	SIRMAURI LETTER KA

	ᳵ	SIRMAURI LETTER KHA
	ᳶ	SIRMAURI LETTER GA
	᳷	SIRMAURI LETTER GHA
	᳸	SIRMAURI LETTER CA
	᳹	SIRMAURI LETTER CHA
	ᳺ	SIRMAURI LETTER JA
	᳻	SIRMAURI LETTER JHA
	᳼	SIRMAURI LETTER TTA
	᳽	SIRMAURI LETTER TTHA
	᳾	SIRMAURI LETTER DDA
	᳿	SIRMAURI LETTER DDHA
	᳠	SIRMAURI LETTER NNA
	᳡	SIRMAURI LETTER TA
	᳢	SIRMAURI LETTER THA
	᳣	SIRMAURI LETTER DA
	᳤	SIRMAURI LETTER DHA
	᳥	SIRMAURI LETTER NA
	᳦	SIRMAURI LETTER PA
	᳧	SIRMAURI LETTER PHA
	᳨	SIRMAURI LETTER BA
	ᳩ	SIRMAURI LETTER BHA
	ᳪ	SIRMAURI LETTER MA
	ᳫ	SIRMAURI LETTER YA
	ᳬ	SIRMAURI LETTER RA
	᳭	SIRMAURI LETTER LA
	ᳮ	SIRMAURI LETTER SA
	ᳯ	SIRMAURI LETTER HA
Various signs (4)	ᳰ	SIRMAURI SIGN ANUSVARA
	ᳱ	SIRMAURI SIGN VISARGA
	ᳲ	SIRMAURI SIGN VIRAMA
	ᳳ	SIRMAURI SIGN NUKTA

Consonant sign (1)	◌̣	SIRMAURI CONSONANT SIGN MEDIAL RA
Auspicious sign (1)	◌̤	SIRMAURI EKAM
Punctuation (2)	◌̥	SIRMAURI SECTION MARK
	◌̦	SIRMAURI DOUBLE SECTION MARK
Digits (10)	◌̧	SIRMAURI DIGIT ZERO
	◌̨	SIRMAURI DIGIT ONE
	◌̩	SIRMAURI DIGIT TWO
	◌̪	SIRMAURI DIGIT THREE
	◌̫	SIRMAURI DIGIT FOUR
	◌̬	SIRMAURI DIGIT FIVE
	◌̭	SIRMAURI DIGIT SIX
	◌̮	SIRMAURI DIGIT SEVEN
	◌̯	SIRMAURI DIGIT EIGHT
	◌̰	SIRMAURI DIGIT NINE

8 Character Data

Character Properties: UnicodeData.txt

```

11850;SIRMAURI LETTER A;Lo;0;L;;;;N;;;;;
11851;SIRMAURI LETTER AA;Lo;0;L;;;;N;;;;;
11852;SIRMAURI LETTER I;Lo;0;L;;;;N;;;;;
11853;SIRMAURI LETTER II;Lo;0;L;;;;N;;;;;
11855;SIRMAURI LETTER U;Lo;0;L;;;;N;;;;;
11855;SIRMAURI LETTER UU;Lo;0;L;;;;N;;;;;
11856;SIRMAURI LETTER E;Lo;0;L;;;;N;;;;;
11858;SIRMAURI LETTER ALTERNATE E;Lo;0;L;;;;N;;;;;
11857;SIRMAURI LETTER AI;Lo;0;L;;;;N;;;;;
11859;SIRMAURI LETTER O;Lo;0;L;;;;N;;;;;
1185A;SIRMAURI LETTER AU;Lo;0;L;;;;N;;;;;
1185B;SIRMAURI LETTER KA;Lo;0;L;;;;N;;;;;
1185C;SIRMAURI LETTER KHA;Lo;0;L;;;;N;;;;;
1185D;SIRMAURI LETTER GA;Lo;0;L;;;;N;;;;;
1185E;SIRMAURI LETTER GHA;Lo;0;L;;;;N;;;;;
1185F;<reserved>
11860;SIRMAURI LETTER CA;Lo;0;L;;;;N;;;;;
11861;SIRMAURI LETTER CHA;Lo;0;L;;;;N;;;;;
11862;SIRMAURI LETTER JA;Lo;0;L;;;;N;;;;;

```

```

11863;SIRMAURI LETTER JHA;Lo;0;L;;;;N;;;;;
11864;<reserved>
11865;SIRMAURI LETTER TTA;Lo;0;L;;;;N;;;;;
11866;SIRMAURI LETTER TTHA;Lo;0;L;;;;N;;;;;
11867;SIRMAURI LETTER DDA;Lo;0;L;;;;N;;;;;
11868;SIRMAURI LETTER DDHA;Lo;0;L;;;;N;;;;;
11869;SIRMAURI LETTER NNA;Lo;0;L;;;;N;;;;;
1186A;SIRMAURI LETTER TA;Lo;0;L;;;;N;;;;;
1186B;SIRMAURI LETTER THA;Lo;0;L;;;;N;;;;;
1186C;SIRMAURI LETTER DA;Lo;0;L;;;;N;;;;;
1186D;SIRMAURI LETTER DHA;Lo;0;L;;;;N;;;;;
1186E;SIRMAURI LETTER NA;Lo;0;L;;;;N;;;;;
1186F;SIRMAURI LETTER PA;Lo;0;L;;;;N;;;;;
11870;SIRMAURI LETTER PHA;Lo;0;L;;;;N;;;;;
11871;SIRMAURI LETTER BA;Lo;0;L;;;;N;;;;;
11872;SIRMAURI LETTER BHA;Lo;0;L;;;;N;;;;;
11873;SIRMAURI LETTER MA;Lo;0;L;;;;N;;;;;
11874;SIRMAURI LETTER YA;Lo;0;L;;;;N;;;;;
11875;SIRMAURI LETTER RA;Lo;0;L;;;;N;;;;;
11876;SIRMAURI LETTER LA;Lo;0;L;;;;N;;;;;
11877;<reserved>
11878;<reserved>
11879;<reserved>
1187A;SIRMAURI LETTER SA;Lo;0;L;;;;N;;;;;
1187B;SIRMAURI LETTER HA;Lo;0;L;;;;N;;;;;
1187C;SIRMAURI VOWEL SIGN AA;Mc;0;L;;;;N;;;;;
1187D;<reserved>
1187E;SIRMAURI VOWEL SIGN II;Mc;0;L;;;;N;;;;;
1187F;SIRMAURI VOWEL SIGN U;Mn;0;NSM;;;;N;;;;;
11880;<reserved>
11881;SIRMAURI VOWEL SIGN E;Mn;0;NSM;;;;N;;;;;
11882;SIRMAURI VOWEL SIGN AI;Mn;0;NSM;;;;N;;;;;
11883;SIRMAURI VOWEL SIGN O;Mc;0;L;1187C 11881;;;;N;;;;;
11884;SIRMAURI VOWEL SIGN AU;Mc;0;L;1187C 11882;;;;N;;;;;
11885;SIRMAURI SIGN ANUSVARA;Mn;0;NSM;;;;N;;;;;
11886;SIRMAURI SIGN VISARGA;Mc;0;L;;;;N;;;;;
11887;SIRMAURI SIGN VIRAMA;Mn;9;NSM;;;;N;;;;;
11888;SIRMAURI SIGN NUKTA;Mn;0;NSM;;;;N;;;;;
11889;SIRMAURI CONSONANT SIGN MEDIAL RA;Mn;0;NSM;;;;N;;;;;
1188A;SIRMAURI EKAM;So;0;L;;;;N;;;;;
1188B;SIRMAURI SECTION MARK;Po;0;L;;;;N;;;;;
1188C;SIRMAURI DOUBLE SECTION MARK;Po;0;L;;;;N;;;;;
11890;SIRMAURI DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
11891;SIRMAURI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
11892;SIRMAURI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;8
11893;SIRMAURI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
11894;SIRMAURI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
11895;SIRMAURI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
11896;SIRMAURI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
11897;SIRMAURI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;

```

```
11898;SIRMAURI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
11899;SIRMAURI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
```

8.1 Linebreaking Properties: LineBreak.txt

```
11850..1185E ; AL # Lo [4] SIRMAURI LETTER A..SIRMAURI LETTER GH
11860..11863 ; AL # Lo [4] SIRMAURI LETTER CA..SIRMAURI LETTER JH
11865..11876 ; AL # Lo [18] SIRMAURI LETTER TTA..SIRMAURI LETTER LA
1187A..1187B ; AL # Lo [2] SIRMAURI LETTER SA..SIRMAURI LETTER HA
1187C ; CM # Mc SIRMAURI VOWEL SIGN AA
1187E ; CM # Mc SIRMAURI VOWEL SIGN II
1187F ; CM # Mn SIRMAURI VOWEL SIGN U
11881..11882 ; CM # Mn [2] SIRMAURI VOWEL SIGN E..SIRMAURI VOWEL SIGN AI
11883..11884 ; CM # Mc [2] SIRMAURI VOWEL SIGN O..SIRMAURI VOWEL SIGN AU
11885 ; CM # Mn SIRMAURI SIGN ANUSVARA
11886 ; CM # Mc SIRMAURI SIGN VISARGA
11887..11889 ; CM # Mn [2] SIRMAURI SIGN VIRAMA..SIRMAURI CONSONANT SIGN MEDIAL RA
1188A ; AL # Lo SIRMAURI EKAM
1188B..1188C ; BA # Po [2] SIRMAURI SECTION MARK..SIRMAURI DOUBLE SECTION MARK
11890..11899 ; NU # Nd [10] SIRMAURI DIGIT ZERO..SIRMAURI DIGIT NINE
```

8.2 Syllabic Categories: IndicSyllabicCategory.txt

```
# Indic_Syllabic_Category=Bindu
11885 ; Bindu # Mc SIRMAURI SIGN ANUSVARA

# Indic_Syllabic_Category=Visarga
11886 ; Visarga # Mc SIRMAURI SIGN VISARGA

# Indic_Syllabic_Category=Pure_Killer
11887 ; Pure_Killer # Mn SIRMAURI SIGN VIRAMA

# Indic_Syllabic_Category=Vowel_Independent
11850..1185A ; Vowel_Independent # Lo [11] SIRMAURI LETTER A..SIRMAURI LETTER AU

# Indic_Syllabic_Category=Vowel_Dependent
1187C ; Vowel_Dependent # Mc SIRMAURI VOWEL SIGN AA
1187E ; Vowel_Dependent # Mc SIRMAURI VOWEL SIGN II
1187F ; Vowel_Dependent # Mc SIRMAURI VOWEL SIGN U
11881..11884 ; Vowel_Dependent # Mn [4] SIRMAURI VOWEL SIGN E..SIRMAURI VOWEL SIGN AU

# Indic_Syllabic_Category=Consonant
1185B..1185E ; Consonant # Lo [4] SIRMAURI LETTER KA..SIRMAURI LETTER GH
11860..11863 ; Consonant # Lo [4] SIRMAURI LETTER CA..SIRMAURI LETTER JH
11865..11876 ; Consonant # Lo [18] SIRMAURI LETTER TTA..SIRMAURI LETTER LA
1187A..1187B ; Consonant # Lo [2] SIRMAURI LETTER SA..SIRMAURI LETTER HA

# Indic_Syllabic_Category=Consonant_Medial
```

11889 ; Consonant_Medial # Mn SIRMAURI CONSONANT SIGN MEDIAL RA

8.3 Positional Categories: IndicPositionalCategory.txt

```
# Indic_Positional_Category=Top
11885 ; Top # Mn SIRMAURI SIGN ANUSVARA
11881..11882 ; Top # Mn [2] SIRMAURI VOWEL SIGN E..SIRMAURI VOWEL SIGN AI

# Indic_Positional_Category=Bottom
1187F ; Bottom # Mn SIRMAURI VOWEL SIGN U ..
11887..11888 ; Bottom # Mn [2] SIRMAURI SIGN VIRAMA..SIRMAURI SIGN NUKTA
11889 ; Bottom # Mn SIRMAURI CONSONANT SIGN MEDIAL RA

# Indic_Positional_Category=Right
1187C ; Right # Mc SIRMAURI VOWEL SIGN AA
1187E ; Right # Mc SIRMAURI VOWEL SIGN II
11886 ; Right # Mc SIRMAURI SIGN VISARGA

# Indic_Positional_Category=Top_And_Right
11883..11884 ; Top_And_Right # Mc [2] SIRMAURI VOWEL SIGN O..SIRMAURI VOWEL SIGN AU
```

8.4 Script Extensions: ScriptExtensions.txt

```
0964 ; ... # Po DEVANGARI DANDA
0965 ; ... # Po DEVANGARI DOUBLE DANDA
```

9 References

- Grierson, George A. 1916a. *The Linguistic Survey of India*. Vol. IX. Indo-Aryan Family. Central Group. Part I. Specimens of Western Hindī and Pañjābī. Calcutta: Office of the Superintendent of Government Printing, India.
- . 1916b. *The Linguistic Survey of India*. Vol. IX. Indo-Aryan Family. Central Group. Part IV. Specimens of Pahārī Languages and Gujuri. Calcutta: Office of the Superintendent of Government Printing, India.
- Jensen, Hans. 1969. *Die Schrift: In Vergangenheit und Gegenwart*. Reprint der 3. Auflage. Berlin: Deutscher Verlag der Wissenschaften.
- Leitner, Gottlieb William. 1882. “A Collection of Specimens of Commercial and Other Alphabets and Hand-writings as also of Multiplication Tables Current in Various Parts of the Panjab, Sind and the North West Provinces”. Published as Appendix V of *History of Indigenous Education in the Punjab*. Lahore: Anjuman-i-Punjab Press.
- Pandey, Anshuman. 2009a. “Proposal to encode the Sharada script in Unicode” (L2/09-074R2). <https://www.unicode.org/L2/L2009/09074r2-n3595-sharada.pdf>
- . 2009b. “Proposal to Encode the Takri Script in ISO/IEC 10646” (L2/09-424).

<http://www.unicode.org/L2/L2009/09424-takri.pdf>

———. 2011a. “Final Proposal to Encode the Khojki Script in ISO/IEC 10646” (L2/11-021).

<https://www.unicode.org/L2/L2011/11021-khojki.pdf>

———. 2011b. “Proposal to Encode the Mahajani Script in ISO/IEC 10646” (L2/11-274).

<https://www.unicode.org/L2/L2011/11274-n4126-mahajani.pdf>

———. 2015. “Proposal to encode the Dogra script in Unicode” (L2/15-234R).

<http://www.unicode.org/L2/L2015/15234r-dogra.pdf>

———. 2018. “Preliminary proposal to encode Sirmauri in Unicode” (L2/18-085).

<https://www.unicode.org/L2/L2018/18085-sirmauri.pdf>

———. 2021. “Proposal to encode the Khojki letter SHORT I in Unicode” (L2/21-104).

<https://www.unicode.org/L2/L2021/21104-khojki-letter-short-i.pdf>

Pandey, Anshuman; Mandal, Biswajit. 2025. “Proposal to encode Sirmauri in Unicode” (L2/25-038).

<https://www.unicode.org/L2/L2025/25038-sirmauri.pdf>

Punjab Government. 1907. *Sirmur State Gazetteer, 1904*. Part A. With Maps. Lahore: The Punjab Government Press.










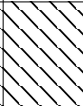






10 Acknowledgments

Our work to encode Sirmauri would not be possible without assistance from the following, to whom we express our heartfelt gratitude:

- Nikhil Bhardwaj (‘Project Pahari’: http://x.com/project_pahari)
- Sherjung Chauhan (Sirmaur)
- Bhupal Singh Chauhan (Nohradhar, Sirmaur)
- The family of Natiram Negi (Maleata, Jaunsar-Bawar)
- Gangaram Pabuch (Sirmaur)
- Yatin Pandit (Kullu)
- Dr. Om Prakash Sharma (Shimla)
- Dr. Dilip Singh Tilkan (Sirmaur)
- The family of Har Singh Tilkan (Shilla village, Sirmaur)
- The family of Ved Singh Tilkan (Shilla village, Sirmaur)

This project has been made possible in part by grants from:

- The U.S. National Endowment for the Humanities, which funded the Universal Scripts Project PR-253360-17 (part of the Script Encoding Initiative at UC Berkeley)
- The Mellon Foundation to the Script Encoding Initiative at the University of California, Berkeley

	1185	1186	1187	1188	1189
0	3। 11850	3 11860	3n 11870		o 11890
1	3।। 11851	4 11861	on 11871	ȯ 11881	। 11891
2	6 11852	5 11862	e 11872	ȯ̇ 11882	2 11892
3	6 11853	uy 11863	on 11873	ȯ। 11883	3 11893
4	6 11854		2 11874	ȯ। 11884	4 11894
5	6 11855	U 11865	7 11875	ȯ 11885	4 11895
6	2 11856	c 11866	m 11876	ȯ: 11886	E 11896
7	3। 11857	6 11867		ȯ 11887	7 11897
8	3। 11858	6 11868		ȯ 11888	2 11898
9	3। 11859	u 11869		ȯ 11889	2 11899
A	3। 1185A	n 1186A	4 1187A	2 1188A	
B	4 1185B	on 1186B	5 1187B	2 1188B	
C	on 1185C	E 1186C	ȯ। 1187C	2 1188C	
D	7। 1185D	2 1186D			
E	uy 1185E	on 1186E	ȯ2 1187E		
F		5 1186F	ȯ 1187F		

Independent vowels

11850	3i	SIRMAURI LETTER A
11851	3ii	SIRMAURI LETTER AA
11852	6	SIRMAURI LETTER I
11853	6o	SIRMAURI LETTER II
11854	6u	SIRMAURI LETTER U
11855	6uu	SIRMAURI LETTER UU
11856	6e	SIRMAURI LETTER E
11857	3ii	SIRMAURI LETTER ALTERNATE E
11858	3i	SIRMAURI LETTER AI
11859	3ii	SIRMAURI LETTER O
1185A	3ii	SIRMAURI LETTER AU

Consonants

1185B	5	SIRMAURI LETTER KA
1185C	5	SIRMAURI LETTER KHA
1185D	7i	SIRMAURI LETTER GA
1185E	7ii	SIRMAURI LETTER GHA
1185F	7ii	<reserved>
11860	7ii	SIRMAURI LETTER CA
11861	7ii	SIRMAURI LETTER CHA
11862	7ii	SIRMAURI LETTER JA
11863	7ii	SIRMAURI LETTER JHA
11864	7ii	<reserved>
11865	7ii	SIRMAURI LETTER TTA
11866	7ii	SIRMAURI LETTER TTHA
11867	7ii	SIRMAURI LETTER DDA
11868	7ii	SIRMAURI LETTER DDHA
11869	7ii	SIRMAURI LETTER NNA
1186A	7ii	SIRMAURI LETTER TA
1186B	7ii	SIRMAURI LETTER THA
1186C	7ii	SIRMAURI LETTER DA
1186D	7ii	SIRMAURI LETTER DHA
1186E	7ii	SIRMAURI LETTER NA
1186F	7ii	SIRMAURI LETTER PA
11870	7ii	SIRMAURI LETTER PHA
11871	7ii	SIRMAURI LETTER BA
11872	7ii	SIRMAURI LETTER BHA
11873	7ii	SIRMAURI LETTER MA
11874	7ii	SIRMAURI LETTER YA
11875	7ii	SIRMAURI LETTER RA
11876	7ii	SIRMAURI LETTER LA
11877	7ii	<reserved>
11878	7ii	<reserved>
11879	7ii	<reserved>
1187A	7ii	SIRMAURI LETTER SA
1187B	7ii	SIRMAURI LETTER HA

Dependent vowel signs

1187C	o	SIRMAURI VOWEL SIGN AA
1187D	7ii	<reserved>
1187E	o	SIRMAURI VOWEL SIGN II
1187F	o	SIRMAURI VOWEL SIGN U
11880	7ii	<reserved>
11881	o	SIRMAURI VOWEL SIGN E
11882	o	SIRMAURI VOWEL SIGN AI
11883	o	SIRMAURI VOWEL SIGN O
11884	o	SIRMAURI VOWEL SIGN AU

Various signs

11885	o	SIRMAURI SIGN ANUSVARA
11886	o:	SIRMAURI SIGN VISARGA
11887	o	SIRMAURI SIGN VIRAMA
		• vowel silencer
		• does not control conjunct formation
11888	o	SIRMAURI SIGN NUKTA

Consonant sign

11889	o	SIRMAURI CONSONANT SIGN MEDIAL RA
		= ra-kara

Auspicious sign

1188A	o	SIRMAURI EKAM
		= anji

Punctuation

1188B	o	SIRMAURI SECTION MARK
1188C	o	SIRMAURI DOUBLE SECTION MARK

Digits

11890	o	SIRMAURI DIGIT ZERO
11891	1	SIRMAURI DIGIT ONE
11892	2	SIRMAURI DIGIT TWO
11893	3	SIRMAURI DIGIT THREE
11894	4	SIRMAURI DIGIT FOUR
11895	5	SIRMAURI DIGIT FIVE
11896	6	SIRMAURI DIGIT SIX
11897	7	SIRMAURI DIGIT SEVEN
11898	8	SIRMAURI DIGIT EIGHT
11899	9	SIRMAURI DIGIT NINE

The Sirmauri Alphabet.

a	3	da	ε
ā	3)	dha	εs
i, ī & ē	6 or 5	na	7
u & ū	3	pa	5
ē	3 or 5	pha	εh
ai	3)	ba	9
ō & ȳ	3)	bha	9
au	3)	ma	7
ka	x	ya	x
kha	h	ra	7
ga	u	la	3
gha	ε	wa	9
cha	h	śa	9
chha	8	sha	h
ja	7	sa	9
jha	7	ha	ε
ṭa	ε	kā	x
ṭha	6	ki, kī	x
ḍa	5	ku, kū	x
ḍha	u	ke, kē	x
ṇa	7	kai	x
ta	7	ko, kō	x
tha	ε	kau	x

Figure 1: Chart of the Sirmauri form of Takri (from Grierson 1916b: 459).

WESTERN PAHĀRĪ (SIRMAURĪ).

DHĀRTHĪ DIALECT.

SPECIMEN I.

ਤਖ਼ਾ ਭੀਠੇਠੇ ਏ ਭੀਠੇ ਭੀਠੇ "
 ਖਾਠੇਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ "
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ "
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ "
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ
 ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ ਭੀਠੇ

Figure 2: Specimen of the Sirmauri form of Takri (from Grierson 1916b: 468).

JAUNSĀRĪ (SIRMAURĪ) ALPHABET.

3	a	⌒ x)	kau	3	tha
3), 3))	ā	ḡ	kā	4	da
6	i	x	ka	5	dha
6	ī	h	kha	7	na
6	u	7)	ga	8	pa
6	ū	6)	gha	9	pha
7, x	ě, ē	—	ṇa	0	ba
6	ai	5	cha	1	bha
3)	ō, ō	6	chha	2	ma
3)	au	7	ja	3	ya
.	~	4	jha	4	ra
x	ka	—	ṅa	5	la
7)	ku	✓	tḡ	6	wa
x)	ki	✓ 6	tha	7	śa
x)	kī	1 6	ḍa	8	sha
6	ku	6	ṛa	9	sa
6	kū	6	ḍha	0	ha
6	kě, kē	6	ṛha		
6	kai	7)	ṇa		
6)	kō, kō	7)	ta		

When used before consonants of their own class, nasals are, as usual, represented by Anusvāra (◌ṃ). As ṇ and ṅ occur only in this position, they have no special character

Figure 3: Chart of the Jaunsari form of Takri (from Grierson 1916b: 384).

DISTRICT DERA DUN.

[illegible]

Figure 4: Specimen of the Jaunsari form of Takri (from Grierson 1916b: 401).

Lautwert	Śāradā 804	Kasch- mirī	Ṭākri		Laṇḍā		Multani	Gur- mukhi
			Jaun- sari	Cha- meāli	Khu- dāwādi	Sindhi- Schrift		
a	अ	अ	3	3	ṁ	ṁ	ṁ	ਅ
i	इ	इ	४	४	"	४	४	਀
u	उ	उ	५	५	"	५	५	ਁ
e		ए	६	६	"	६	६	ਂ
o		ओ	७	७	"	७	७	ਃ
ā	आ	आ	८	८	"	८	८	਄
ka	क	क	९	९	क	क	क	ਕ
kha	ख	ख	१०	१०	ख	ख	ख	ਖ
ga	ग	ग	११	११	ग	ग	ग	ਗ
gha	घ	घ	१२	१२	घ	घ	घ	ਘ
ṇa	ण	ण	१३	१३	ṇ	ṇ	ṇ	ਙ
ḥa	ह	ह	१४	१४	ह	ह	ह	ਹ
ḥha	ह्रस्व	ह्रस्व	१५	१५	ह्रस्व	ह्रस्व	ह्रस्व	ਠ
ḡa	गुरु	गुरु	१६	१६	गुरु	गुरु	गुरु	ਗੁਰੂ
ḡha	गुरुह्रस्व	गुरुह्रस्व	१७	१७	गुरुह्रस्व	गुरुह्रस्व	गुरुह्रस्व	ਗੁਰੂਠ
ṇa	ण	ण	१८	१८	ṇ	ṇ	ṇ	ਙ
ṭa	ट	ट	१९	१९	ट	ट	ट	ਟ
ṭha	ठ	ठ	२०	२०	ठ	ठ	ठ	ਠ
ḍa	ड	ड	२१	२१	ड	ड	ड	ਡ
ḍha	ढ	ढ	२२	२२	ढ	ढ	ढ	ਢ
na	न	न	२३	२३	न	न	न	ਨ
ta	त	त	२४	२४	त	त	त	ਤ
tha	थ	थ	२५	२५	थ	थ	थ	ਥ
da	द	द	२६	२६	द	द	द	ਦ
dha	ध	ध	२७	२७	ध	ध	ध	ਧ
na	न	न	२८	२८	न	न	न	ਨ
pa	प	प	२९	२९	प	प	प	ਪ
pha	फ	फ	३०	३०	फ	फ	फ	ਫ
ba	ब	ब	३१	३१	ब	ब	ब	ਬ
bha	भ	भ	३२	३२	भ	भ	भ	ਭ
ma	म	म	३३	३३	म	म	म	ਮ
ya	य	य	३४	३४	य	य	य	ਯ
ra	र	र	३५	३५	र	र	र	ਰ
la	ल	ल	३६	३६	ल	ल	ल	ਲ
va	व	व	३७	३७	व	व	व	ਵ
śa	श	श	३८	३८	श	श	श	ਸ਼
ṣa	ष	ष	३९	३९	ष	ष	ष	ਸ਼
sa	स	स	४०	४०	स	स	स	ਸ
ha	ह	ह	४१	४१	ह	ह	ह	ਹ

Figure 5: Comparison of Sirmauri (Jaunsari) with related scripts (from Jensen 1969: 366).

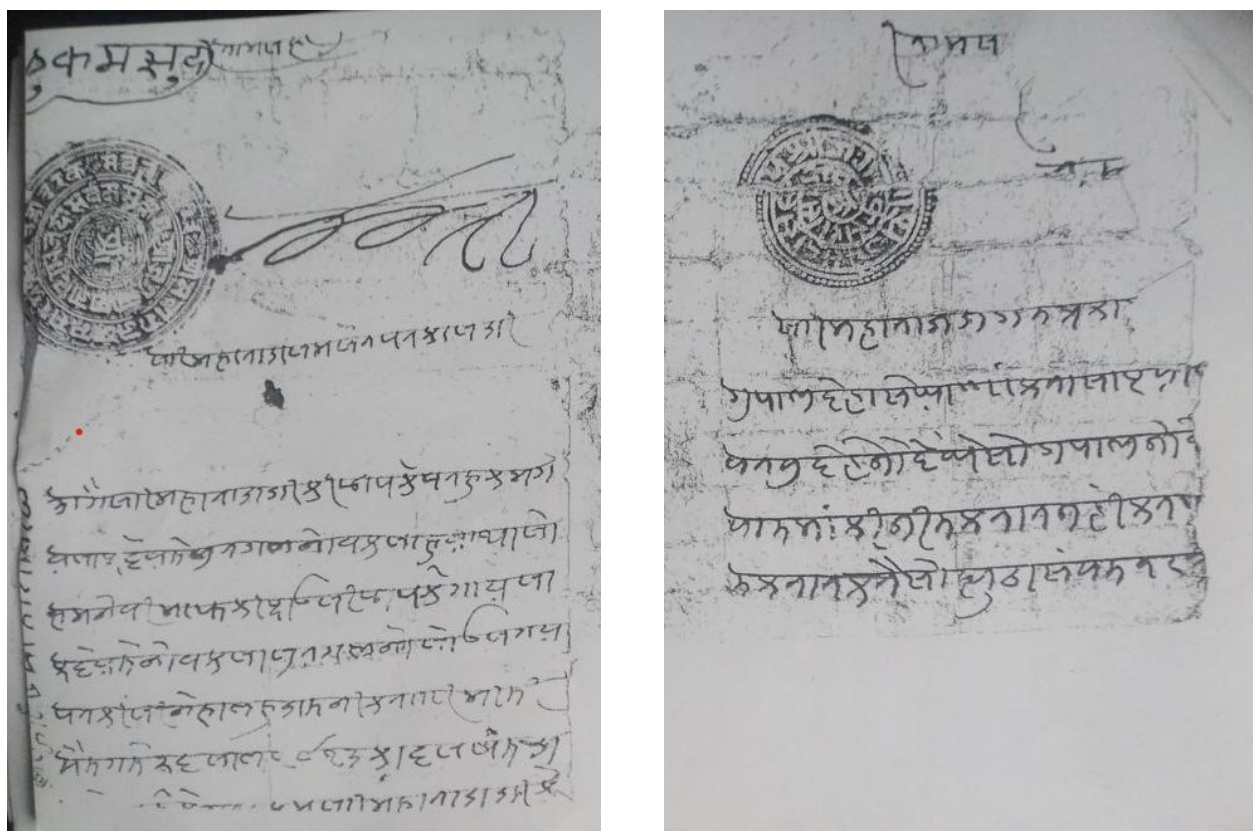


Figure 6: Documents recording land donations to the Shirgul Devta temple by *rājā* Shamsher Prakash (1770–1789) (left) and *rājā* Jagat Prakash (1856–1898) (right).

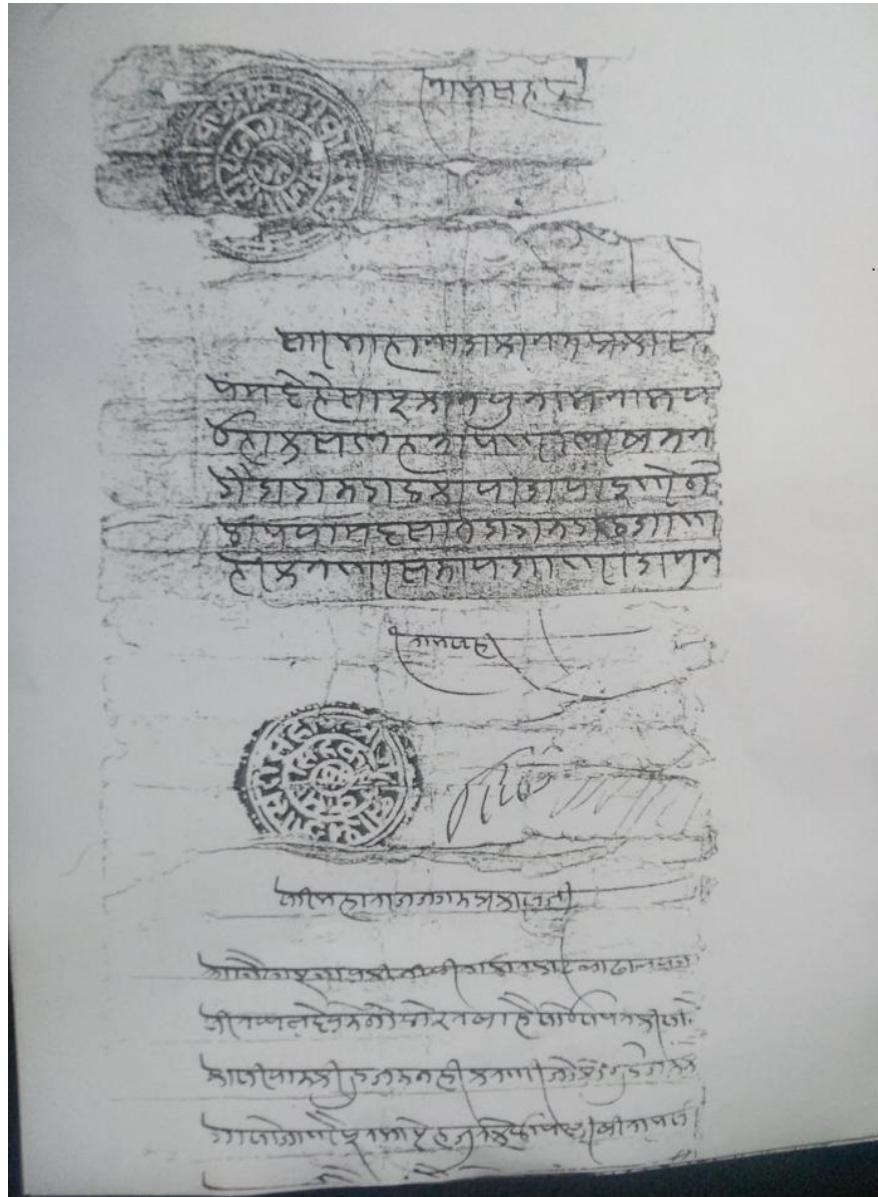


Figure 7: Record of land donation to the Shirgul Devta temple by *rājā* Karma Prakash (1804–1815).

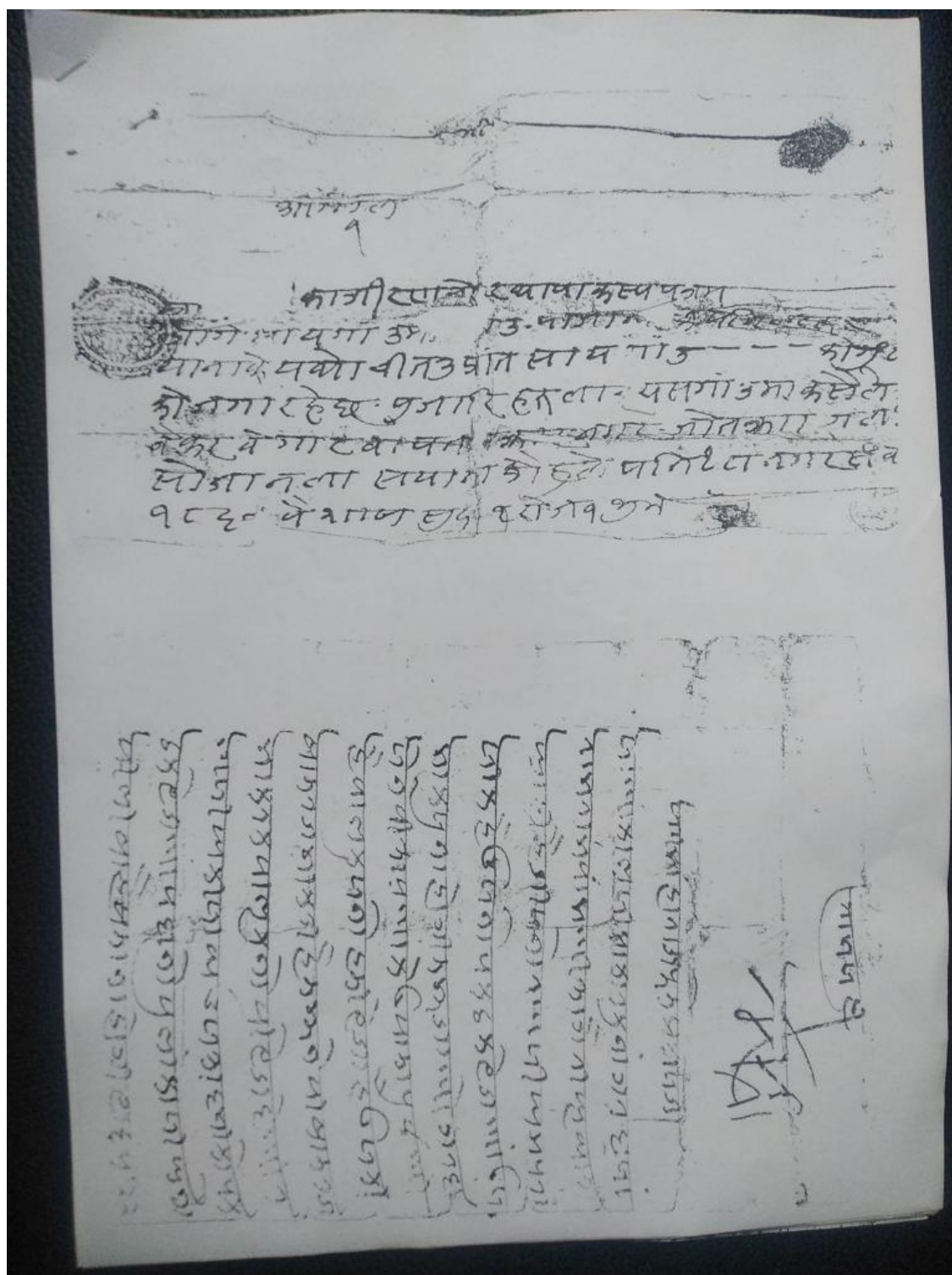


Figure 8: Record of land donation to the Shirgul Devta temple by *rājā* Fateh Prakash (1815–1850).

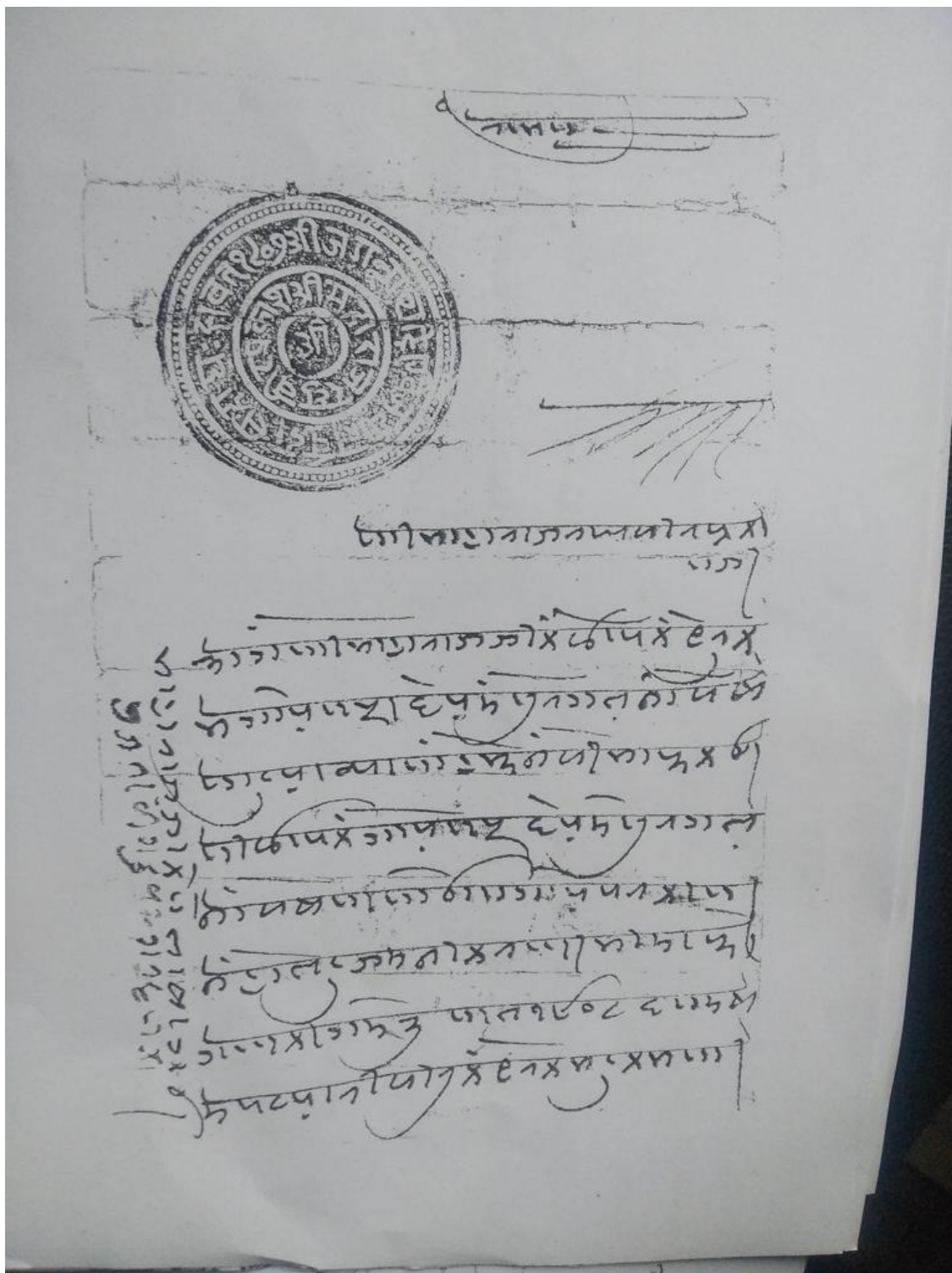


Figure 9: Record of land donation to the Shirgul Devta temple by Raghbir Prakash II (1850–56).

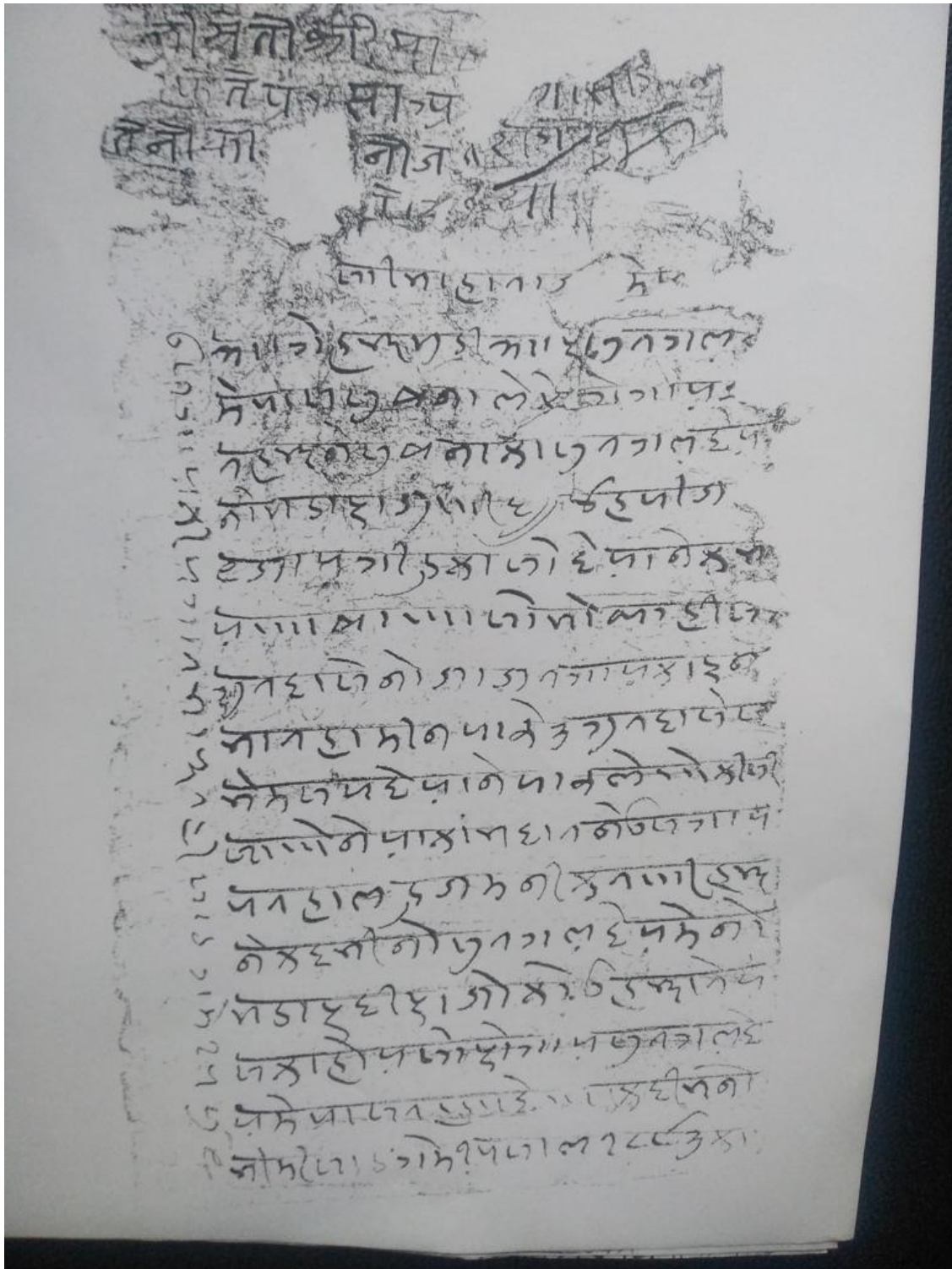


Figure 10: Record of land donation to the Shirgul Devta temple by *rājā* Raghbir Prakash II (1850–56).

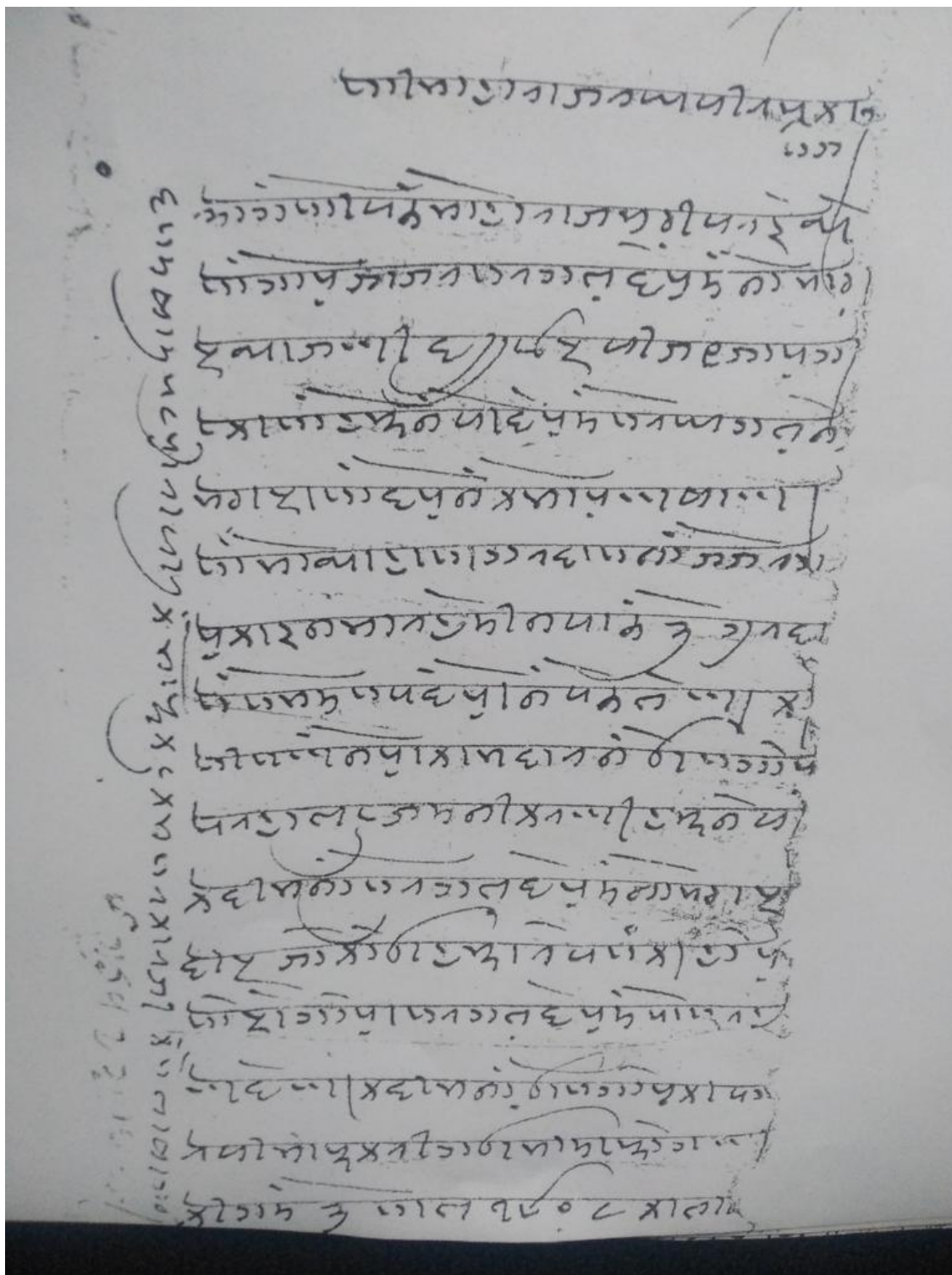


Figure 11: Record of land donation to the Shirgul Devta temple by *rājā* Raghbir Prakash II (1850–56).

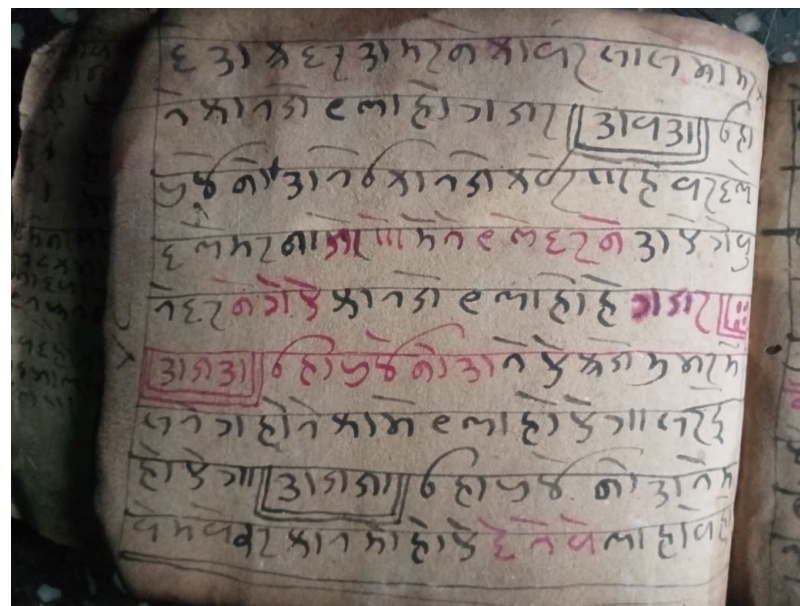
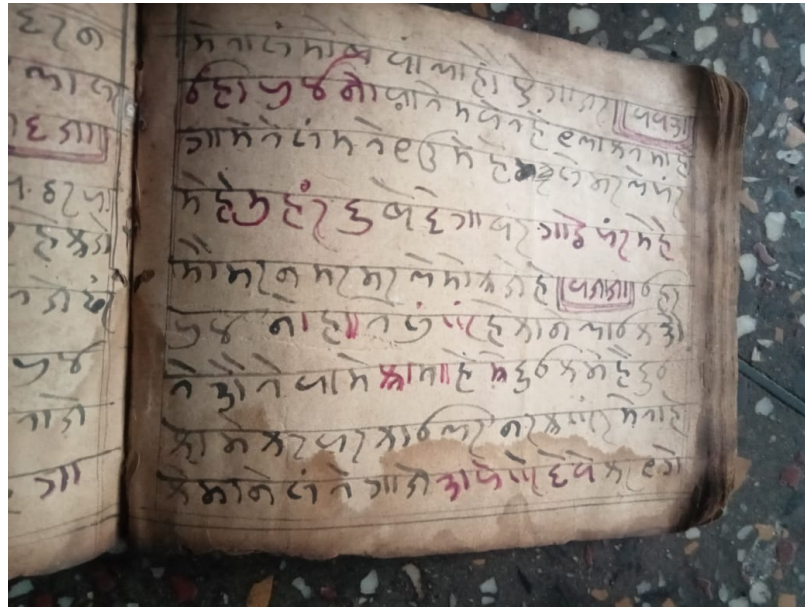


Figure 13: Sancha text written in the Jaunsari variety by Natiram Negi.

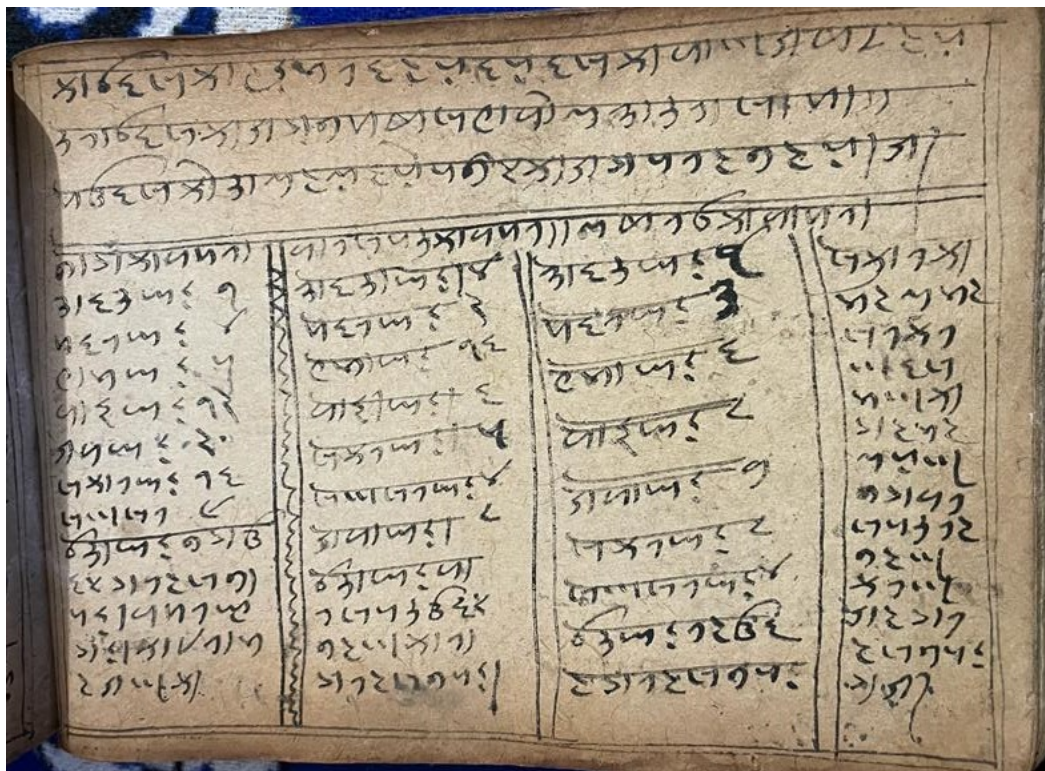
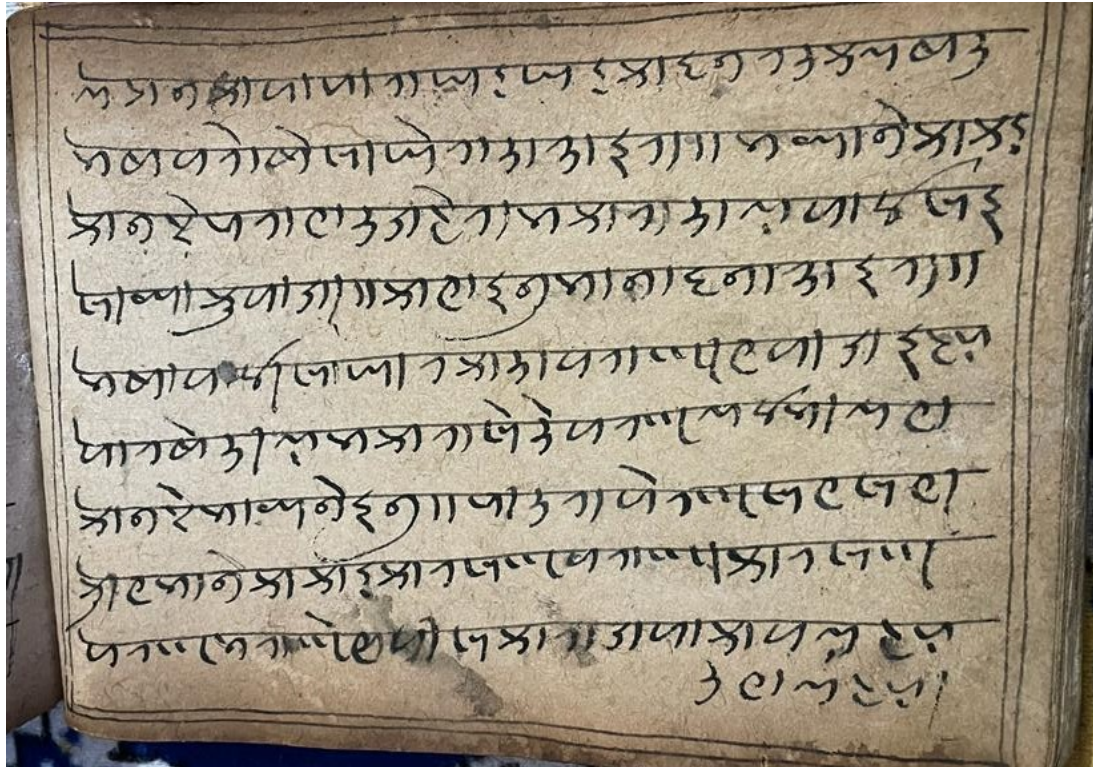


Figure 14: Sancha text written in Sirmauri by Gangaram Pabuch.

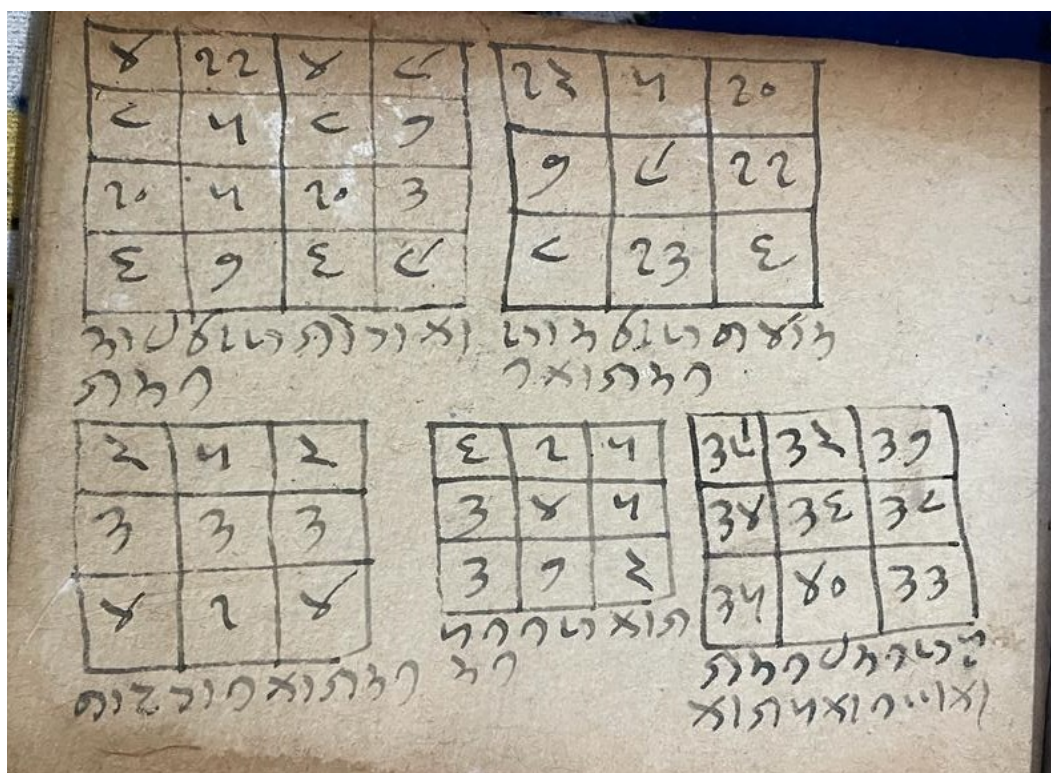
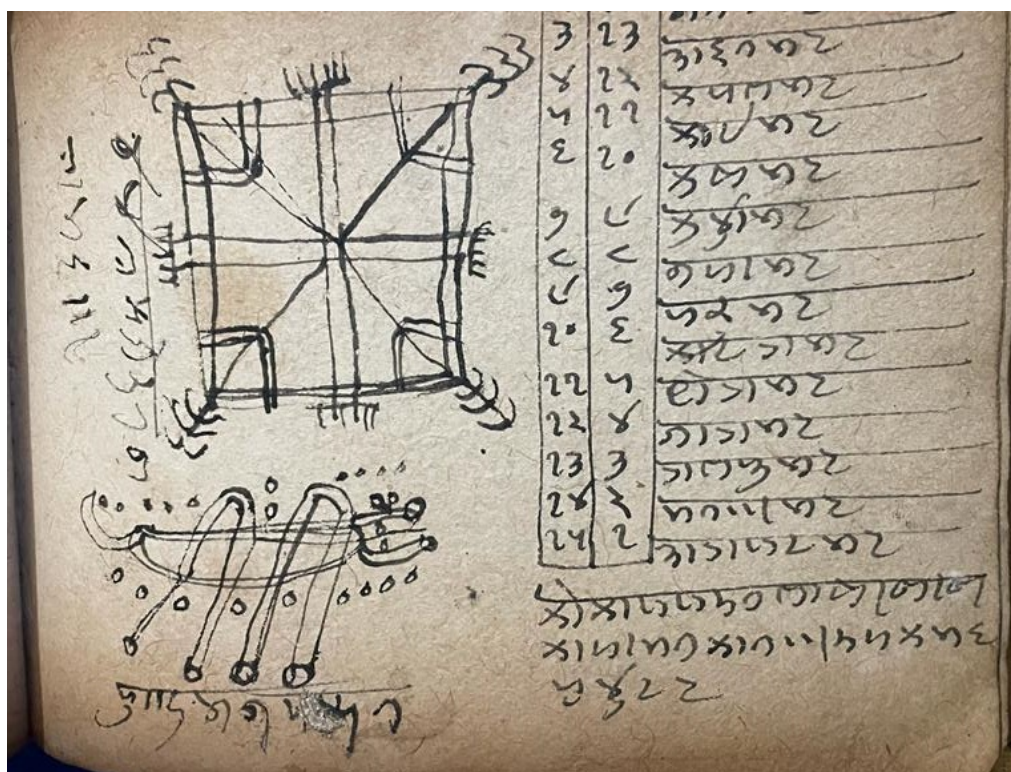


Figure 15: Sancha text written in Sirmauri by Gangaram Pabuch.

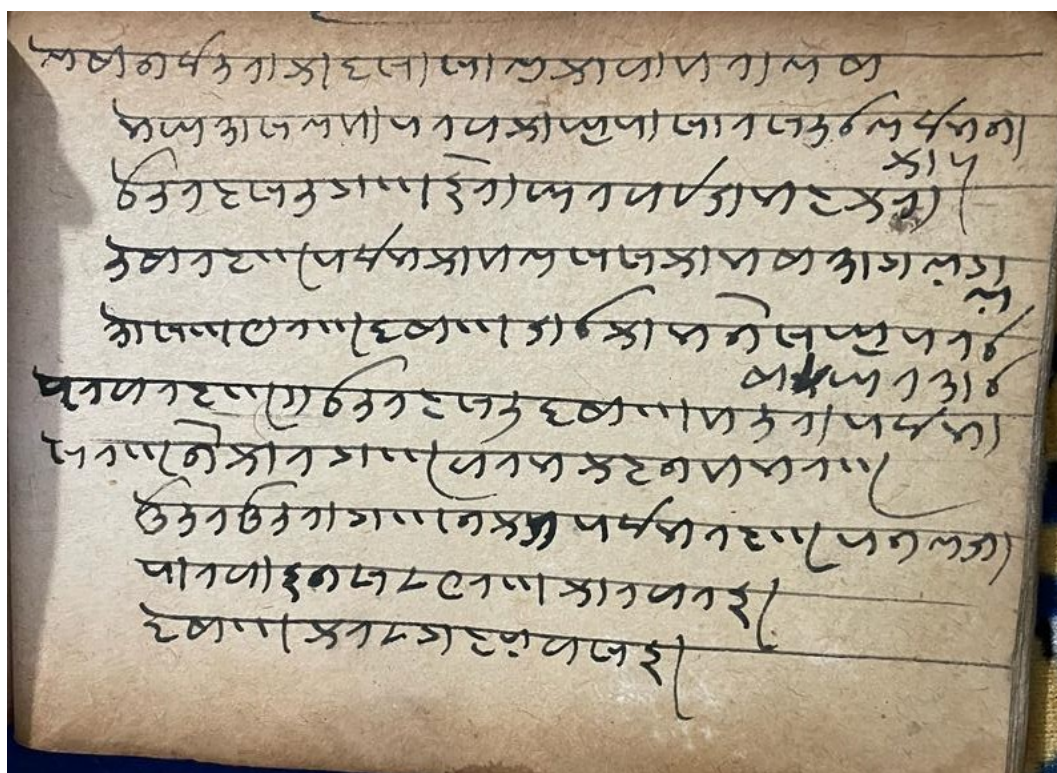


Figure 16: Sancha text written in Sirmauri by Gangaram Pabuch.

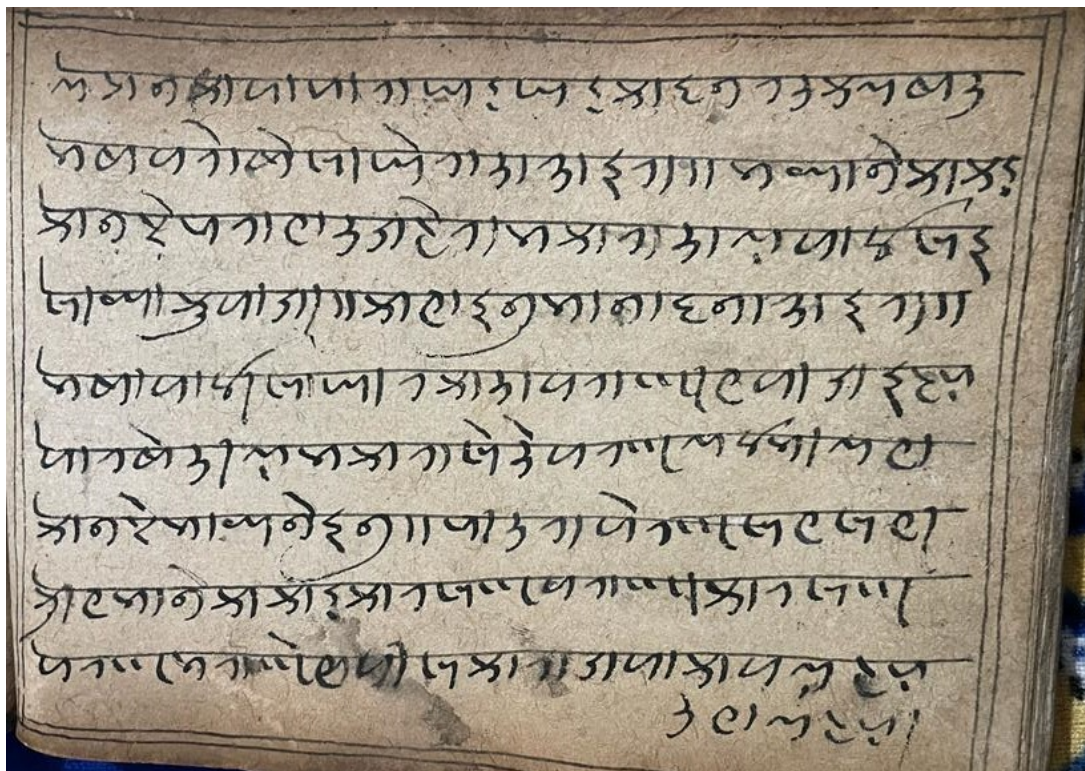
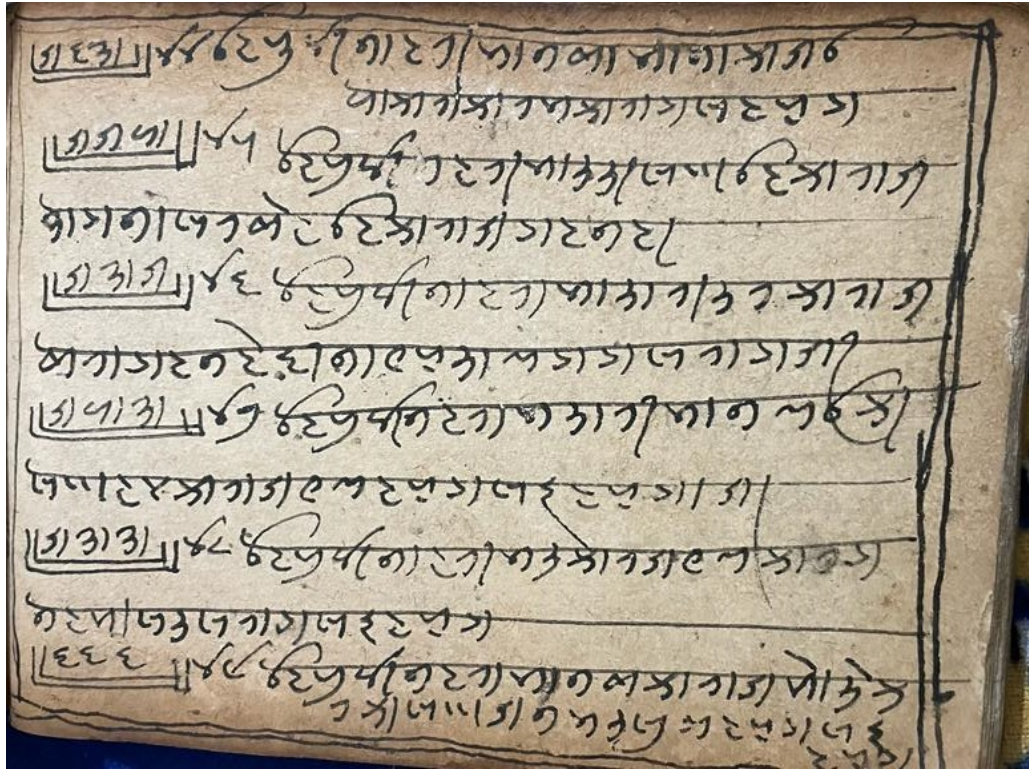


Figure 17: Sancha text written in Sirmauri by Gangaram Pabuch.

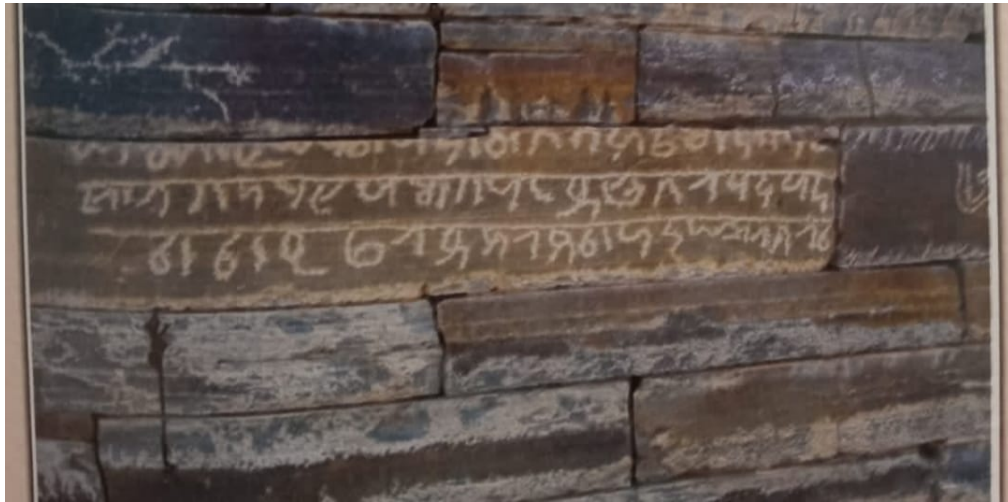


Figure 18: Inscription found on a tower (*killā*) adjacent to a house

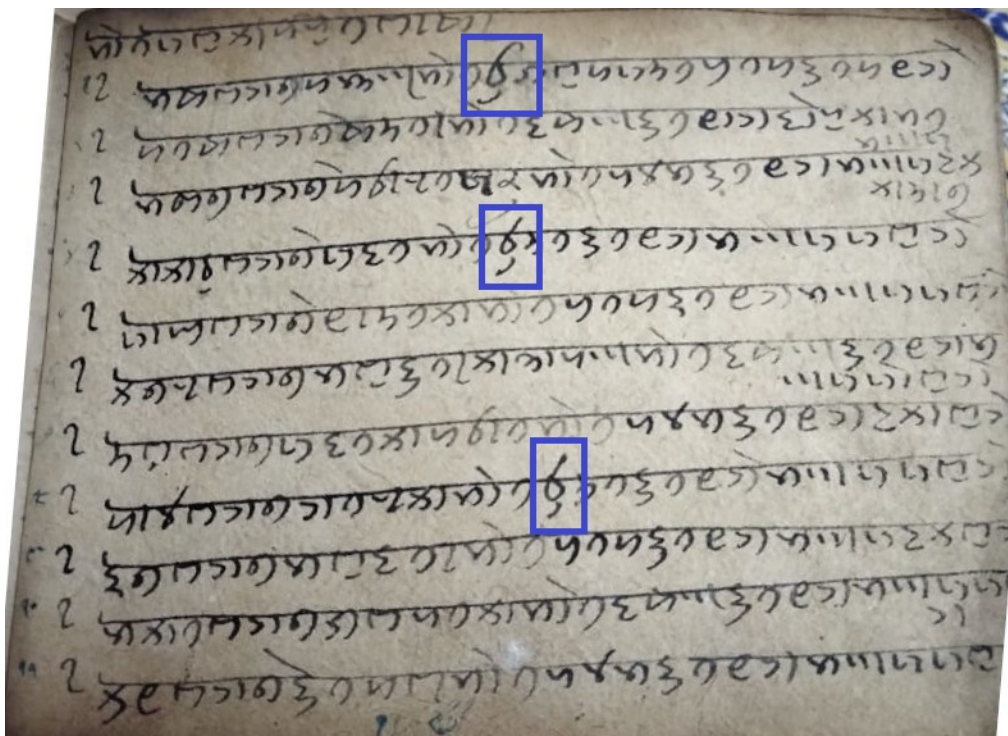


Figure 19: Examples of *᳚* for writing *ī*.

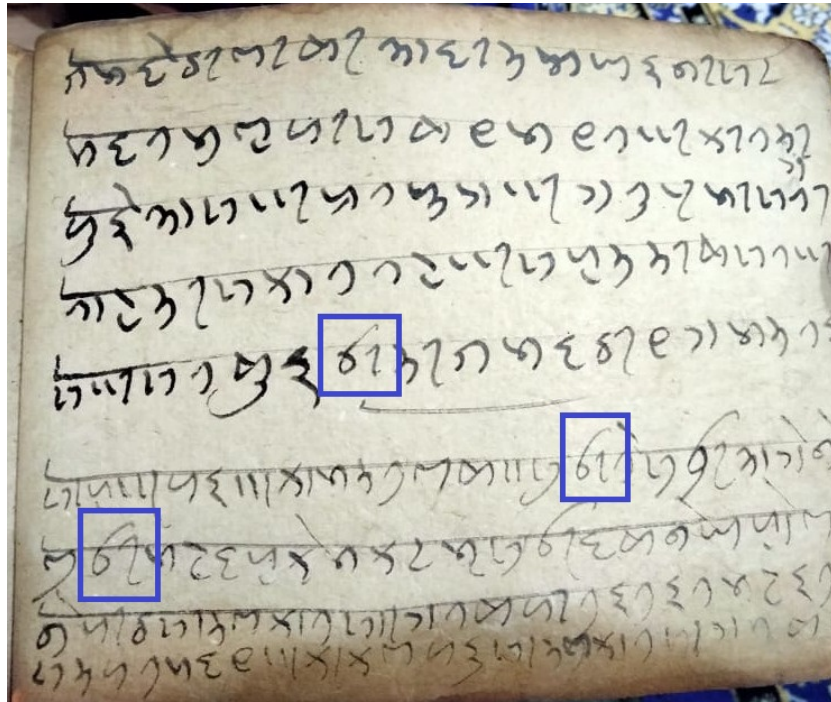


Figure 20: Example of 6ॐ for representing *ī*, illustrating usage of 6 used as a vowel carrier bearing the vowel sign ॐ.

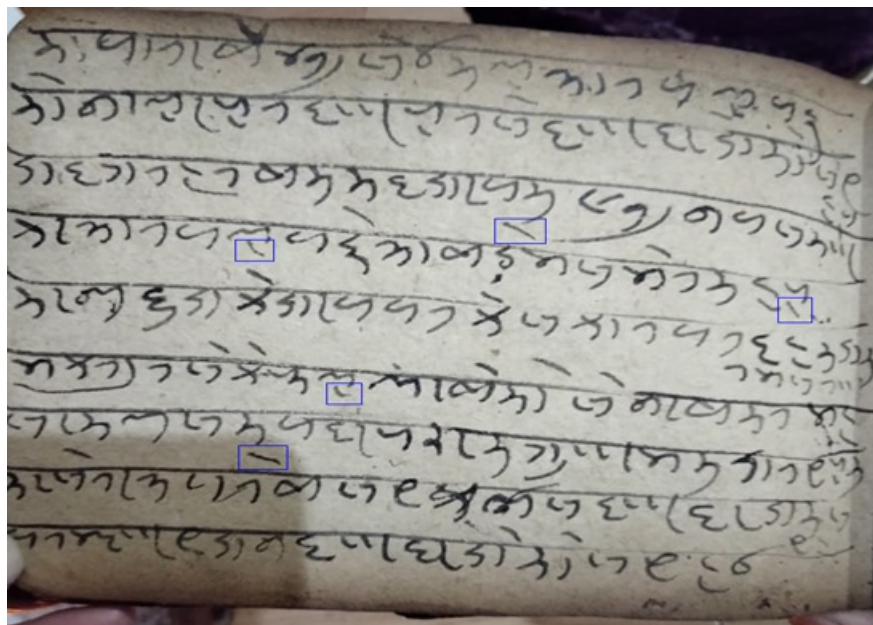
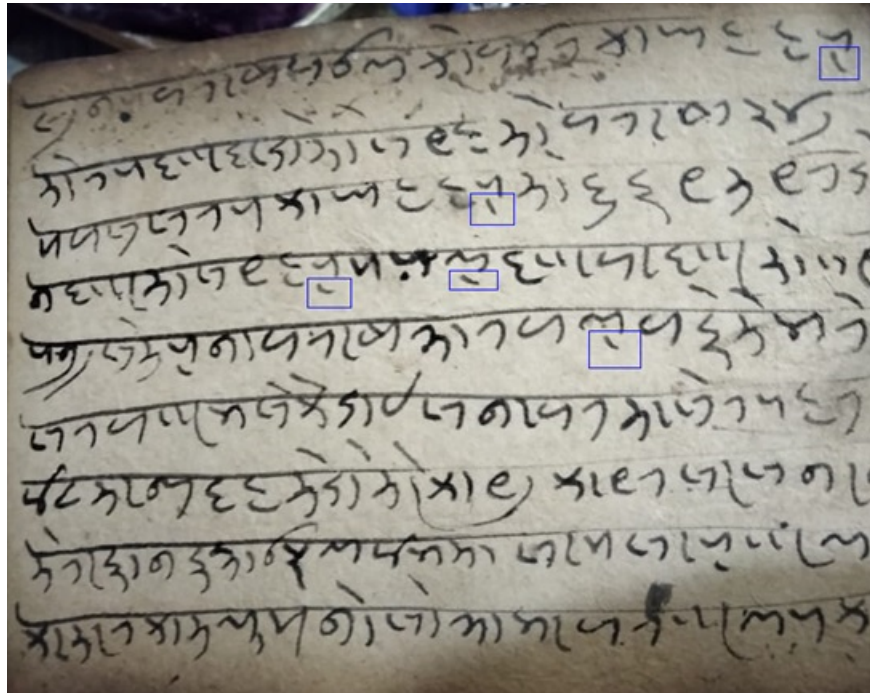
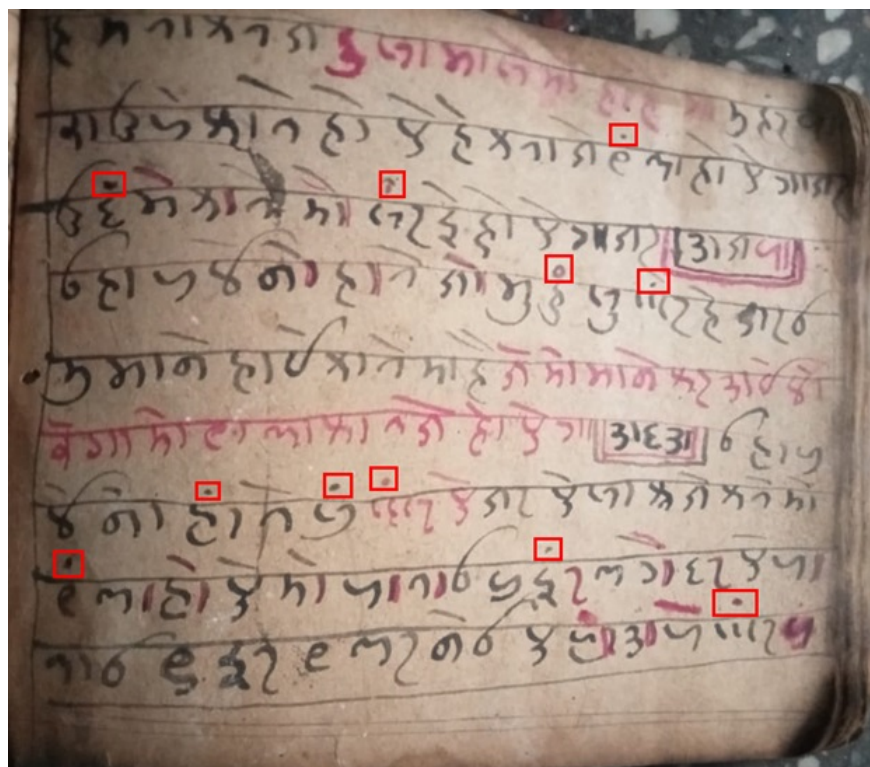


Figure 21: Usage of *halanta*.

Figure 22: Additional examples of *halanta*.Figure 23: Usage of *anusvāra*

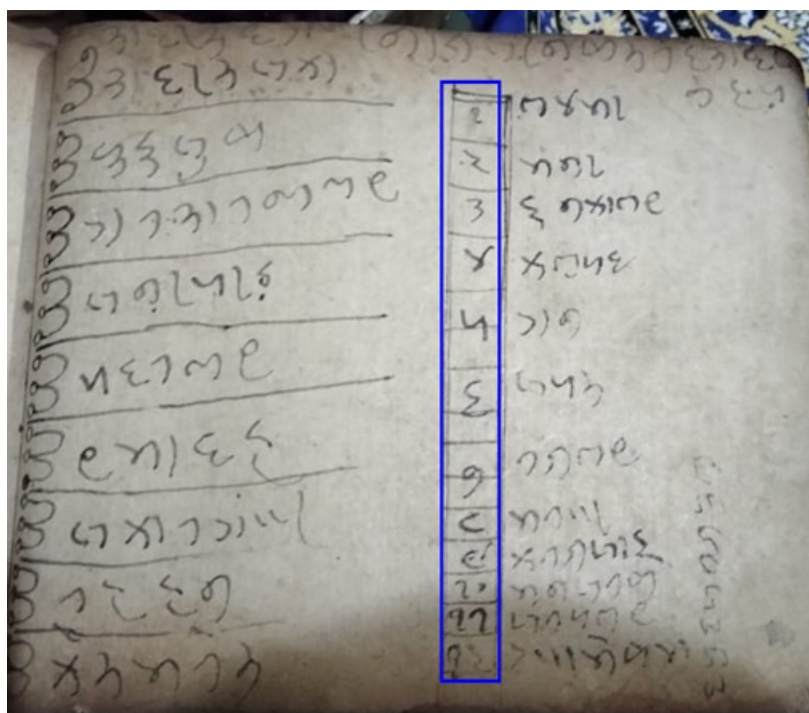


Figure 24: Examples of digits.

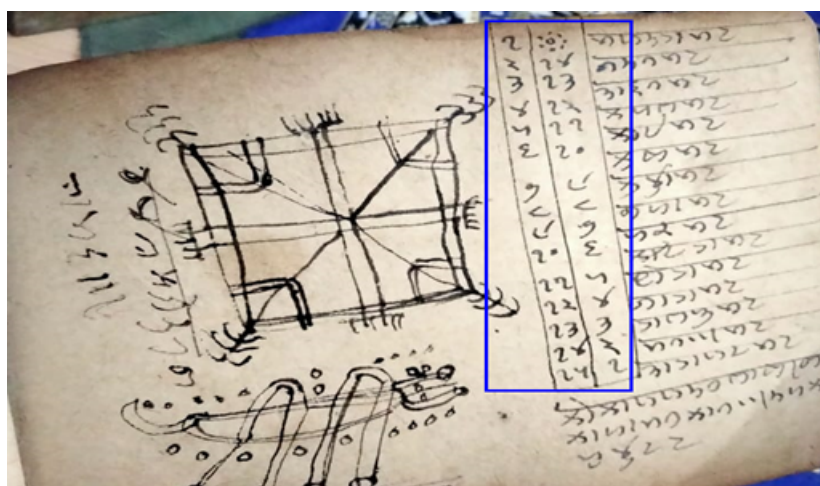


Figure 25: Examples of digits.

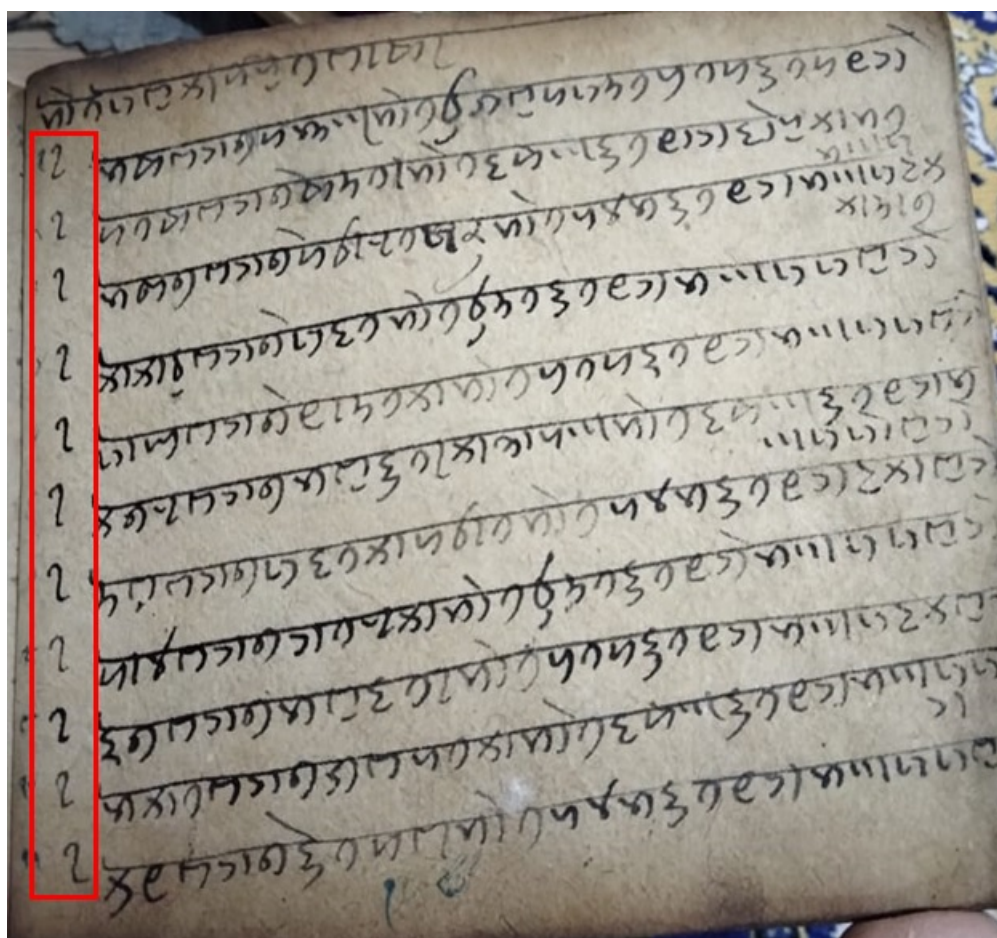


Figure 26: Usage of the auspicious sign ॐ ekam

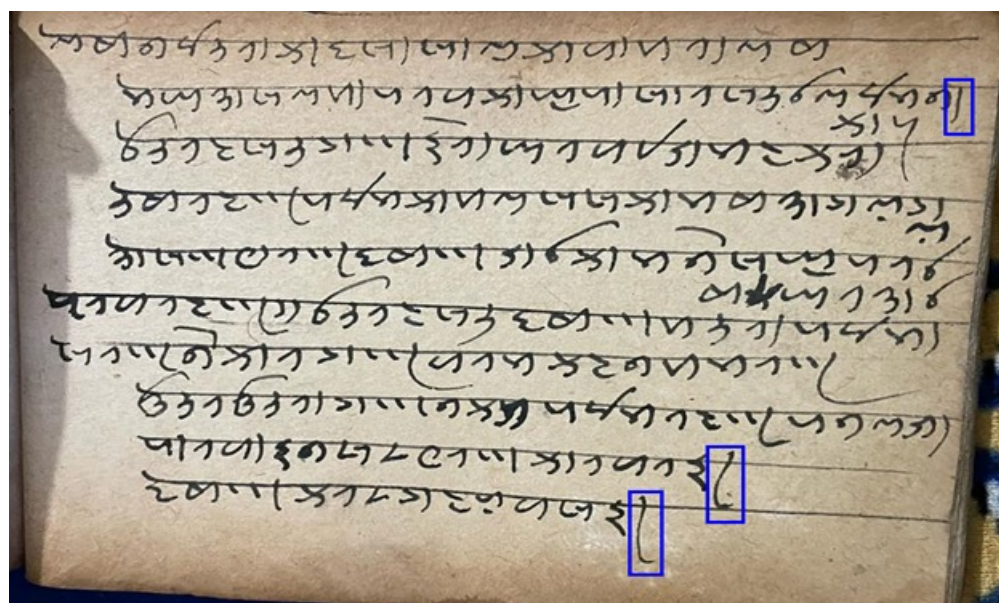


Figure 27: Usage of daṇḍā-s.

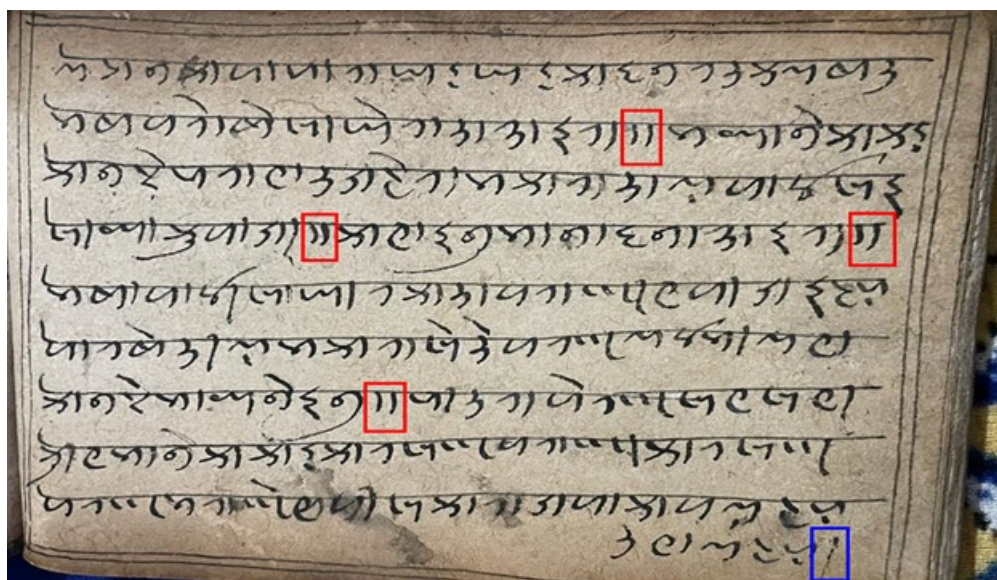
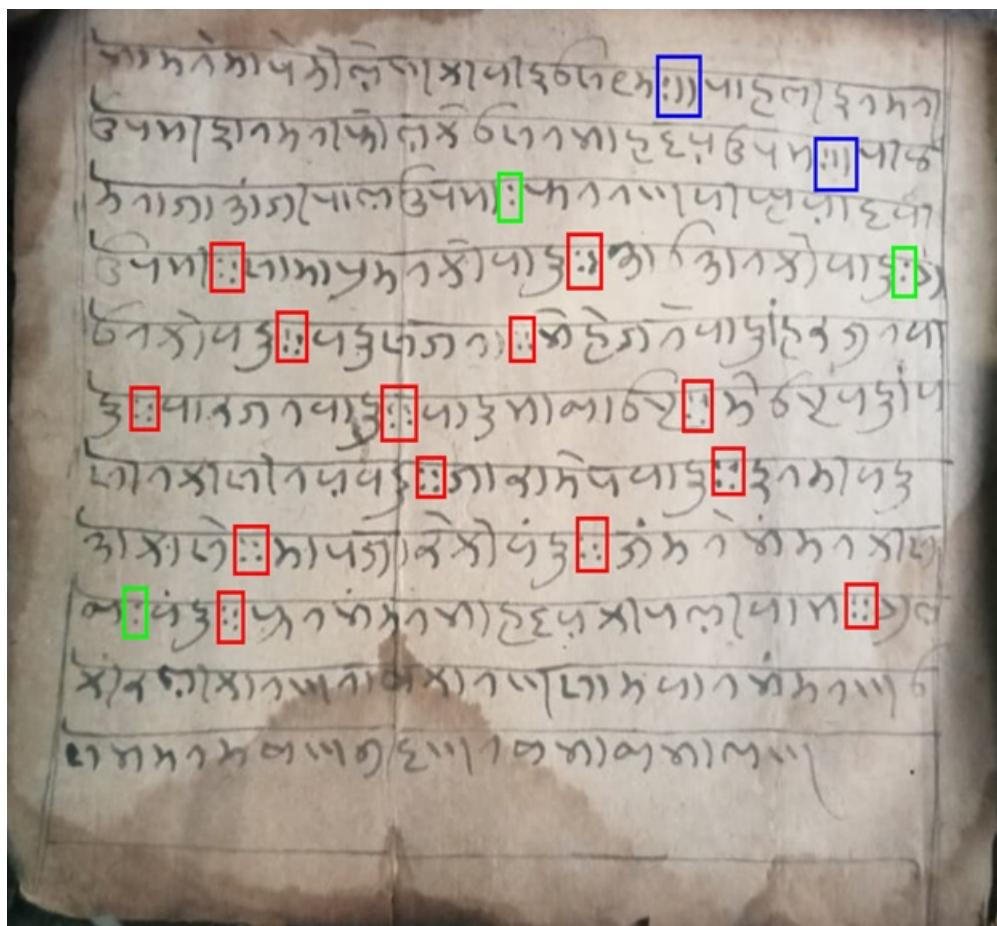
Figure 28: Single (blue) and double (red) *danḍā*-s.

Figure 29: Various punctuation signs.

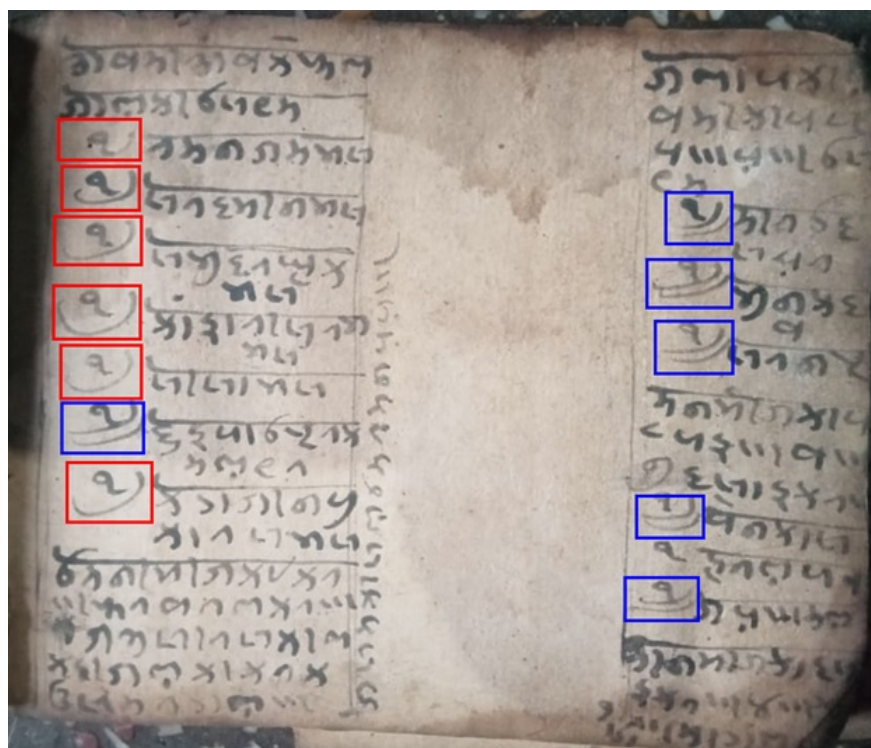


Figure 30: Section signs.

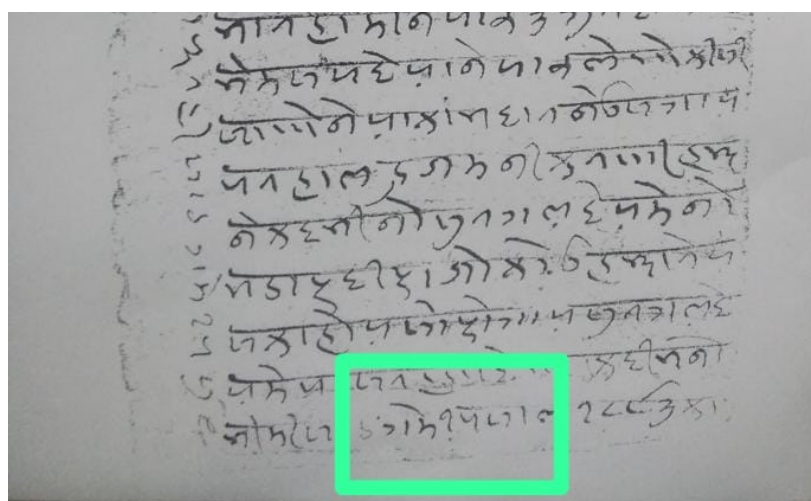


Figure 31: Usage of a Latin question mark in a Sirmauri document.

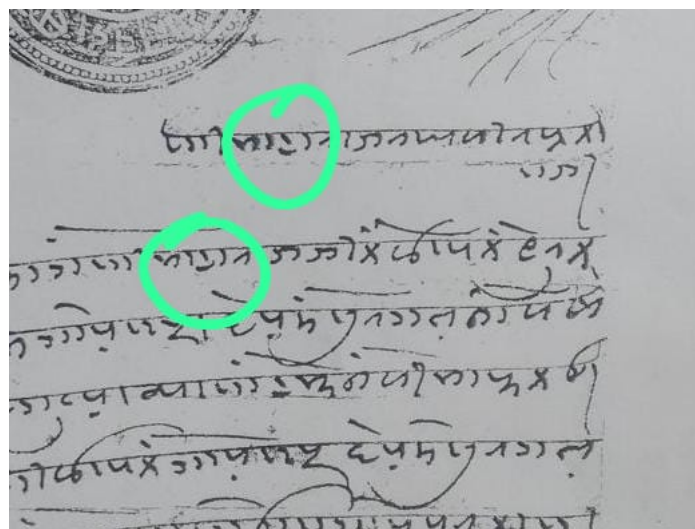


Figure 32: Document showing variant form ॐ of ॐ ha.

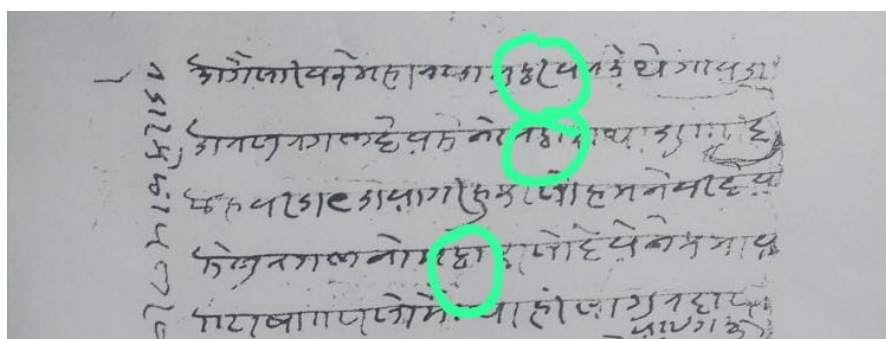


Figure 33: Document showing Jaunsari form ॐ of ॐ dha.

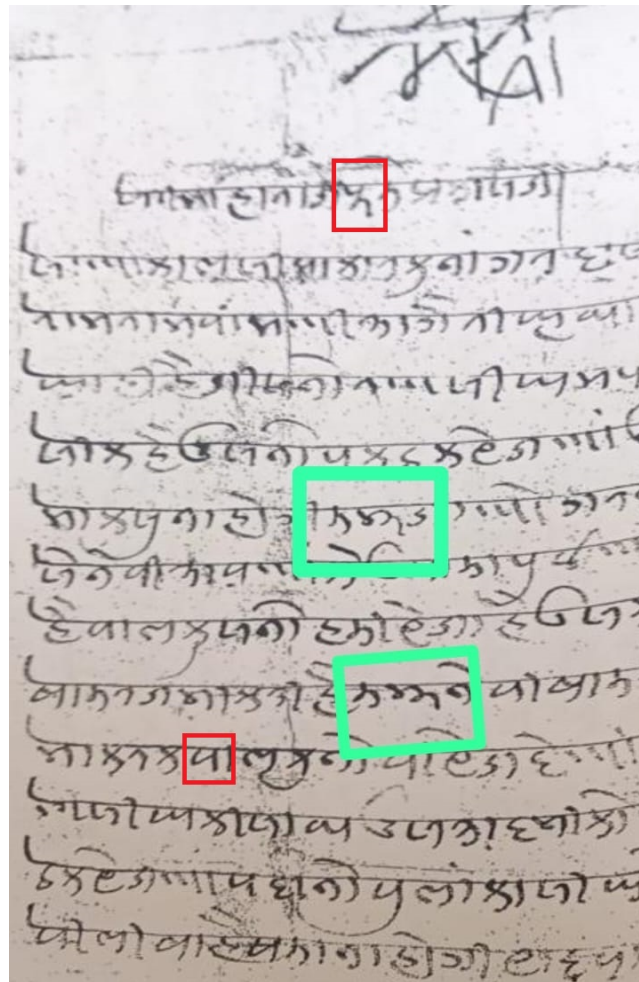


Figure 34: Document showing usage of the conjunct *mma* (green) and Devanagari-like forms for *pha* (red, top) and *pa* (red, bottom).

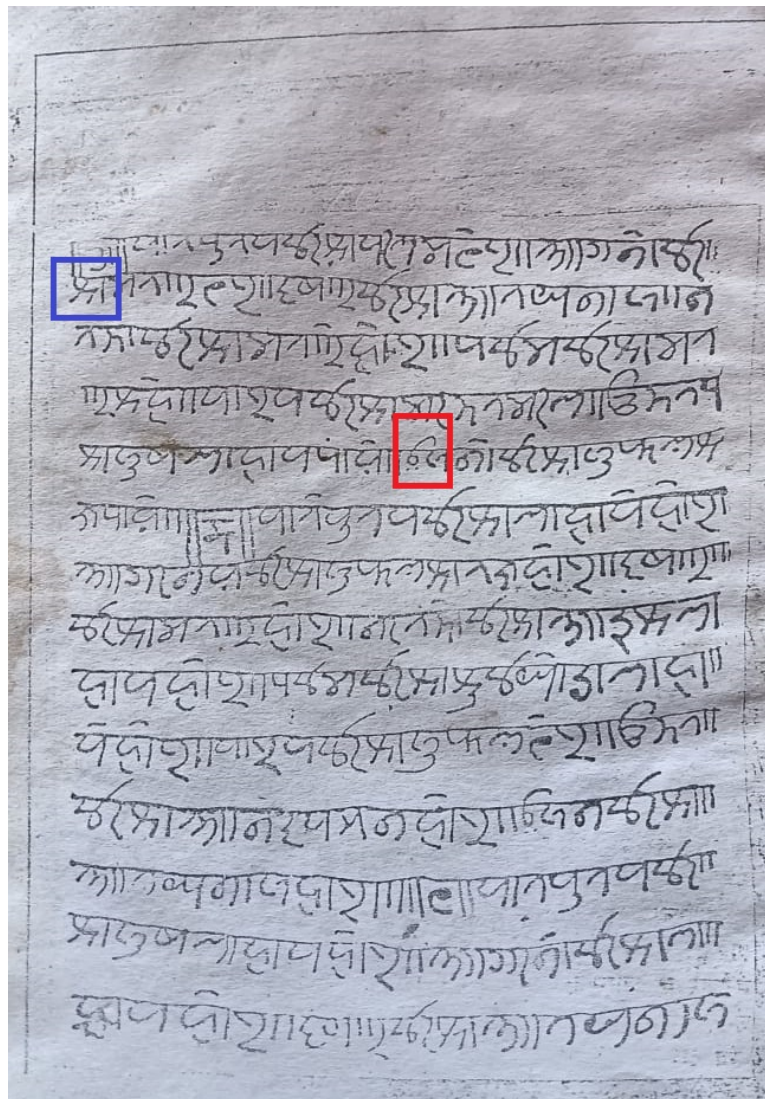


Figure 35: Alternate form २ of ३ ca (blue); usage of variant ६ for ६ for ī.

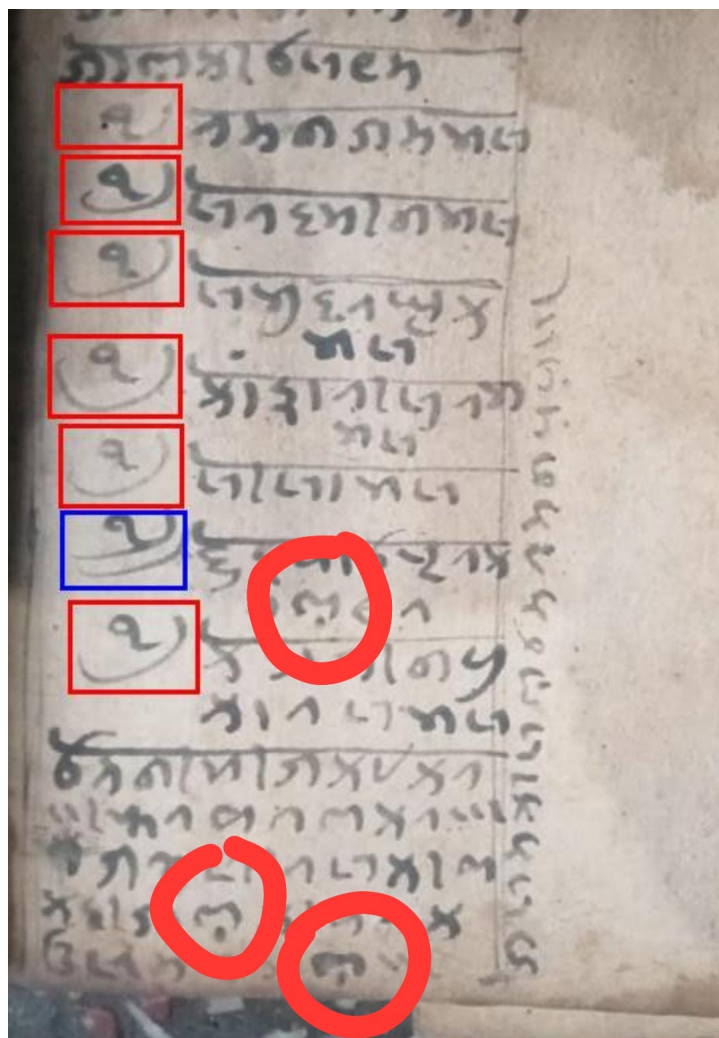


Figure 36: Usage of *ekam* + section marks; also *la* + *nukta* for writing *la*.

ੋੜਦਾ ੋਲੂ ੜੂਰਤਸਿੰਗਾ, ਤੇਰੀ ਲਾਣੀ ਓ ਬਾਤੋ ਰੇ,
 ਬੋਧਦਾ ਬੋਲੂ ੜੂਰਤਸਿੰਗਾ, ਤੇਰੀ ਲਾਣੀ ਓ ਬਾਤੋ ਰੇ,
 ਜਾਦੀ ਰੀ ਤਾਂਝੀ ਹਾਂਡਦਾ ਫਿਰਾ, ਦੇਸ ਦੇਖੋ ਨ ਰਾਤੋ ਰੇ॥
 ਰੋਹਣਾ ਬੇ ਛੋੜਾ, ਸੋਹਣਾ ਬੇ ਛੋੜਾ, ਛੋੜਾ ਤਿਧਿਏ ਠਾਠੋ ਰੇ,
 ਜਾਦੀ ਰੀ ਤਾਂਝੀ ਹਾਂਡਦਾ ਫਿਰਾ, ਦੇਸ ਦੇਖੋ ਨ ਰਾਤੋ ਰੇ॥
 ਪੜ੍ਹੋਤੇ ਰੇ ਨੋਝਏ ਮਾਂਛੇ ਬੇ ਦੇਤਾ, ਤੇਸਰਾ ਪੂਰਾ ਸਾਥੋ ਰੇ॥
 ਰਾਜੇ ਤਾਂਝੀ ਪੈਂਚੀ ਝਨੋ ਮਾਂਛੋ ਰੀ, ਬਗਾਵਤੀ ਰੀ ਬਾਤੋ ਰੇ,
 ਤਿਧਿਏ ਬੇ ਕੀ ਝੰਦਰੀ ਬੋਲੋ, ਬੋਂਦ ਸਾਰੀ ਬਾਟੋ ਰੇ॥
 ਬੁਕਾ ਨ ਦੋਬਾ ਤੋਬੇ ਬੇ ਰਾਜਾ ਰੇ, ਕਿਆ ਹੁਕਮੋਂ ਜਾਰੀ ਰੇ,
 ਹੋਸਧੋ ਡੋਏ ਖੇਲੇਧੋ ਡੋਏ, ਜੋਧਲ ਕੇ ਬਾਰੀ-ਬਾਰੀ ਰੇ॥
 ਛੇਡੂ ਬੇ ਛੋੜੀ ਛੋਟੂ ਬੇ ਛੋੜੀ, ਧੋਰੋ ਛੋੜੀ ਤਿਨੇ ਬਾਰੋ ਰੇ,
 ਉਮ੍ਹਰੇ ਭੋਰੇ ਖੇ ਜੋਧਲੋ ਲੋਏ, ਮਾਨੇ ਨ ਤਿਨੇਂ ਬੋਲੋ ਹਾਰੋ ਰੇ॥
 ਸਨਾਨੀ ਥੇਏ ਸੇ ਜਾਦੀ ਰੇ ਬੋਲੋ, ਓਓ ਧਿੰਦਰਾ ਭੋਲਾ ਰੇ,
 ਚਿਤੇ ਨੀ ਰਾਖੋ ਜੂ ਆਜ ਤਿਨੋਂ, ਬੋਧਰੇ ਦੇਸ਼ੋ ਰਾ ਹੋਲਾ ਰੇ॥
 ਬੋਧਦਾ ਬੋਲੂ ੜੂਰਤਸਿੰਗਾ, ਤੇਰੀ ਲਾਣੀ ਓ ਬਾਤੋ ਰੇ,
 ਜਾਦੀ ਰੀ ਤਾਂਝੀ ਹਾਂਡਦਾ ਫਿਰਾ, ਦੇਸ ਦੇਖੋ ਨ ਰਾਤੋ ਰੇ॥

Figure 39: A poem by Sherjung Chauhan typed in both Sirmauri and Devanagari using digital fonts.

**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 Sirmauri in Unicode
2. Requester's name:	Anshuman Pandey <pandey@umich.edu> Biswajit Mandal <biswajitmandal.bm90@gmail.com>
3. Requester type (Member body/Liaison/Individual contribution):	Expert contribution
4. Submission date:	2025-03-28
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:	Sirmauri
b. The proposal is for addition of character(s) to an existing block:	
Name of the existing block:	
2. Number of characters in proposal:	64
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? If YES explain	No
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom?	Yes
<p>Nikhil Bhardwaj Sherjung Chauhan Bhupal Singh Caouhan The family of Natiram Negi Gangaram Pabuch Yatin Pandit Dr. Om Prakash Sharma Dr. Dilip Singh Tilkan The family of Har Singh Tilkan The family of Ved Singh Tilkan</p>	
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? Reference:	Yes See text of proposal
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare See text of proposal
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes Used by Sirmauri language speakers for ritual and personal use
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	No
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? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	No
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? If YES, is a rationale for its inclusion provided? If YES, reference:	No
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference:	No No
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	No
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	Yes Virama
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No