

Proposal to encode Sirmauri in Unicode

Anshuman Pandey

Biswajit Mandal

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

biswajitmandal.bm90@gmail.com

October 25, 2024

1 Introduction

This is a proposal to encode the Sirmauri script in the Unicode standard. It replaces the following document:

- L2/18-085: “Preliminary proposal to encode Sirmauri in Unicode” (Pandey)

The encoding of Sirmauri in Unicode was hindered by limited access to users and sources. Pandey introduced the script to the Unicode Technical Committee (UTC) in L2/18-085, but relied solely upon charts and specimens published in *The Linguistic Survey of India* (1916b) by George A. Grierson. He found it impractical to continue with a proposal based upon a single source. Proper research on Sirmauri would require direct interaction with users of the script in Himachal Pradesh. A few years ago, Mandal independently began to seek out users of Sirmauri. He was able to connect with active users and experts of the script. These contacts provided Mandal with information about Sirmauri and gave him access to manuscripts.

Although Pandey and Mandal were separately working to encode Sirmauri in Unicode, they have merged their efforts in this joint proposal in order to provide a singular voice for the script community. The character repertoire has been validated by analysis of manuscripts, such as accounting (*huṇḍī*), astrological (*sāñcā*), calendrical (*pañcāṅgam*), and other ritual texts, as well as personal diaries and royal orders. These materials are more representative of conventional Sirmauri than the specimens in Grierson. The letterforms used in this proposal are based upon these sources, and were drawn by Pandey and vetted by Mandal. Most importantly, this proposal has been reviewed and approved by experts and native users, who are acknowledged in § 9.

2 Background

Sirmauri (𑂦𑂧𑂱𑂰𑂱𑂲) is a Brahmi-based script that is written from left to right. It was used as the official script in Sirmaur (also spelled as ‘Sirmour’ and ‘Sirmur’ in English sources), a former 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 Uttarakand. It was used for writing Sirmauri (ISO 639-3: srx), a Western Pahari language, and is very similar to the script used for writing Jaunsari (ISO 639-3: jns), spoken in Jaunsar-Bawar. The script of Jaunsar, identified by Grierson as ‘Jaunsari’, is considered a variant of Sirmauri, and should be unified in Unicode with ‘Sirmauri’.

The origins of this script are unknown. It bears similarities to Kaithi (U+11080), Mahajani (U+11150), and Chambeali — the ‘standard’ form of Takri (U+11680), which is encoded in Unicode — and mercantile scripts known as *muḍiā* ‘turned, rounded’, or possibly ‘shorn, tonsured’, referring to the absence of the head-stroke. Local legends state that on account of marital alliances between the royal families of Sirmaur and Rajasthan, the families from the latter brought their accountants, clerks, and administrators, who, in turn, brought the Mahajani script used in their native lands. When this Mahajani was combined with the local form of Takri, a distinctive Sirmauri script emerged.

Sirmauri was used as an official script in royal gazettes for recording land records (see fig. 6–12), revenue records, accounts, etc. According to the *Sirmur State Gazetteer, 1904*, there was a school in Kufarmand of Pachhad, where the Sirmauri and Hindi languages were taught in the Sirmauri and Devanagari scripts, respectively. But, from the early through mid 20th century, the growing influence of Urdu and Persian, then eventually Hindi, transformed the orthographic culture of Sirmaur.

Knowledge about the Sirmauri script has been maintained in Western scholarly literature for more than a century, when George A. Grierson published specimens of it in *The Linguistic Survey of India*, vol. IX (1916b). However, beyond the few samples and charts provided by Grierson, there is very little coverage of Sirmauri in the secondary literature. The letters of the Jaunsari variety appears in a chart in *Die Schrift* (1969) by Hans Jensen, but, those forms are based on those in Grierson.

Apart from the brief descriptions in Grierson and Jensen, there is a paucity of knowledge about Sirmauri in both Indian and Western scholarly literature, and an absence of published primary sources. This is due to three reasons. Firstly, as a popular and current hand, the script was used for routine activities such as letters and accounting, which may not have been preserved after the documents were no longer needed. Secondly, Sirmauri was also used for recording religious texts, horoscopes, and other sacred documents. These manuscripts were kept in private family archives and rarely provided to outsiders. Thirdly, conventional usage of the script was supplanted by the promotion of Devanagari after the 1950s. Thus, while Sirmauri was maintained within private domains, its usage diminished in the public sphere, as did the availability of materials in the script.

Today, Sirmauri is an endangered, but living script. Although no longer used widely by speakers of Sirmauri, it is read and written by those who have maintained knowledge of the script through their cultural practices and the traditions of their families. Sirmauri is actively used by these present-day stewards of the script, used for religious, ritual, and astrological practices, known as *sāñcā vidyā*, which are maintained in manuscripts known as *sāñcā grantha* (see fig. 13–17).

3 The Script

There are 4 basic vowel letters. The letter *i* represents both *i* and *ī*; and *u* represents both *u* and *ū*. The letters *a*, *i*, and *u* also function as vowel carriers for representing independent vowels.

3)	6	6	ξ
<i>a</i>	<i>i, ī</i>	<i>u, ū</i>	<i>e</i>

These letters are accompanied by 7 dependent vowel signs. Length distinction for *-i*, *-ī* and *-u*, *-ū* is not maintained in the orthography. Like other Indic scripts, there is no dependent form *-a*; the sound is inherent in each consonant letter.

◌̄)	◌̄᳚	◌̄᳚	◌̄	◌̄̄	◌̄)	◌̄̄)
-ā	-ī, -ī̄	-u, -ū	-e	-ai	-o	-au

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

3᳚)	3᳚̄)	3᳚)	3᳚̄̄)
3᳚, ◌̄)	3᳚, ◌̄̄)	3᳚, ◌̄)	3᳚, ◌̄̄)
a, -ā	a, -ai	a, -o	a, -au
ā	ai	o	au

Distinctive, independent forms of the vowels ī, ū, e are written using vowel letters and combining vowel signs (see fig. 19, 20). Both the independent letter and the vowel-carrier form are commonly used for writing e.

᳚	᳚᳚	᳚᳚	3᳚)
᳚, ◌̄᳚	᳚, ◌̄᳚	᳚, ◌̄᳚	3᳚, ◌̄)
i, -u	i, -ī	u, -u	a, -e
ī	ī	ū	e

The variant form ᳚ of ī is attested (see fig. 35). The usage of two dots to distinguish i and ī is a common pattern in related scripts; compare Chambali ᳚ i and ᳚ ī; Dogra ᳚ i and ᳚ ī.

There are 28 consonant letters, arranged according to the typical Brahmi paradigm. The script lacks distinctive letters for representing the nasals ṅa and ña, the sibilants śa and sa, and the semi-vowel va.

𑂀	𑂁	𑂂)	𑂃	—	𑂄	𑂅	𑂆)	𑂇	—
ka	kha	ga	gha	ṅa	ca	cha	ja	jha	ñā
𑂈	𑂉	𑂊	𑂋	𑂌)	𑂍	𑂎	𑂏	𑂐	𑂑
ṭa	ṭha	ḍa	ḍha	ṇa	ta	tha	da	dha	na
𑂒	𑂓	𑂔)	𑂕	𑂖)	𑂗	𑂘)	𑂙	—	
pa	pha	ba	bha	ma	ya	ra	la	va	
—	—	𑂚)	𑂛)						
śa	ṣa	sa	ha						

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

ε	२	ॐ	४	॥	५, ५, २	६, ७, ८, ९
<i>e</i>	<i>ca</i>	<i>ṭa</i>	<i>ḍha</i>	<i>ta</i>	<i>ya</i>	<i>ha</i>

Additionally, some documents show Devangari-like forms for letters, such as *tha*, *pa*, *pha* (see fig. 34–35). Usage of these is likely due to external orthographic influence.

Consonant clusters are typically represented by writing the sign *halanta* beneath the consonant whose vowel is silenced. There are no atomic conjuncts. Although rare, usage of half or contextual forms of letters is attested:

४	ॐ
half <i>ma</i>	<i>ra-kāra</i>

- The ४ half *ma* occurs in the cluster ४० *mma* in a single document (see fig. 34)
- The *ra-kāra* occurs in a few documents (see fig. 38). It is a contextual form of ॐ *ra* used in a cluster-final position, eg. ५ *kra*, २ *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.

The following modifier signs are used:

◌̣	◌̣:	◌̣	◌̣
<i>anusvāra</i>	<i>visarga</i>	<i>halanta</i>	<i>nukta</i>

- *anusvāra* indicates vowel nasalization or is a proxy for a nasal consonant, eg. ५ *kaṃ*
- *visarga* indicates post-vocalic aspiration and occurs in Sanskrit contexts, eg. ००: *namaḥ* (see fig. 37)
- *halanta* silences the inherent vowel *a* of a consonant letter, and is used for representing consonant clusters, eg. ५ *k* (see fig. 21, 22), ५५ *kka*
- *nukta* is used for extending the script to represent sounds for which distinctive letters do not exist:

४	४	५	२	०	०
<i>ḍa, nukta</i>	<i>ḍha, nukta</i>	<i>ṣa, nukta</i>	<i>pa, nukta</i>	<i>ba, nukta</i>	<i>la, nukta</i>
<i>ṛa</i>	<i>ṛha</i>	<i>śa</i>	<i>va</i>	<i>va</i>	<i>ḷa</i>

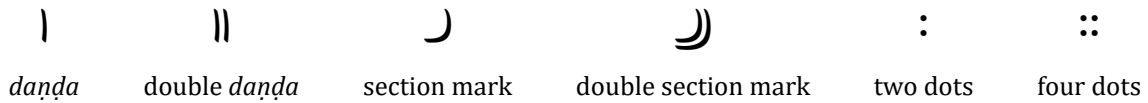
There is a full set of decimal digits (see fig. 24, 25):

०	१	२	३	४	५	६	७	८	९
<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>

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:



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



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).

4 Comparison with Grierson

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	3, 6	3)	3)
3)	3))	6, ५,	6, 6, 6?	3	3	५, 3)	3)	3))	3))
<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))	५	—	५	५	५	५	—
५	५)	५	—	५	५	५	५	—
<i>ka</i>	<i>kha</i>	<i>ga</i>	<i>gha</i>	<i>ṅa</i>	<i>ca</i>	<i>cha</i>	<i>ja</i>	<i>jha</i>	<i>ṅa</i>
८	6	5	५)))	५	५	ε	ε	५
५	८	5	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>

𑂒	𑂓	𑂔	𑂕	𑂖	𑂗	𑂘	𑂙	𑂚
𑂒	𑂓	𑂔	𑂕	𑂖	𑂗	𑂘	𑂙	(𑂚)
pa	pha	ba	bha	ma	ya	ra	la	va
𑂛	𑂜	𑂝	𑂞					
(𑂛)	(𑂜)	𑂝	𑂞					
śa	ṣa	sa	ha					

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

- *a* : Grierson's 𑂒 *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 𑂒 form is the basis for 𑂒 *ā*, 𑂒 *e*, 𑂒 *ai*, 𑂒 *o*, 𑂒 *au*. But, despite the consistency of this form across the independent vowel letters, the source for 𑂒 is unknown. It could have been an idiosyncrasy of the scribe of Grierson's text. Experts confirm that 𑂒 is the correct form for *a*. Accordingly, Grierson's forms would be analyzed as follows: 𑂒 *ā* = *a*; 𑂒 *e* and 𑂒 *ai* are invalid forms; 𑂒 *o* = *e*; 𑂒 *au* = *ai*. Nonetheless, his recording of the Jaunsari form 𑂒) as *ā* is correct, as this form has the appropriate 𑂒 *a* as the base combined with the conventional vowel sign 𑂏) for *ā*.
- *ā* : The forms 𑂒) and 𑂒)) are given for Jaunsari *ā*. Of these, 𑂒)) is the correct form for *ā*. It is possible that 𑂒) *a* was used for *ā* in some contexts, just as 𑂏) is used for both *i* and *ī*, and 𑂏) for *u* and *ū*.
- *i*, *ī* : The 𑂏) is the paleographical form for *i*, also used for *ī*. The 𑂗) is the letter *e*, which along with 𑂗) *ya*, is used as a proxy for *i*.
- *e* : The source for 𑂒) is unknown. It is *a*, functioning as a vowel carrier, with the vowel sign *e*. Similarly, 𑂗) is 𑂗) *ya* bearing the vowel sign *e*. The 𑂗) is the letter for *e*. The correct vowel carrier form of *e* is 𑂒).
- *ai* : The Jaunsari 𑂏) *ai* is the letter 𑂏) *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 𑂒) *o* and 𑂒) *au* is unknown. Apart from Grierson, these forms are not attested in other materials.
- *ṭa* : The 𑂕) is a form heavily influenced by Devanagari *ṭa*. It is more commonly used in Sirmauri for *ṭha*, while the representative form for *ṭa* is 𑂕).
- *bha* : The 𑂕) is a reversed form of conventional 𑂕) *bha*.
- *va* : The 𑂚) is not a distinct letter, but the letter *ba* written with a *nukta*.
- *śa* : The 𑂛) is not a distinct letter, but the letter *sa* written with a *nukta*.
- *ṣa* : The 𑂜) shown for *ṣa* is simply a duplication of *kha*, which is commonly used for writing the sibilant.

5 Encoding Model

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

- The ‘Sirmauri’ block unifies the Jaunsari variety. Variant forms are to be handled using fonts.
- Usage of a new *virāma* model, in which the dual functions of the sign ◌̣ *virāma*, as found in the encodings for other Indic scripts, have been separated into two characters:

◌̣	SIRMAURI SIGN HALANTA	visible vowel silencer; no conjunct control functions
◌̤	SIRMAURI SIGN VIRAMA	controls conjunct formation

These signs would be used as follows:

<i>mma</i>	᳚᳚	᳚ <i>ma</i> , ◌̣ <i>halanta</i> , ᳚ <i>ma</i>
<i>mma</i>	᳚᳚	᳚ <i>ma</i> , ◌̤ <i>virāma</i> , ᳚ <i>ma</i>

- The contextual form ◌̣ *ra-kāra* of ᳚ *ra* is encoded as a separate character: SIRMAURI SIGN RA-KARA. This simplifies the representation of clusters by avoiding usage of *virāma*.
- There are two approaches for supporting the variety of independent vowel letters used in the script. The second approach has been selected.

1. Encode only the distinctive vowel letters and define them as vowel carriers:

<i>a</i>	᳚	SIRMAURI LETTER A
<i>i</i>	᳚	SIRMAURI LETTER I
<i>u</i>	᳚	SIRMAURI LETTER U
<i>e</i>	᳚	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>	᳚	᳚ SIRMAURI LETTER A , ◌̣ SIRMAURI VOWEL SIGN AA
<i>e</i>	᳚	᳚ SIRMAURI LETTER A , ◌̣ SIRMAURI VOWEL SIGN E
<i>ai</i>	᳚	᳚ SIRMAURI LETTER A , ◌̣ SIRMAURI VOWEL SIGN AI
<i>o</i>	᳚	᳚ SIRMAURI LETTER A , ◌̣ SIRMAURI VOWEL SIGN O
<i>au</i>	᳚	᳚ SIRMAURI LETTER A , ◌̣ SIRMAURI VOWEL SIGN AU

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

\bar{i}	᳚	᳚ SIRMAURI LETTER I , ◌᳚ SIRMAURI VOWEL SIGN U
\bar{i}	᳚᳚	᳚ SIRMAURI LETTER I , ◌᳚᳚ SIRMAURI VOWEL SIGN II
\bar{u}	᳛	᳛ SIRMAURI LETTER U , ◌᳛ SIRMAURI VOWEL SIGN U

2. Encode all conventional independent vowel forms as characters, following the pattern of other northern Indic scripts:

a	᳚᳚	SIRMAURI LETTER A
\bar{a}	᳚᳚᳚	SIRMAURI LETTER AA
i	᳚	SIRMAURI LETTER I
\bar{i}	᳚᳚	SIRMAURI LETTER II
u	᳛	SIRMAURI LETTER U
\bar{u}	᳛᳛	SIRMAURI LETTER UU
e	᳚᳚	SIRMAURI LETTER E
e	᳚᳚᳚	SIRMAURI LETTER ALTERNATE E
ai	᳚᳚᳚	SIRMAURI LETTER AI
o	᳚᳚᳚	SIRMAURI LETTER O
au	᳚᳚᳚᳚	SIRMAURI LETTER AU

Accordingly, ᳚᳚᳚ \bar{i} and ᳚᳚ \bar{i} would be treated as a glyphic variants of ᳚᳚ SIRMAURI LETTER II. If there is evidence of their concurrent usage with ᳚᳚, then they may be encoded as alternate characters.

- The ◌᳚᳚ is used for both $-i$ and $-\bar{i}$, but it is palaeographically- \bar{i} . For this reason, it has been assigned the name SIRMAURI VOWEL SIGN II.
- Space is reserved for *SIRMAURI VOWEL SIGN I and *SIRMAURI VOWEL SIGN UU in the event that distinctive signs are identified in the future. This is driven by an update for Khojki, which also has a merger of i , \bar{i} and u , \bar{u} . The Khojki ᳚᳚᳚ is palaeographical \bar{i} , but it was encoded as U+11202 KHOJKI LETTER I (see L2/11-021 = Pandey 2011a). Recently, a distinctive independent form ᳚᳚᳚ for i was identified. But as ᳚᳚᳚ was already named KHOJKI LETTER I, the ᳚᳚᳚ was encoded as U+11240 KHOJKI LETTER SHORT I (see L2/21-104 = Pandey 2021). Using character names based on palaeography avoids such complications.
- Space has also been reserved for the following letters in their usual positions in the consonant order: *SIRMAURI LETTER NGA, *SIRMAURI LETTER NYA, *SIRMAURI LETTER SHA, *SIRMAURI LETTER SSA.
- 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.
- The $daṇḍā$ and double $daṇḍā$ are to be unified with the corresponding characters in Devanagari. Script extensions have been specified.

6 Proposed Character Repertoire

The proposed Unicode repertoire for Sirmauri consists of 65 characters:

Category	Character	Proposed Unicode Character Name
Vowel letters (11)	ॐ	SIRMAURI LETTER A
	ॐॐ	SIRMAURI LETTER AA
	ॐ	SIRMAURI LETTER I
	ॐ	SIRMAURI LETTER II
	ॐ	SIRMAURI LETTER U
	ॐ	SIRMAURI LETTER UU
	ॐ	SIRMAURI LETTER E
	ॐ	SIRMAURI LETTER ALTERNATE E
	ॐ	SIRMAURI LETTER AI
	ॐ	SIRMAURI LETTER O
	ॐ	SIRMAURI LETTER AU
Dependent vowel signs (7)	ॐ	SIRMAURI VOWEL SIGN AA
	ॐ	SIRMAURI VOWEL SIGN II
	ॐ	SIRMAURI VOWEL SIGN U
	ॐ	SIRMAURI VOWEL SIGN E
	ॐ	SIRMAURI VOWEL SIGN AI
	ॐ	SIRMAURI VOWEL SIGN O
	ॐ	SIRMAURI VOWEL SIGN AU
Consonant letters (28)	ॐ	SIRMAURI LETTER KA
	ॐ	SIRMAURI LETTER KHA
	ॐ	SIRMAURI LETTER GA
	ॐ	SIRMAURI LETTER GHA
	ॐ	SIRMAURI LETTER CA
	ॐ	SIRMAURI LETTER CHA
	ॐ	SIRMAURI LETTER JA
	ॐ	SIRMAURI LETTER JHA
	ॐ	SIRMAURI LETTER TTA
	ॐ	SIRMAURI LETTER TTHA

	ᵍ	SIRMAURI LETTER DDA
	ᵍ	SIRMAURI LETTER DDHA
	ᵍ	SIRMAURI LETTER NNA
	ᵍ	SIRMAURI LETTER TA
	ᵍ	SIRMAURI LETTER THA
	ᵍ	SIRMAURI LETTER DA
	ᵍ	SIRMAURI LETTER DHA
	ᵍ	SIRMAURI LETTER NA
	ᵍ	SIRMAURI LETTER PA
	ᵍ	SIRMAURI LETTER PHA
	ᵍ	SIRMAURI LETTER BA
	ᵍ	SIRMAURI LETTER BHA
	ᵍ	SIRMAURI LETTER MA
	ᵍ	SIRMAURI LETTER YA
	ᵍ	SIRMAURI LETTER RA
	ᵍ	SIRMAURI LETTER LA
	ᵍ	SIRMAURI LETTER SA
	ᵍ	SIRMAURI LETTER HA
Various signs (4)	ᵍ	SIRMAURI SIGN ANUSVARA
	ᵍ	SIRMAURI SIGN VISARGA
	ᵍ	SIRMAURI SIGN HALANTA
	ᵍ	SIRMAURI SIGN NUKTA
Consonant sign (1)	ᵍ	SIRMAURI SIGN RA-KARA
Auspicious sign (1)	ᵍ	SIRMAURI EKAM
Punctuation (2)	ᵍ	SIRMAURI SECTION MARK
	ᵍ	SIRMAURI DOUBLE SECTION MARK
Virama (1)	ᵍ	SIRMAURI SIGN VIRAMA
Digits (10)	ᵍ	SIRMAURI DIGIT ZERO
	ᵍ	SIRMAURI DIGIT ONE
	ᵍ	SIRMAURI DIGIT TWO

३	SIRMAURI DIGIT THREE
४	SIRMAURI DIGIT FOUR
५	SIRMAURI DIGIT FIVE
६	SIRMAURI DIGIT SIX
७	SIRMAURI DIGIT SEVEN
८	SIRMAURI DIGIT EIGHT
९	SIRMAURI DIGIT NINE

7 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 HALANTA;Mn;9;NSM;;;;;N;;;;;
 11888;SIRMAURI SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;
 11889;SIRMAURI SIGN RA-KARA;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;;;;;
 1188D;SIRMAURI SIGN VIRAMA;Mn;9;NSM;;;;;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;;;;;

7.1 Linebreaking Properties: LineBreak.txt

11850..1185E	; AL # Lo	[4] SIRMAURI LETTER A..SIRMAURI LETTER GHA
11860..11863	; AL # Lo	[4] SIRMAURI LETTER CA..SIRMAURI LETTER JHA
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 HALANTA..SIRMAURI SIGN RA-KARA
1188A      ; AL # Lo      SIRMAURI EKAM
1188B..1188C ; BA # Po      [2] SIRMAURI SECTION MARK..SIRMAURI DOUBLE SECTION MARK
1188D      ; CM # Mn      SIRMAURI SIGN VIRAMA
11890..11899 ; NU # Nd      [10] SIRMAURI DIGIT ZERO..SIRMAURI DIGIT NINE

```

7.2 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=Virama
1188C      ; Virama     # Mn      SIRMAURI SIGN VIRAMA

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

# 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 GHA
11860..11863 ; Consonant      # Lo      [4] SIRMAURI LETTER CA..SIRMAURI LETTER JHA
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 SIGN RA-KARA

```

7.3 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 HALANTA..SIRMAURI SIGN NUKTA

# 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

# Indic_Positional_Category=Bottom_And_Right
11889          ; Bottom_And_Right # Mc          SIRMAURI SIGN RA-KARA
```

7.4 Script Extensions: ScriptExtensions.txt

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

8 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. 2007a. “Proposal to Encode North Indic Number Forms in ISO/IEC 10646” (L2/07-354). <https://www.unicode.org/L2/L2007/07354-north-indic.pdf>
- . 2007b. “Proposal to Encode the Kaithi Script in ISO/IEC 10646” (L2/07-418). <https://www.unicode.org/L2/L2007/07418-kaithi.pdf>

———. 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>

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

















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

Independent vowels

11850	ੳ	SIRMAURI LETTER A
11851	ੲ	SIRMAURI LETTER AA
11852	ੴ	SIRMAURI LETTER I
11853	ੴ	SIRMAURI LETTER II
11854	ੴ	SIRMAURI LETTER U
11855	ੴ	SIRMAURI LETTER UU
11856	ੴ	SIRMAURI LETTER E
11857	ੴ	SIRMAURI LETTER ALTERNATE E
11858	ੴ	SIRMAURI LETTER AI
11859	ੴ	SIRMAURI LETTER O
1185A	ੴ	SIRMAURI LETTER AU

Consonants

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

Dependent vowel signs

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

Various signs

11885	ੴ	SIRMAURI SIGN ANUSVARA
11886	ੴ	SIRMAURI SIGN VISARGA
11887	ੴ	SIRMAURI SIGN HALANTA
		• vowel silencer
		• does not control conjunct formation
11888	ੴ	SIRMAURI SIGN NUKTA

Consonant sign

11889	ੴ	SIRMAURI SIGN RA-KARA
		• cluster-medial ra

Auspicious sign

1188A	ੴ	SIRMAURI EKAM
		= anji

Punctuation

1188B	ੴ	SIRMAURI SECTION MARK
1188C	ੴ	SIRMAURI DOUBLE SECTION MARK

Virama

1188D	ੴ	SIRMAURI SIGN VIRAMA
		• used for controlling conjunct formation

Digits

11890	ੴ	SIRMAURI DIGIT ZERO
11891	ੴ	SIRMAURI DIGIT ONE
11892	ੴ	SIRMAURI DIGIT TWO
11893	ੴ	SIRMAURI DIGIT THREE
11894	ੴ	SIRMAURI DIGIT FOUR
11895	ੴ	SIRMAURI DIGIT FIVE
11896	ੴ	SIRMAURI DIGIT SIX
11897	ੴ	SIRMAURI DIGIT SEVEN
11898	ੴ	SIRMAURI DIGIT EIGHT
11899	ੴ	SIRMAURI DIGIT NINE

The Sirmauri Alphabet.

a	3	da	ε
ā	3)	dha	ε̇
i, i & ð	6 or 5	na	∩
u & ū	⊙	pa	5
ē	3 or 5	pha	5
ai	⊙	ba	9
ō & ȝ	⊙	bha	9
au	⊙	ma	∩
ka	x	ya	x
kha	5	ra	∩
ga	∩	la	ε
gha	5	wa	ε̇
cha	∩	śa	9
chha	∩	sha	5
ja	∩	sa	9
jha	∩	ha	ε
ṭa	∩	kā	x
ṭha	6	ki, ki	x
ḍa	5	ku, kū	x
ḍha	5	ke, kē	x
ṛa	∩	kai	x
ṭa	∩	ko, kō	x
ṭha	5	kau	x

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

WESTERN PAHĀRĪ (SIRMAURĪ).

DHĀRṬHĪ DIALECT.

SPECIMEN I.

3xṛ ʃāṭhē ɖ ʃāṭhē ʃāṭhē "
 xāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē "
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē "
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē "
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē
 ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē ʃāṭhē

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

JAUNSĀRĪ (SIRMAURĪ) ALPHABET.

3	a	⌘	kau	⌘	tha
3, 3, 3	ā	⌘	kā	⌘	da
⌘	i	⌘	ka	⌘	dha
⌘	ī	⌘	kha	⌘	na
⌘	u	⌘	ga	⌘	pa
⌘	ū	⌘	gha	⌘	pha
⌘	ě, ē	—	na	⌘	ba
⌘	ai	⌘	cha	⌘	bha
⌘	ō, ō	⌘	chha	⌘	ma
⌘	au	⌘	ja	⌘	ya
·	~	⌘	jha	⌘	ra
⌘	ka	—	ña	⌘	la
⌘	kū	⌘	tā	⌘	wa
⌘	ki	⌘	tha	⌘	śa
⌘	kī	⌘	ḍa	⌘	sha
⌘	ku	⌘	ra	⌘	sa
⌘	kū	⌘	dha	⌘	ha
⌘	kě, kē	⌘	ṛha		
⌘	kai	⌘	ṇa		
⌘	kō, kō	⌘	ta		

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

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

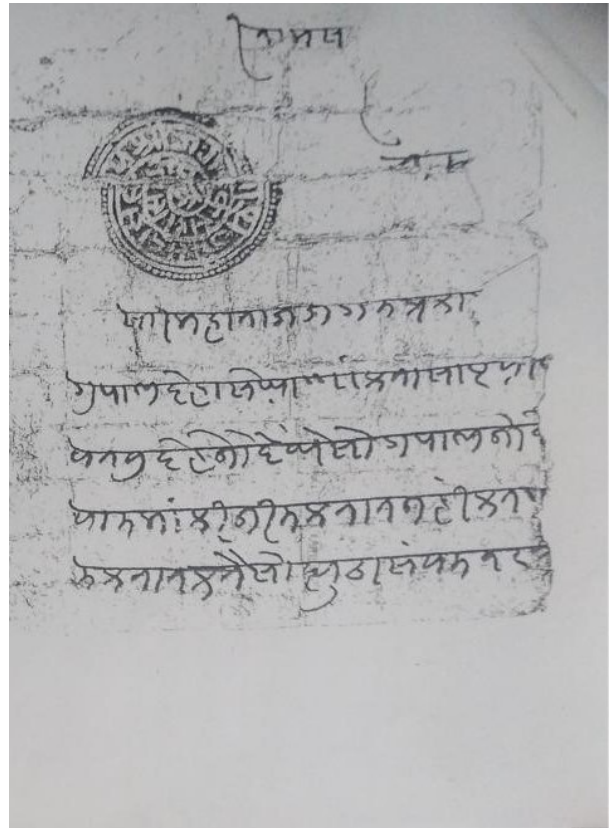
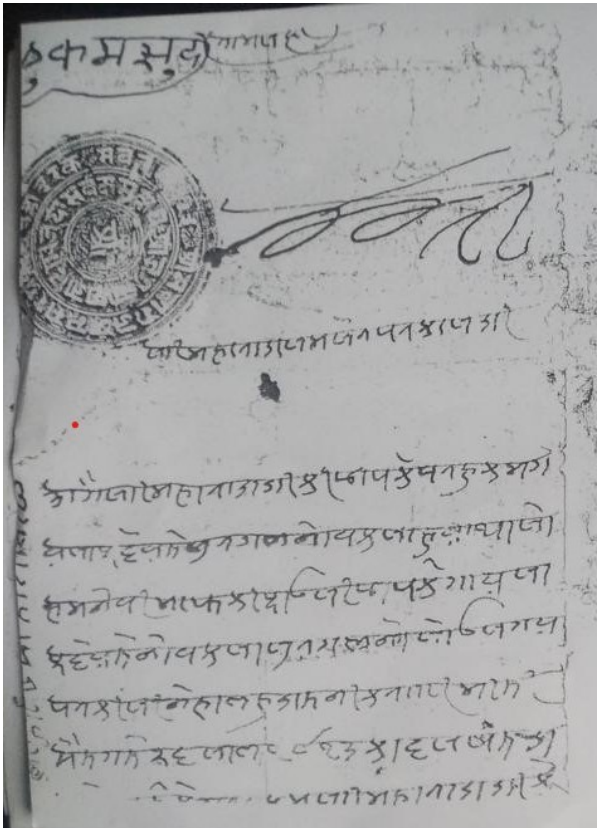


Figure 6: Documents recording land donations to the Shirgul Devta temple by *rājā* Shamsher Prakash (1770–1789) (left) and *rājā* Jagat Prakash (1856–1898) (right).

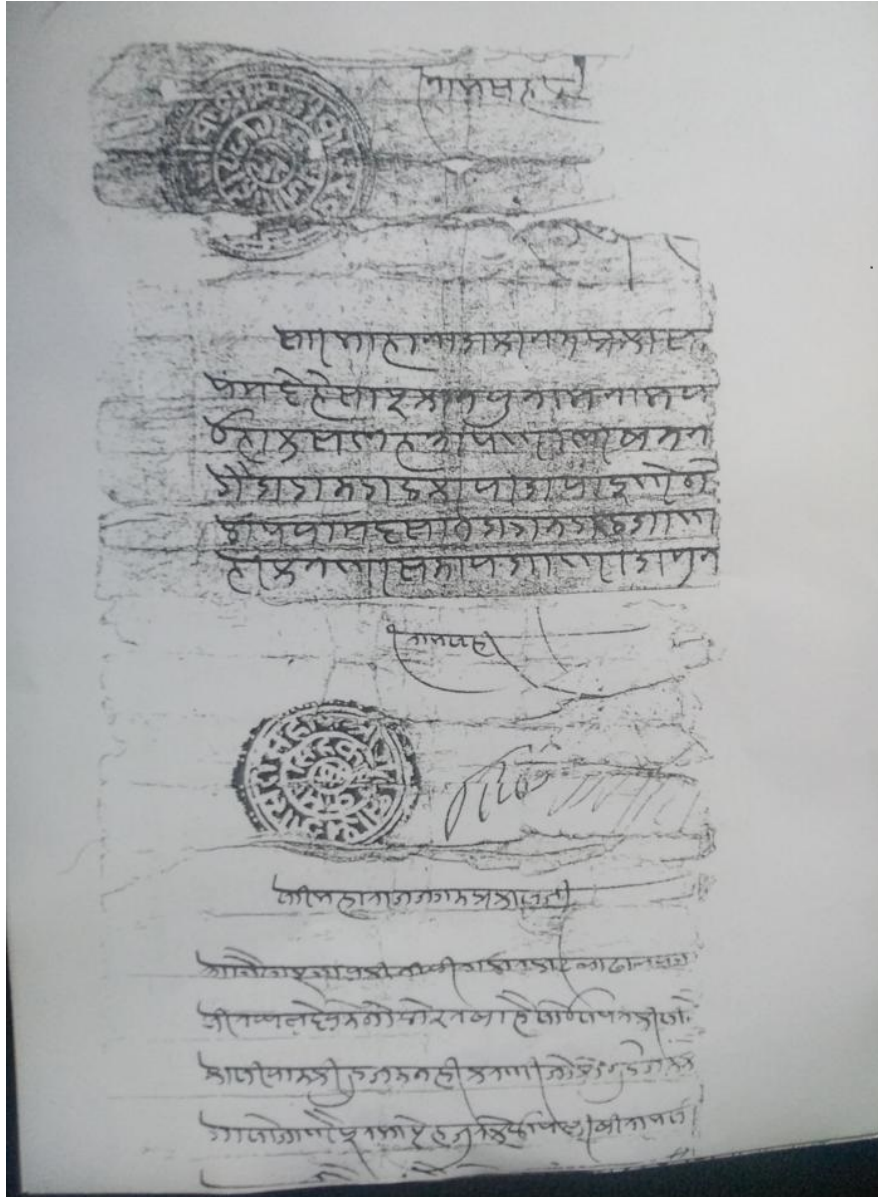


Figure 7: Record of land donation to the Shirgul Devta temple by *rājā* Karma Prakash (1804–1815).

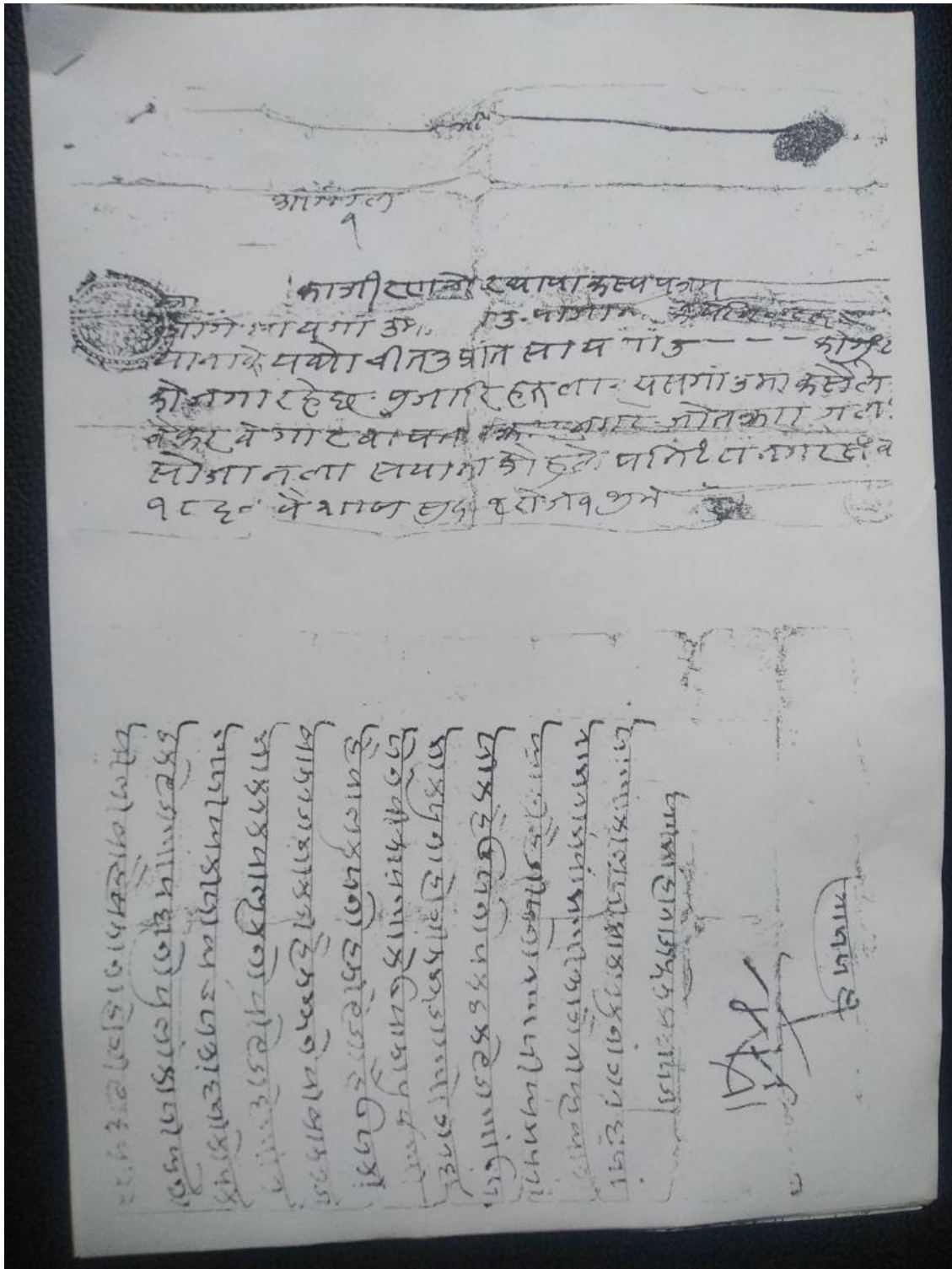


Figure 8: Record of land donation to the Shirgul Devta temple by *rājā* Fateh Prakash (1815–1850).

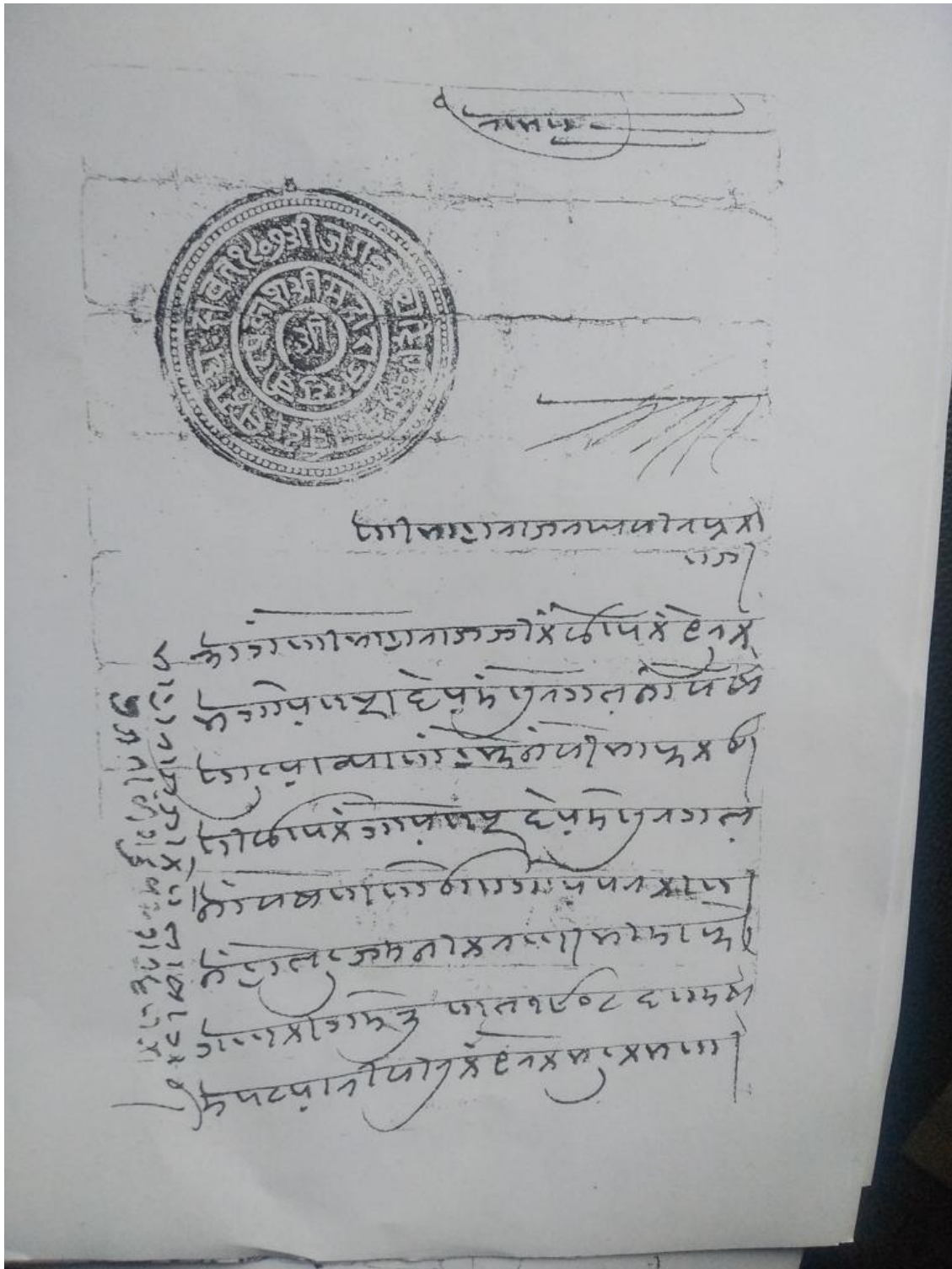


Figure 9: Record of land donation to the Shirgul Devta temple by Raghbir Prakash II (1850–56).

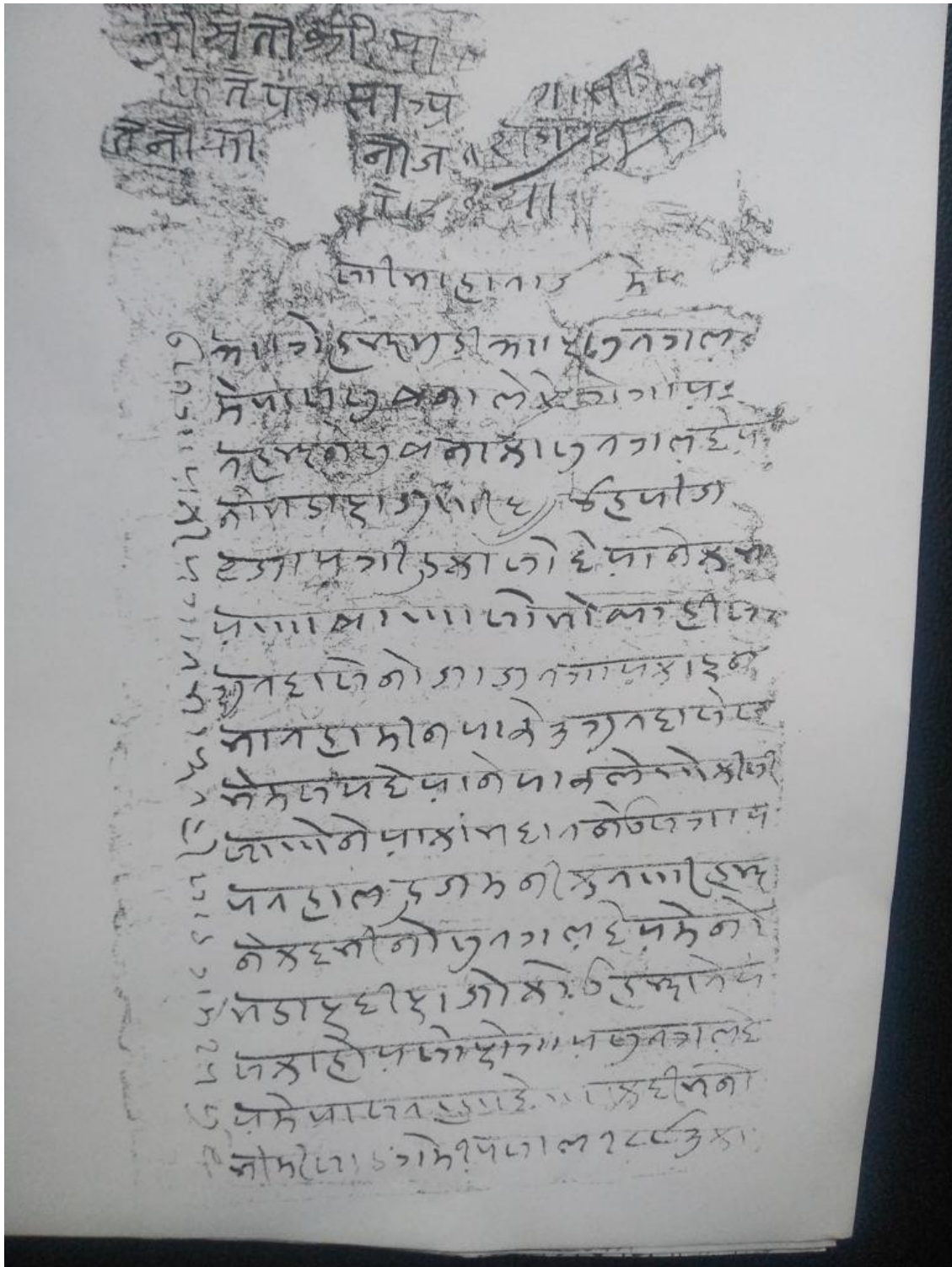


Figure 10: Record of land donation to the Shirgul Devta temple by *rājā* Raghbir Prakash II (1850–56).

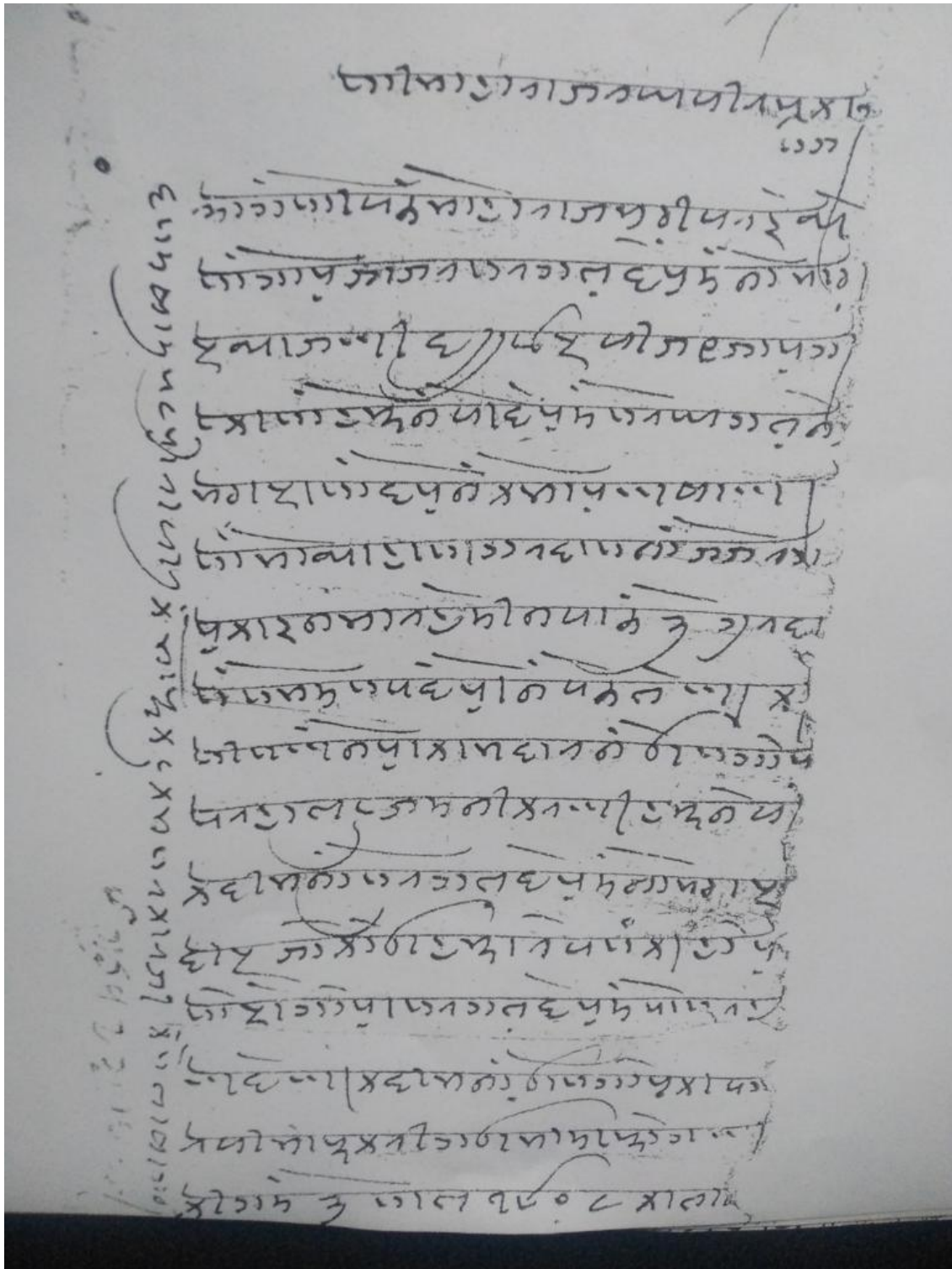


Figure 11: Record of land donation to the Shirgul Devta temple by *rājā* Raghubir Prakash II (1850–56).

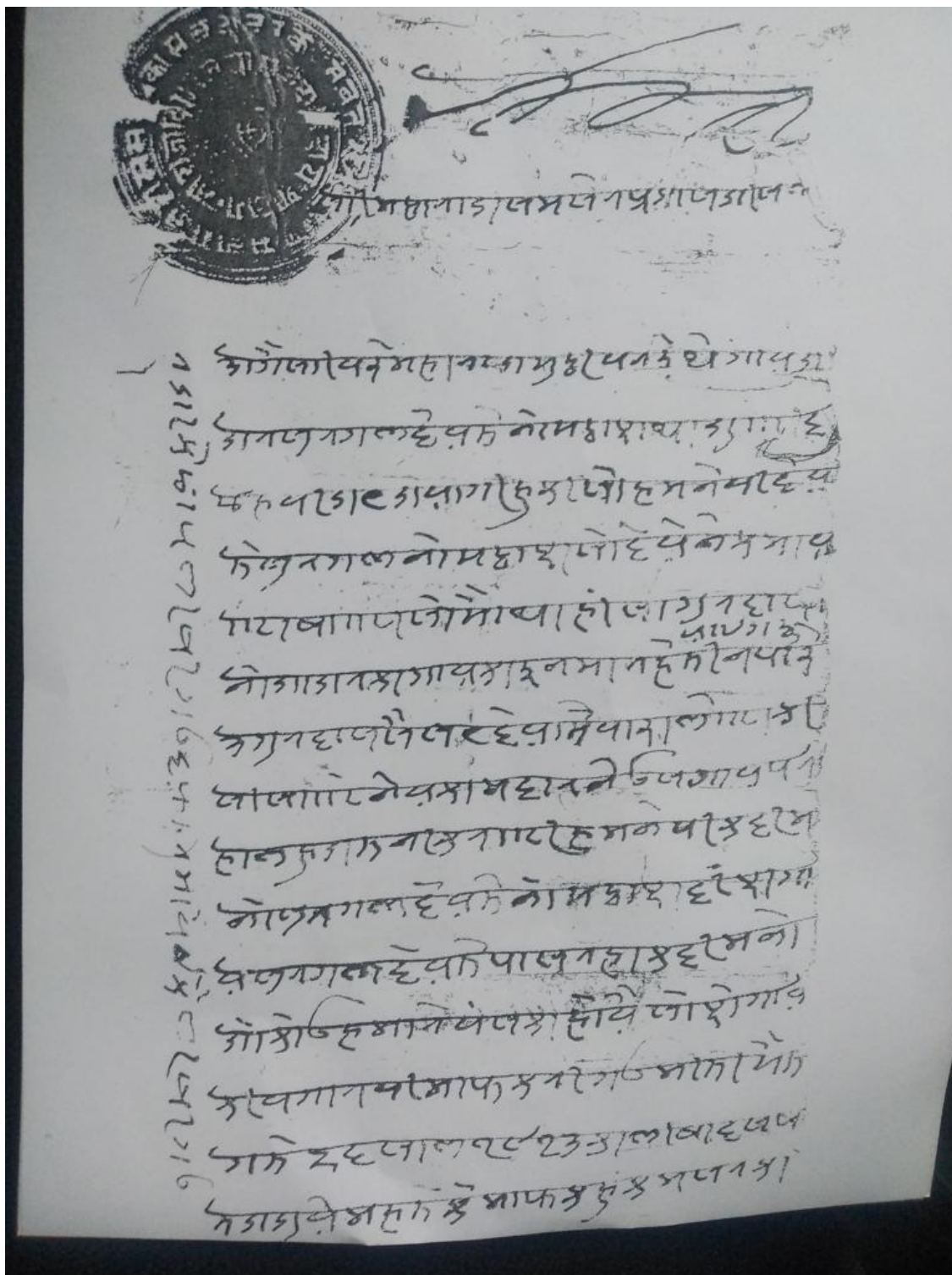


Figure 12: Record of land donation to the Shirgul Devta temple by rājā Shamsher Prakash (1856-98).

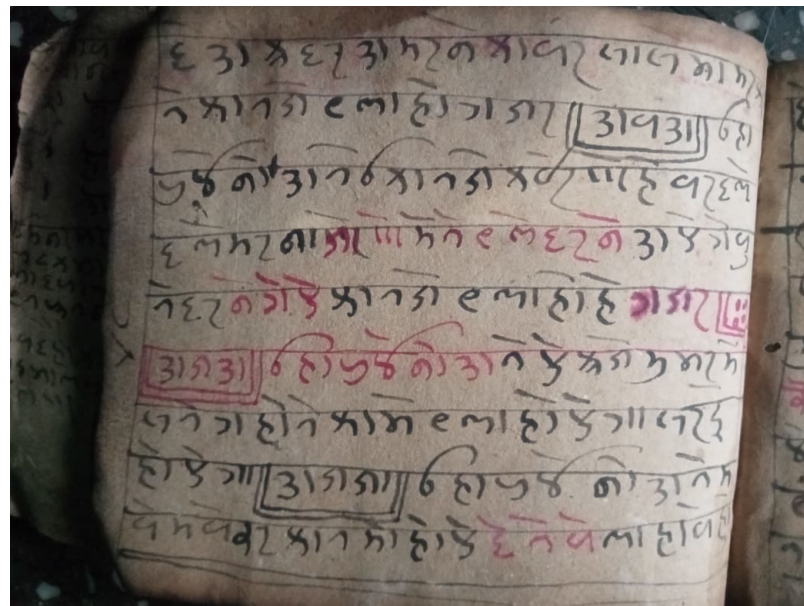
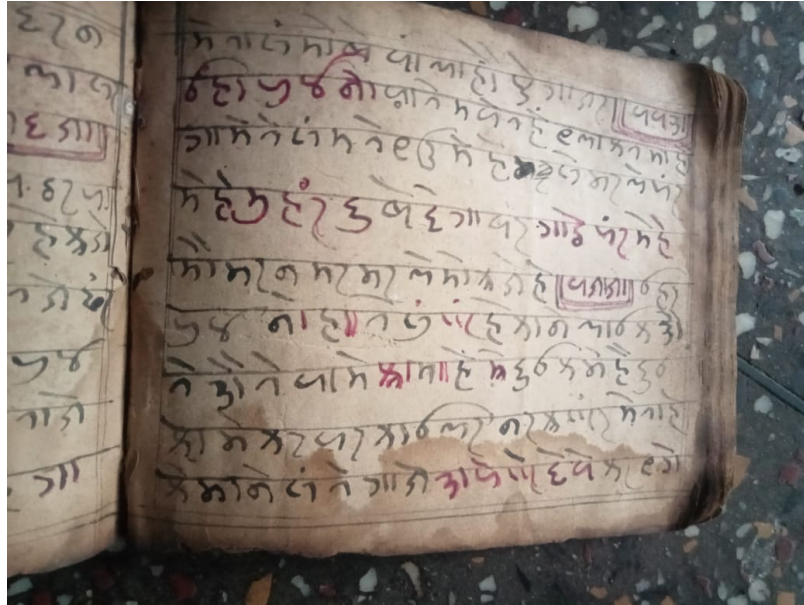


Figure 13: Sancha text written in the Jaunsari variety by Natiram Negi.

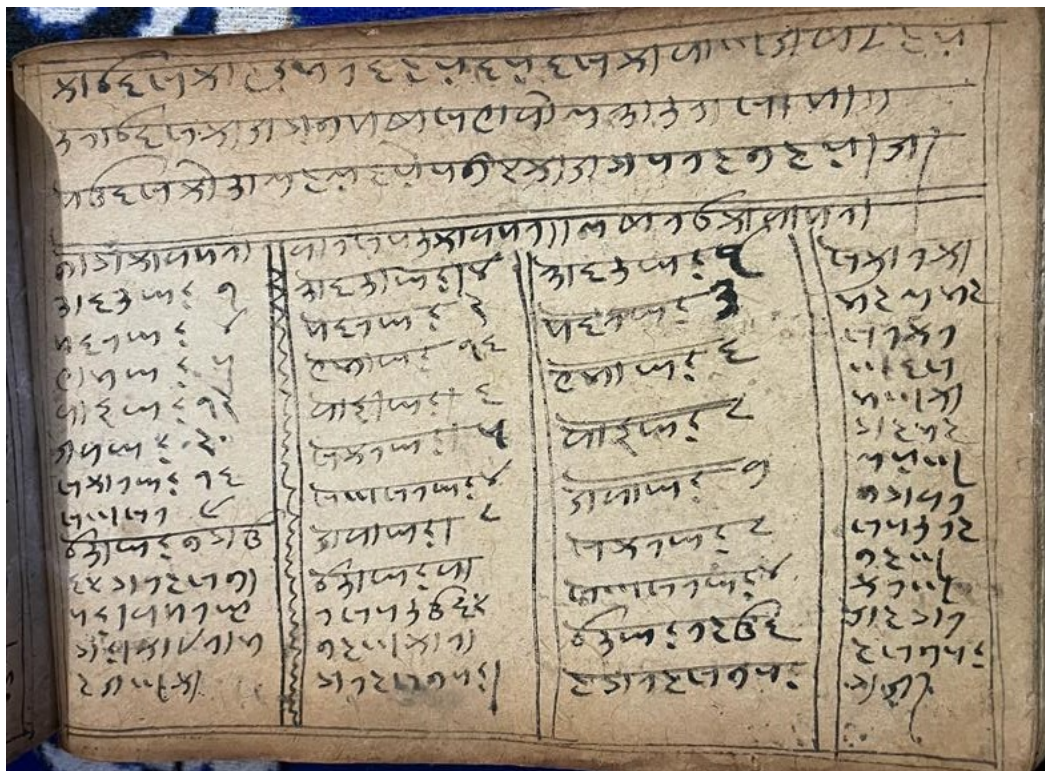
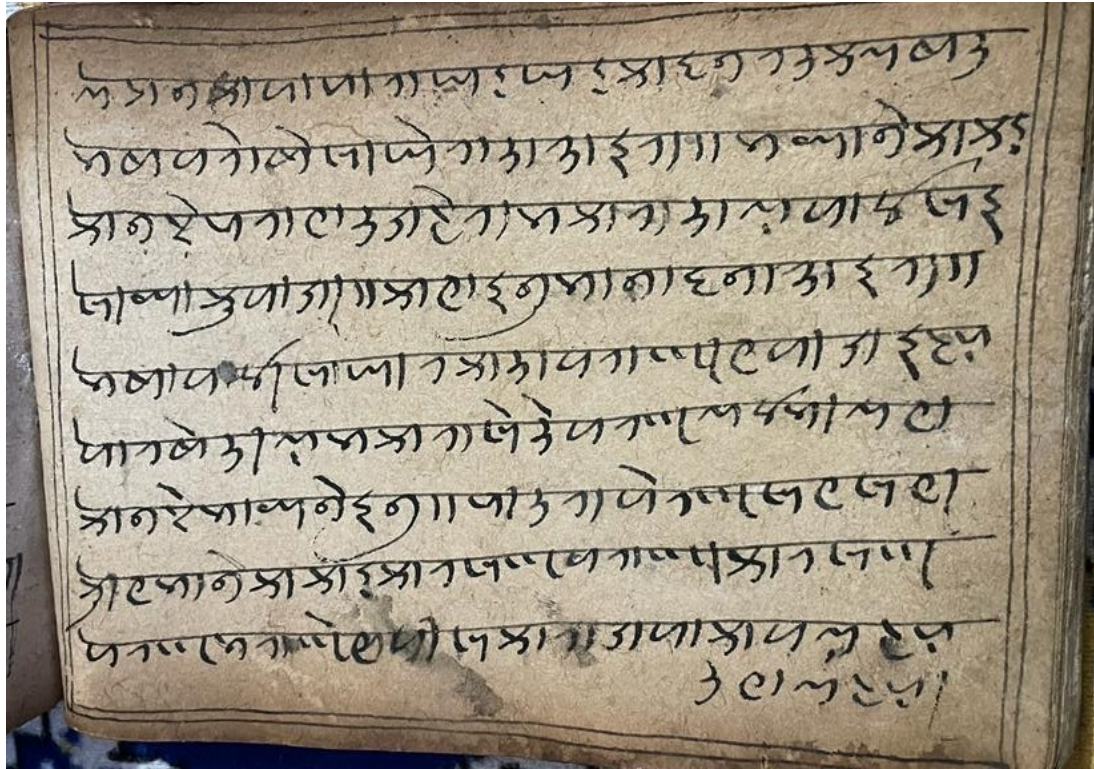


Figure 14: Sancha text written in Sirmauri by Gangaram Pabuch.

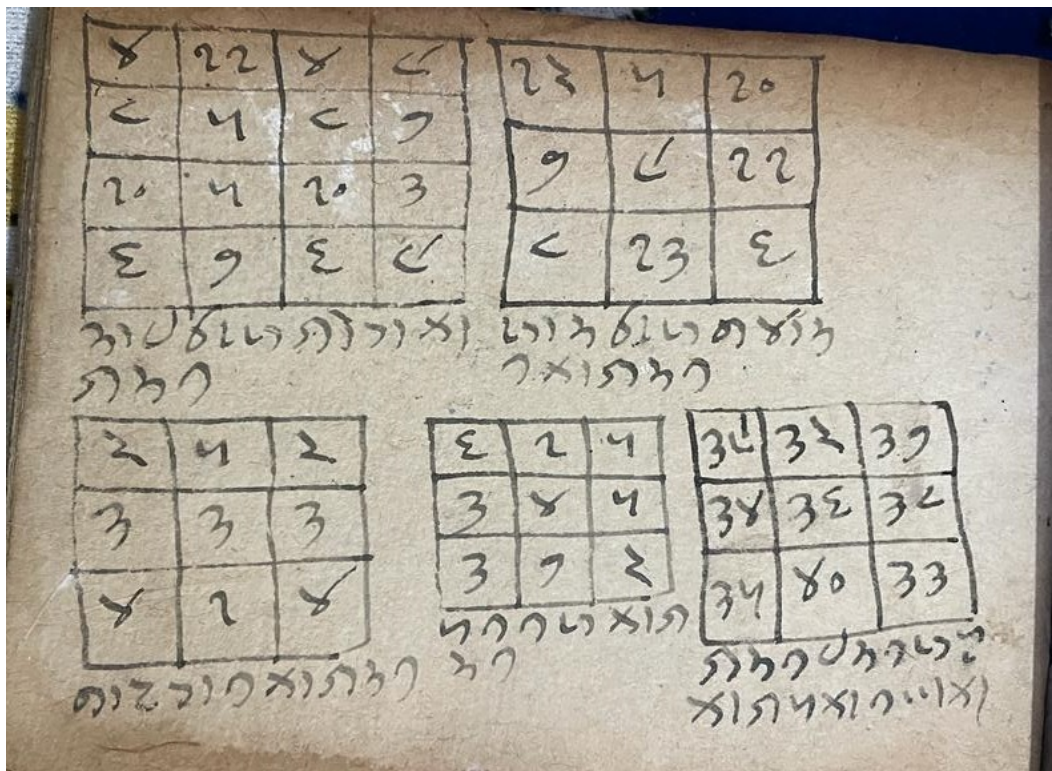
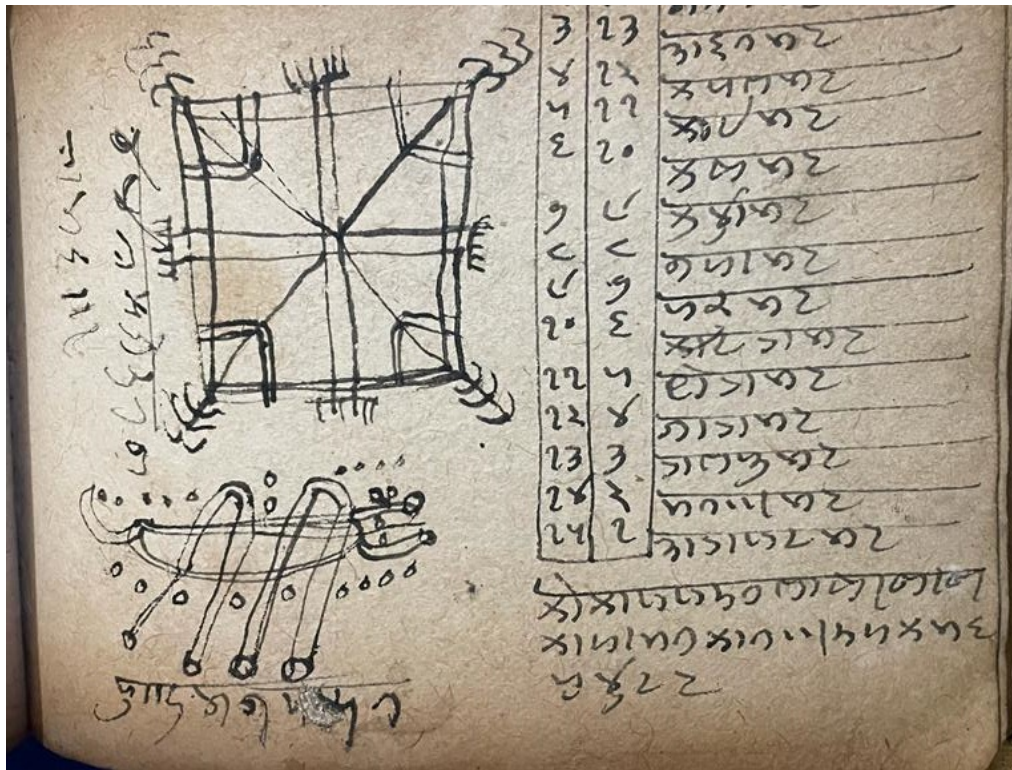


Figure 15: Sancha text written in Sirmauri by Gangaram Pabuch.

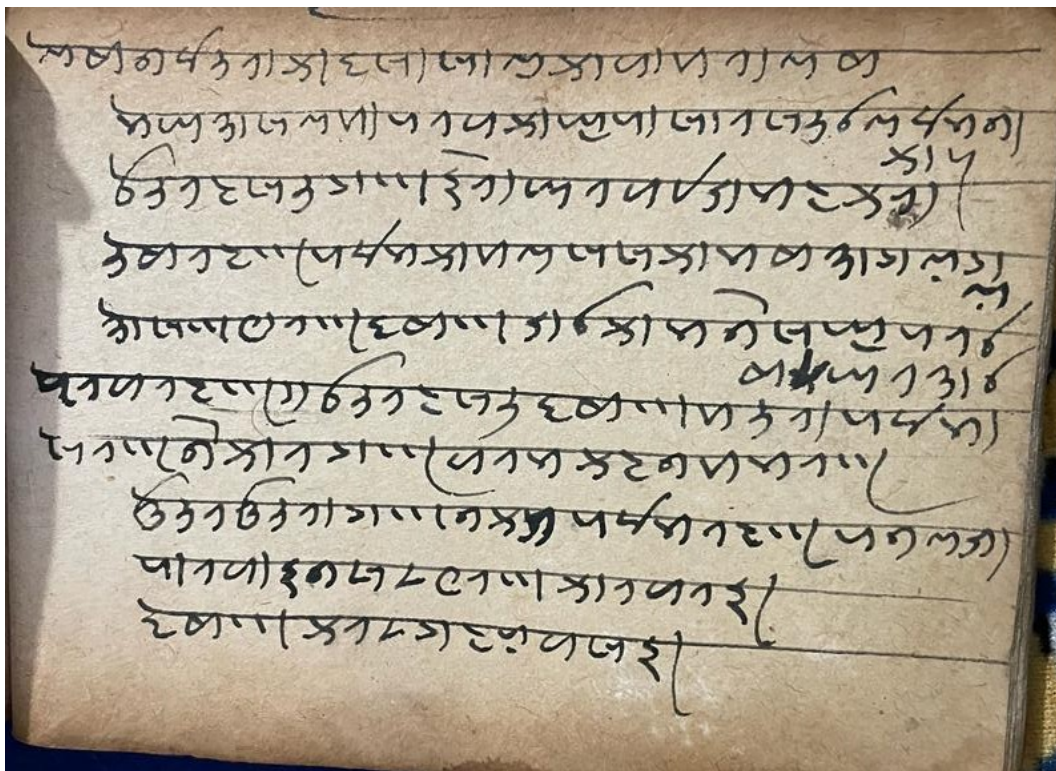


Figure 16: Sancha text written in Sirmauri by Gangaram Pabuch.

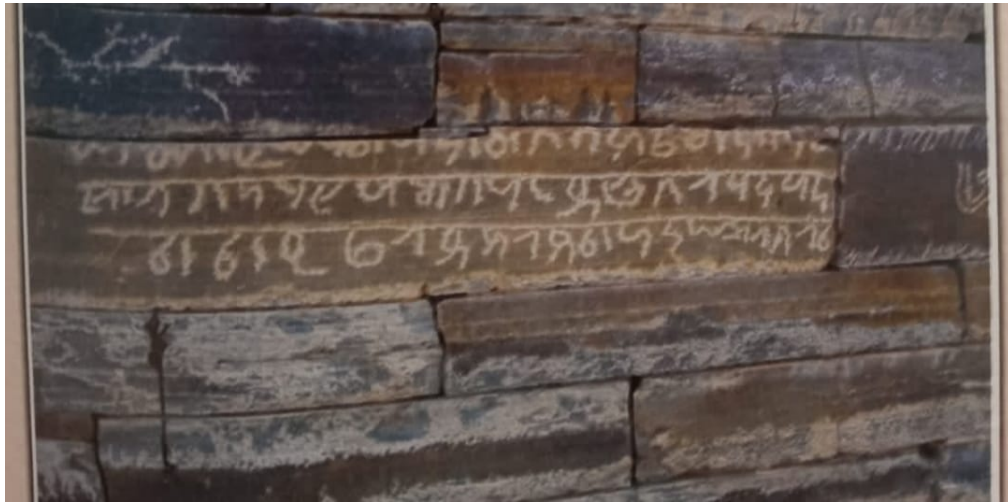


Figure 18: Inscription found on a tower (*killā*) adjacent to a house

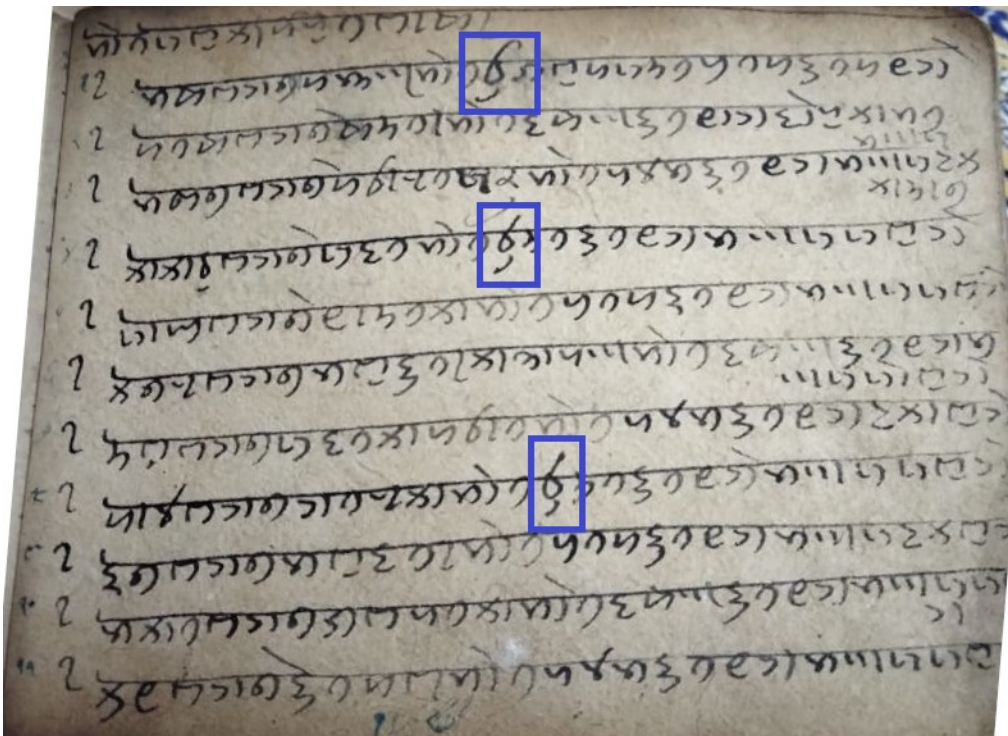


Figure 19: Examples of 6 for writing *ī*.

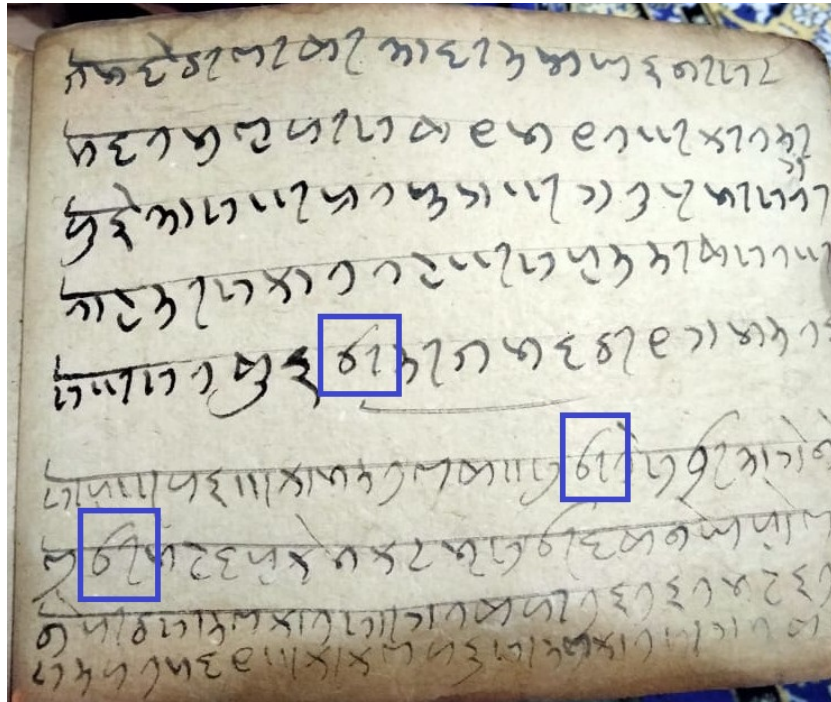


Figure 20: Example of 6 for representing \bar{i} , illustrating usage of 6 used as a vowel carrier bearing the vowel sign \circ .

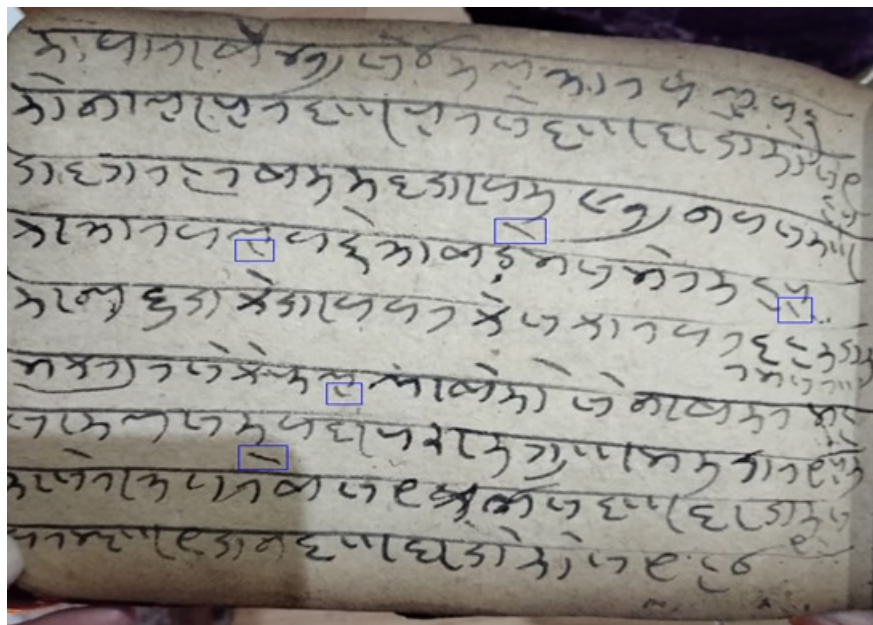


Figure 21: Usage of halanta.

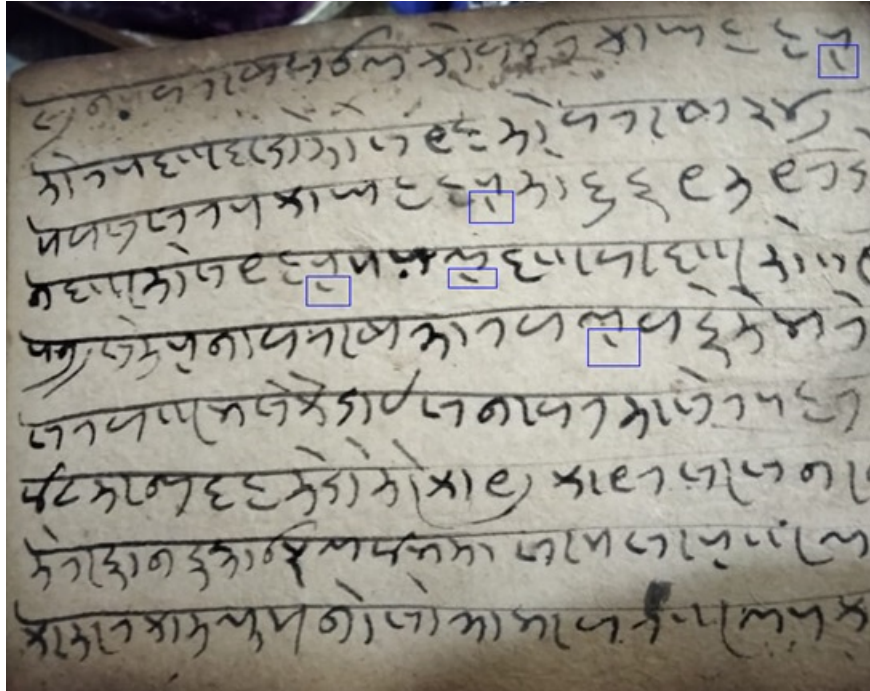


Figure 22: Additional examples of *halanta*.

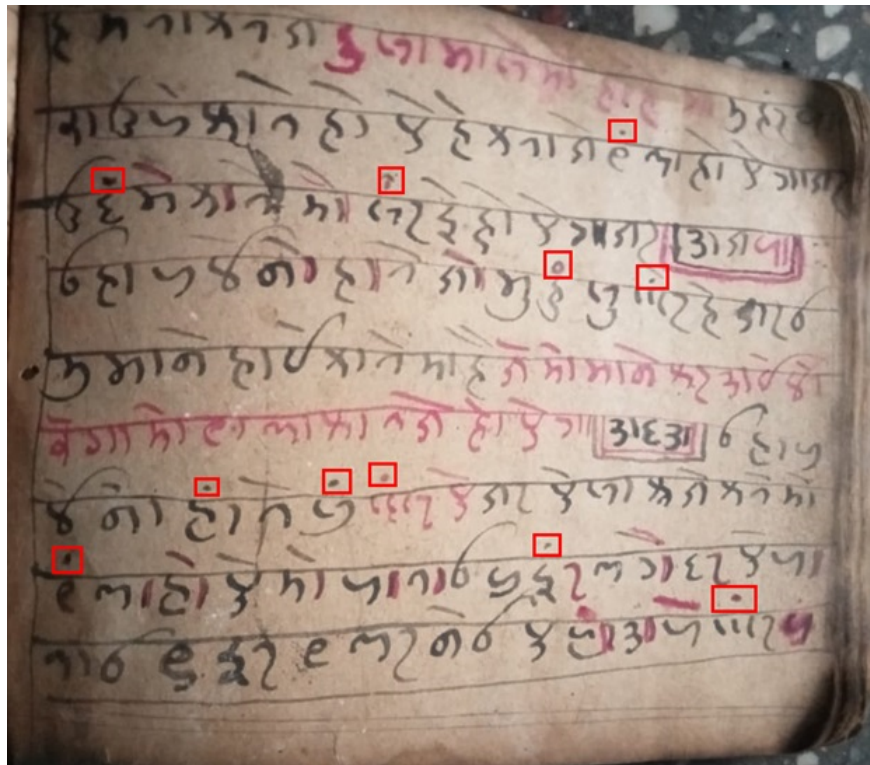


Figure 23: Usage of *anusvāra*

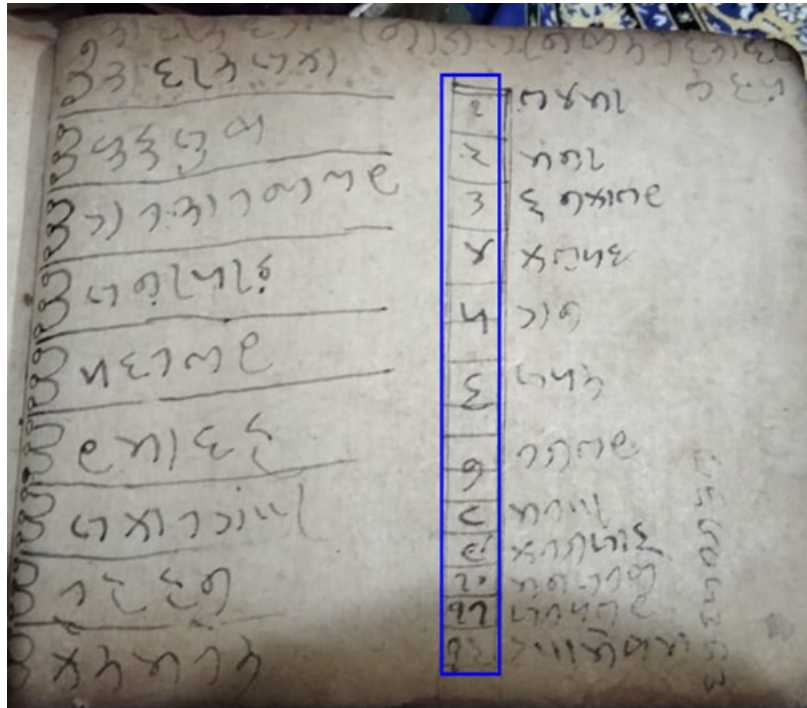


Figure 24: Examples of digits.

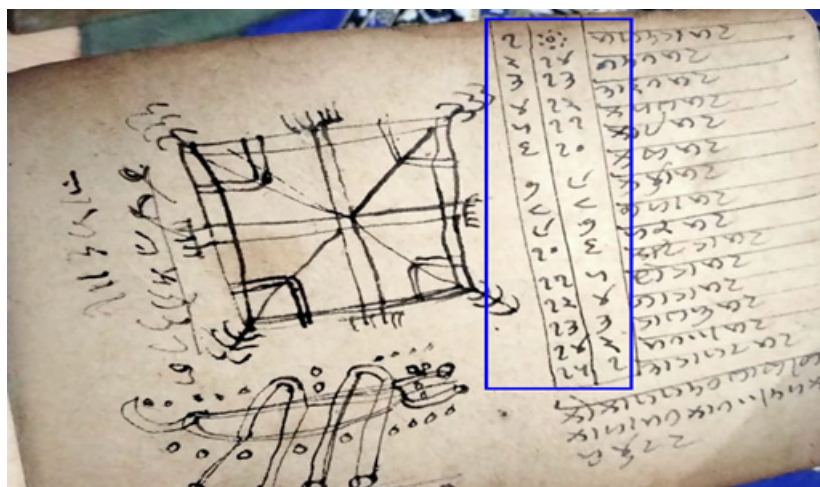


Figure 25: Examples of digits.

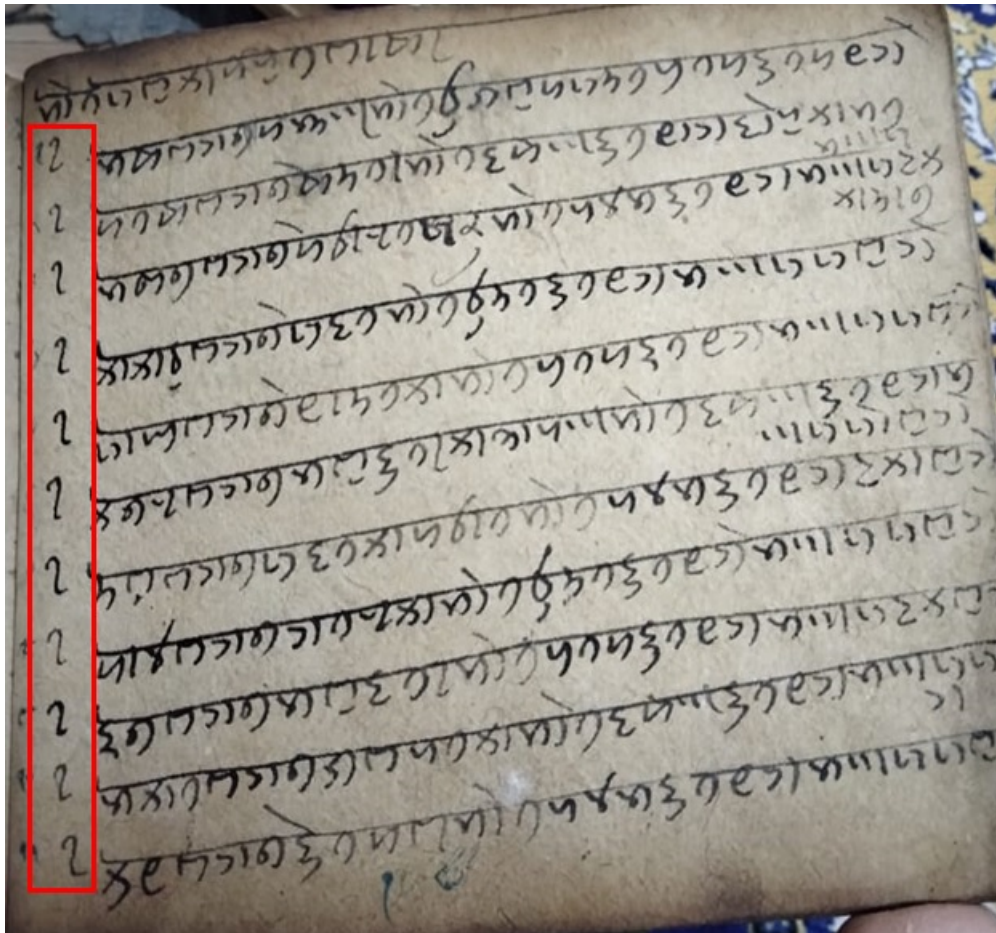


Figure 26: Usage of the auspicious sign ॐ ekam

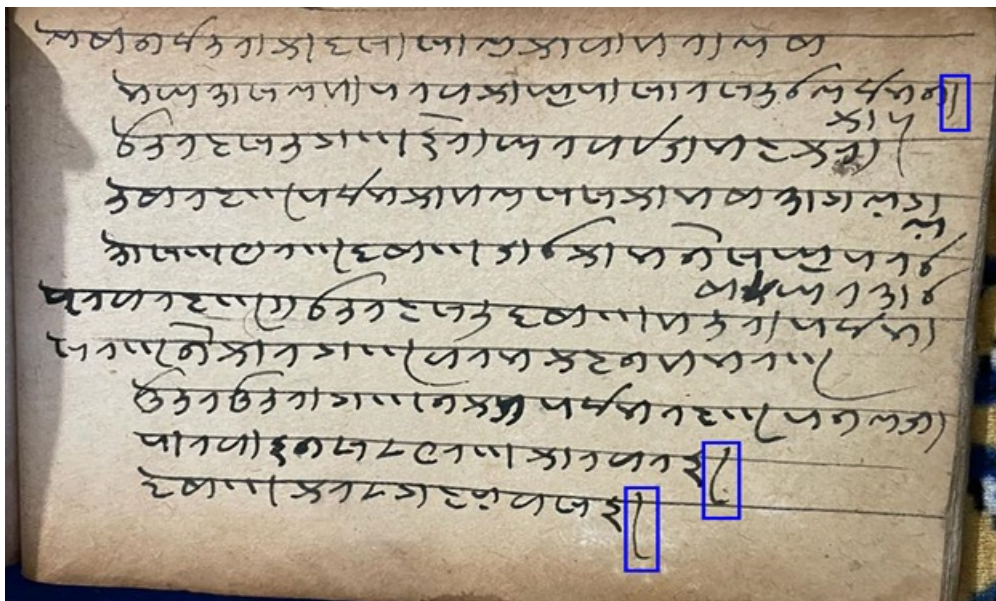


Figure 27: Usage of *dandā*-s.

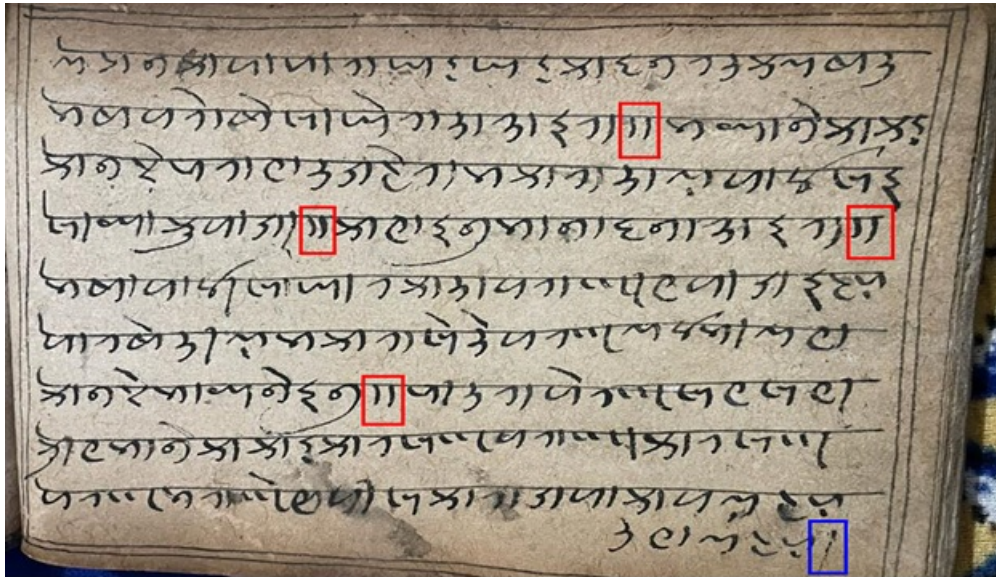


Figure 28: Single (blue) and double (red) *dandā*-s.

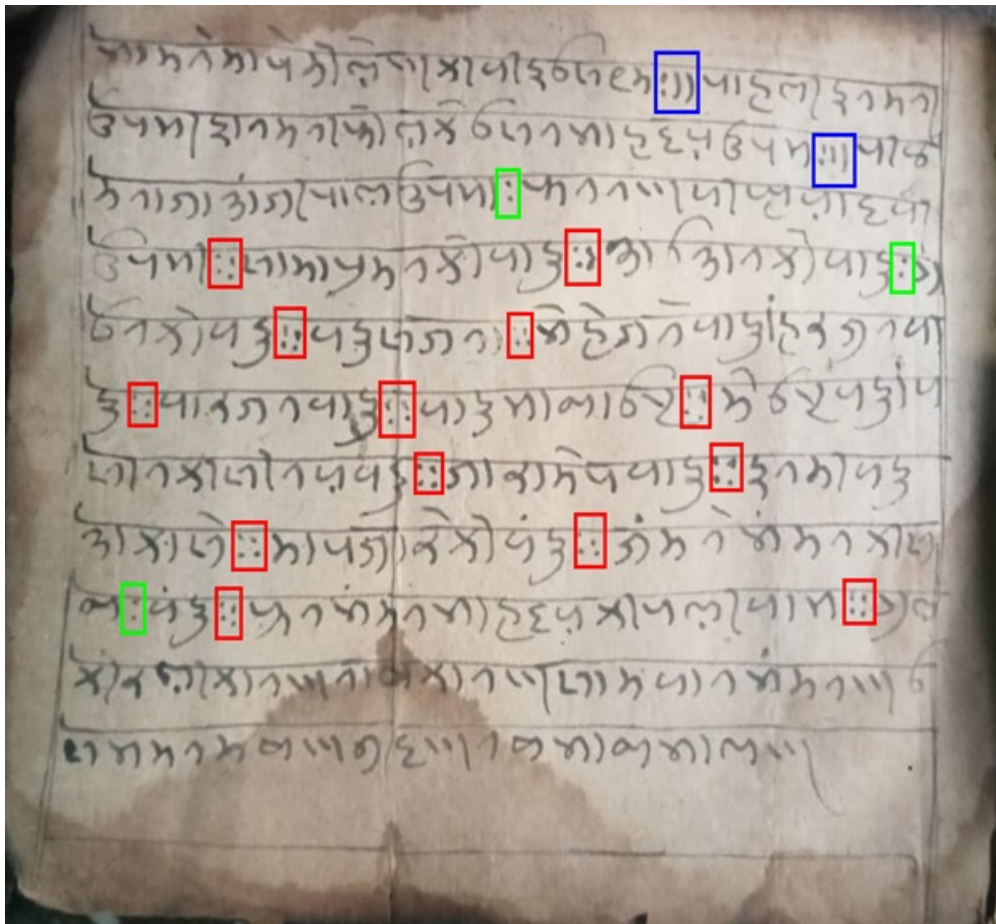


Figure 29: Various punctuation signs.

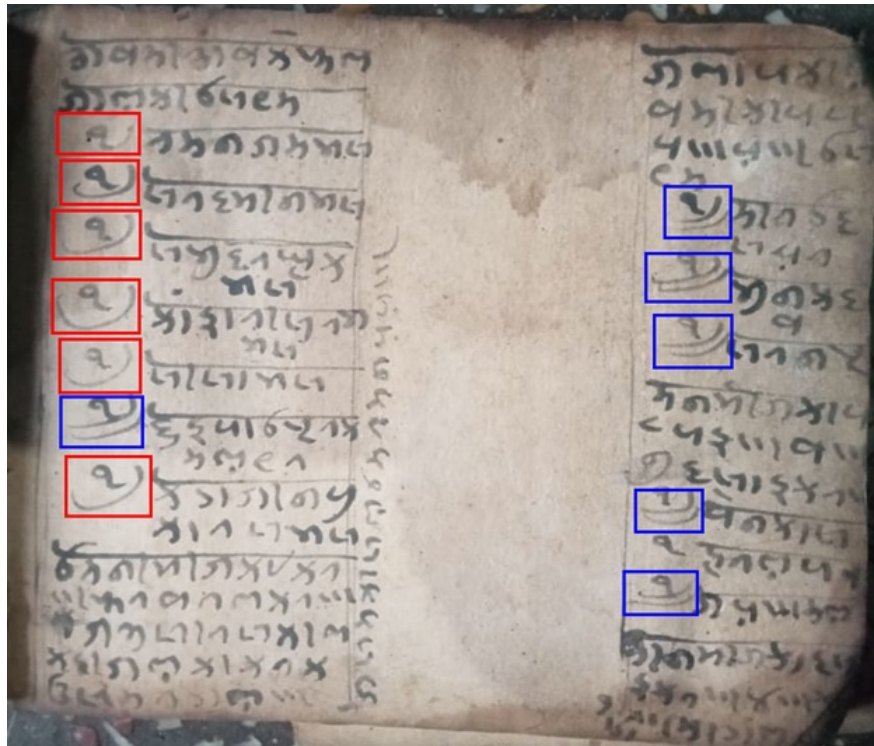


Figure 30: Section signs.

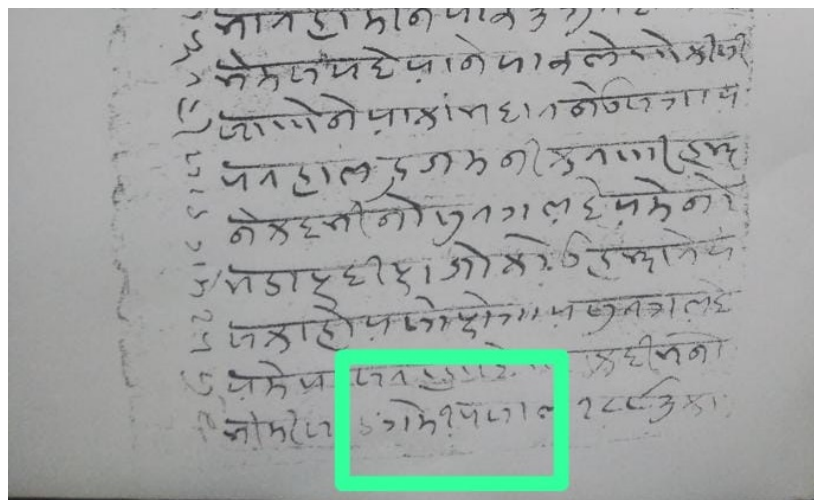


Figure 31: Usage of a Latin question mark in a Sirmauri document.

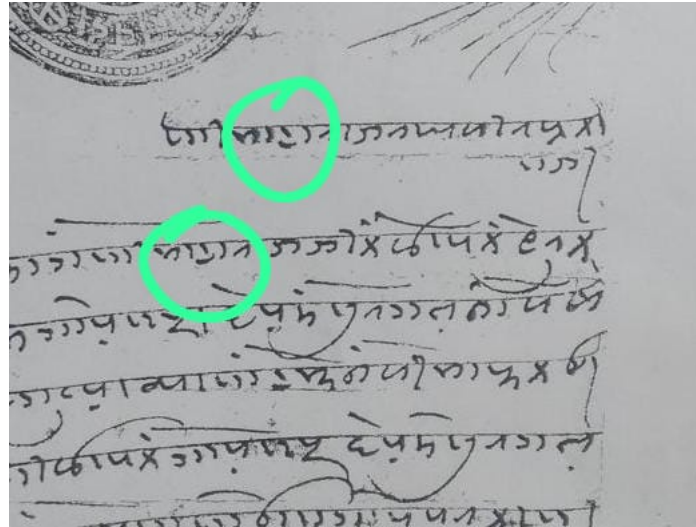


Figure 32: Document showing variant form ३ of ५ ha.

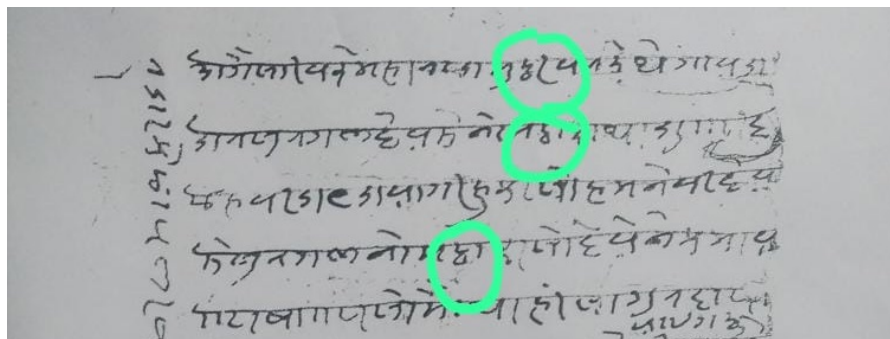


Figure 33: Document showing Jaunsari form ४ of ६ dha.

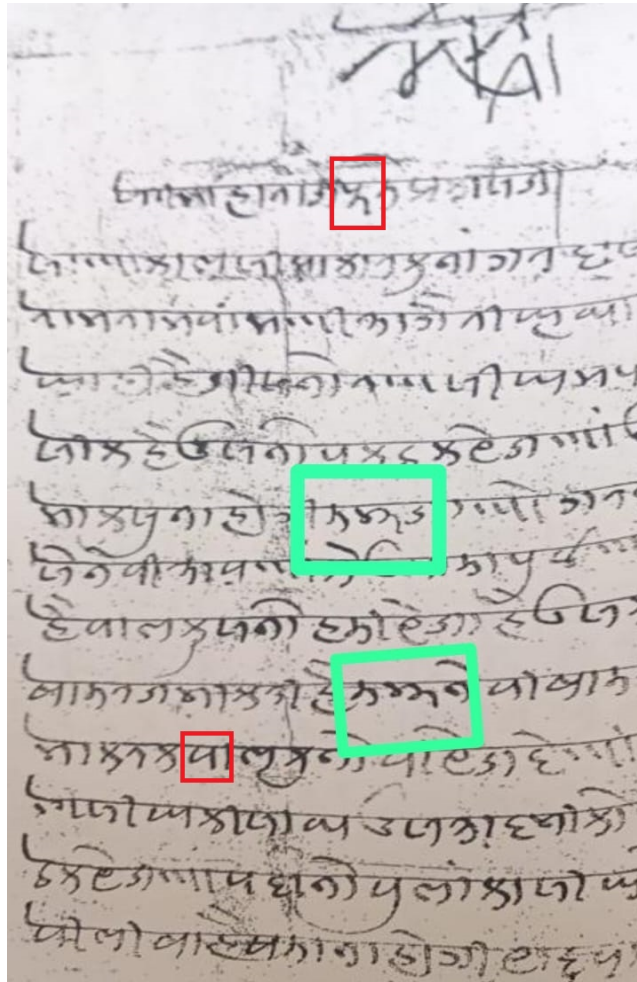


Figure 34: Document showing usage of the conjunct *mma* (green) and Devanagari-like forms for *pha* (red, top) and *pa* (red, bottom).

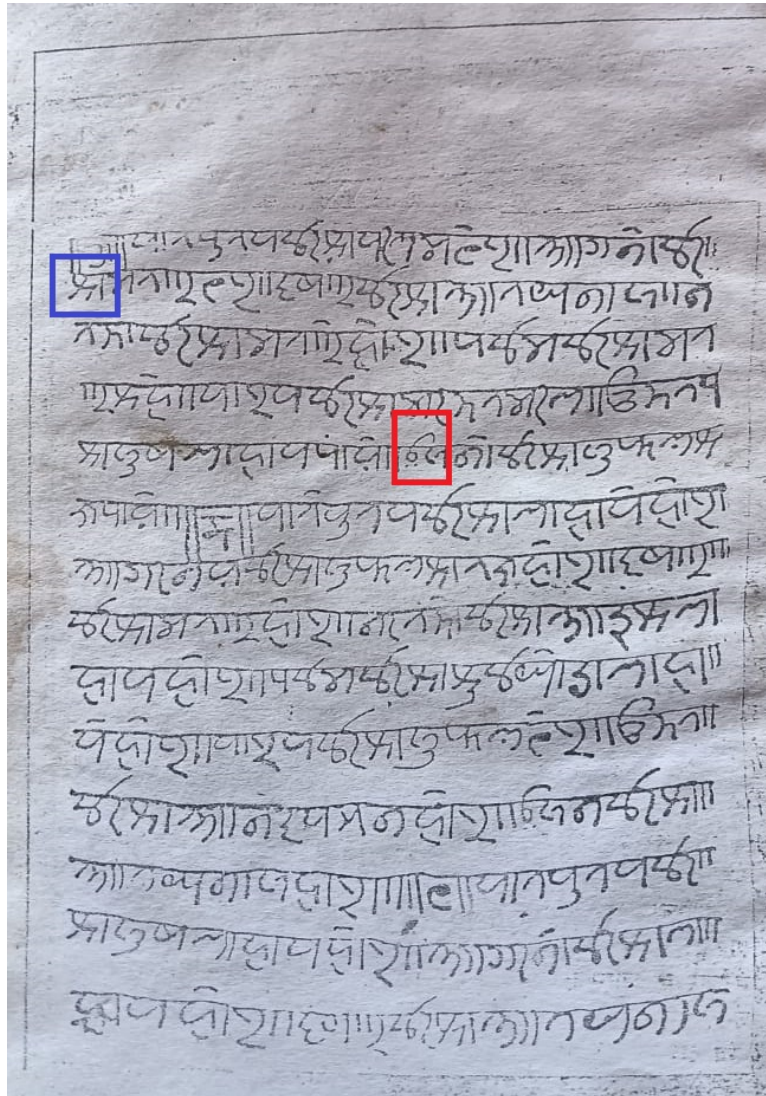


Figure 35: Usage of Devanagari-like letterforms and variants in a Sirmauri document. Alternate form २ of ३ ca (blue); usage of variant ६ for ७ for i.

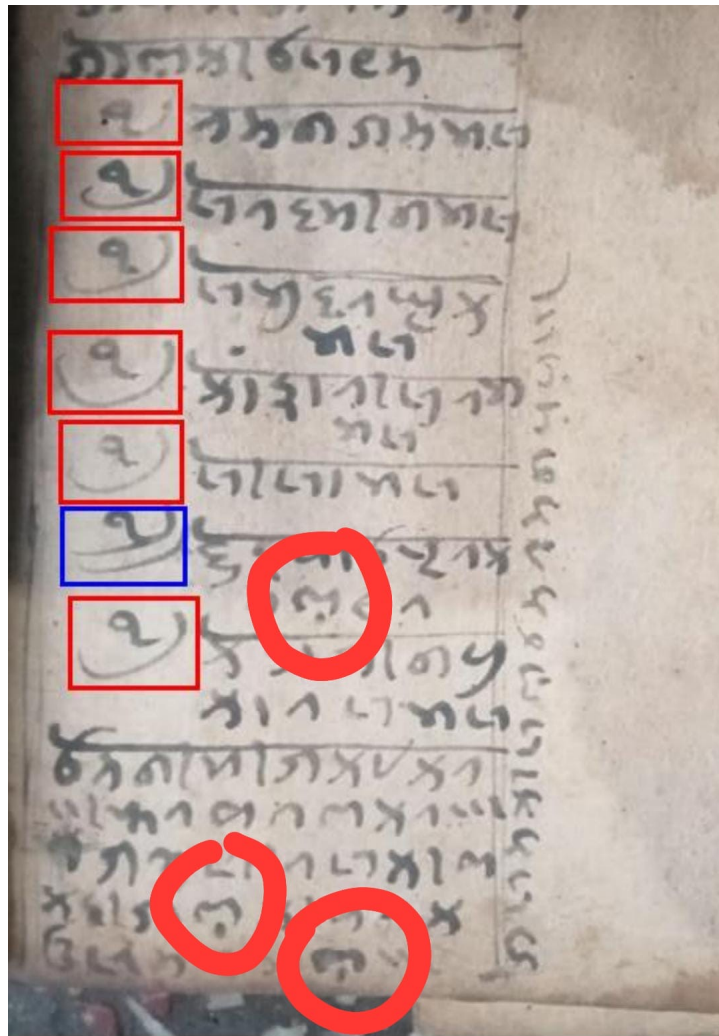


Figure 36: Usage of *ekam* + section marks; also *la* + *nukta* for writing *la*.

**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:	<i>Anshuman Pandey <pandey@umich.edu></i> <i>Biswajit Mandal <biswajitmandal.bm90@gmail.com></i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Expert contribution</i>
4. Submission date:	<i>2024-10-25</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/> <i>Yes</i>
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input checked="" type="checkbox"/> <i>Yes</i>	
Proposed name of script:	<i>Sirmauri</i>	
b. The proposal is for addition of character(s) to an existing block:		
Name of the existing block:		
2. Number of characters in proposal:	<i>65</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" 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?	<input checked="" type="checkbox"/> <i>Yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> <i>Yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> <i>Yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Anshuman Pandey</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Anshuman Pandey</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> <i>Yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> <i>Yes</i>	
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)?	<input checked="" type="checkbox"/> <i>Yes</i>	

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
	<i>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</i>
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 <i>See text of proposal</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare <i>See text of proposal</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>Used by Sirmauri language speakers for ritual and personal use</i>
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 <i>Virama</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No