ISO/IEC JTC1/SC2/WG2 N____
L2/14-036
2014-01-27

Revised Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646

Anshuman Pandey
Department of History
University of Michigan
Ann Arbor, Michigan, U.S.A.
pandey@umich.edu

Dragomir Dimitrov
Fachgebiet Indologie und Tibetologie
Philipps-Universität Marburg
Marburg, Germany
dimitrov@staff.uni-marburg.de

January 27, 2014

1 Introduction

This is a proposal to encode the Bhaiksuki script in the Universal Character Set (ISO/IEC 10646). It replaces the following documents:

- N4121 L2/11-259 "Preliminary Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646"
- N4469 L2/13-167 "Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646"
- N4489 L2/13-194 "Revised Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646"

This document is a revision of N4489 L2/13-194, which provides additional information on contextual forms of vowel signs and consonant letters.

2 Background

Bhaiksuki (𑰦𑰸𑰎𑰿𑰑 *bhaikṣukī*; Devanagari भैक्षुकी) is a Brahmi-based script that was used around the turn of the first millenium CE mainly in the present-day states of Bihar and West Bengal in India, as well as in regions that are now part of Bangladesh. Records have been also located in Tibet, Nepal, and Burma. The script is known variously as the 'Arrow-Headed Script' or 'Point-Headed Script' in English, 'Pfeilspitzenschrift' in German, and 'Śaramāṭṛkā Lipi' in Hindi and modern Sanskrit. An older designation, 'Sindhura', has been used in Tibet for at least three centuries.

The script is attested exclusively in Buddhist textual materials. Only eleven inscriptions and four manuscripts written in this script are presently known to exist. These are the Bhaiksuki manuscripts of the *Abhidharmasamuccayakārikā*, *Maṇicūḍajātaka*, *Candrālaṃkāra*, and at least one more Buddhist canonical text. The codex of the *Abhidharmasamuccayakārikā* was once kept in Tibet, but it is now inaccessible and its exact place of preservation is unknown. The fourth codex was discovered in Tibet and was recently shown in a Chinese documentary; however, information about this manuscript is limited. It is likely that additional materials in Bhaiksuki may become available in the future.

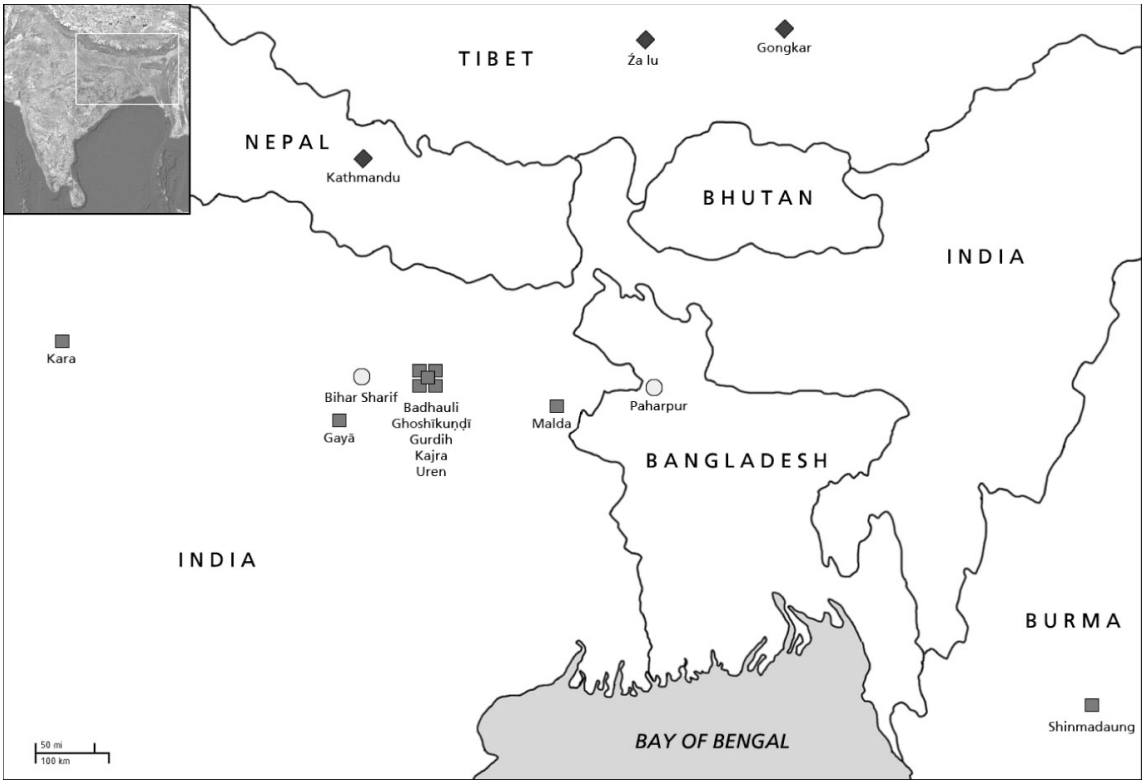


Figure 1: Sites of inscriptions (▩) and manuscripts (◆) and other places (○) where Bhaiksuki has reportedly been used (from Dimitrov 2010: 52).

There has been scholarly interest in Bhaiksuki from the time that Cecil Bendall (1856–1906) presented the script to Western academic communities in the 1880s. In the 1890s, Bruno Liebich (1862–1939) made further advances through his study of the materials available at the time. More recently, the Bhaiksuki manuscript of the *Maṇicūḍajātaka* was studied by Albrecht Hanisch in 2008 and the manuscript of the *Candrālaṃkāra* was presented by Dragomir Dimitrov in 2010. Since 2004 the Arrow-headed Script Project at Philipps-Universität, Marburg, Germany has been engaged in research on Bhaiksuki and in developing resources for the script.

3 Script Details

3.1 Name

The transliterated name of the script, *bhaikṣukī*, is normalized as 'Bhaiksuki'. The proposed block consists of 94 characters, which is tentatively allocated to the SMP at the range U+11C00..11C6F. Character names are aligned with those used in other script encoded in the UCS. A code chart and names list is attached.

3.2 Structure

The general structure (phonetic order, *mātrā* reordering, use of *virāma*, etc.) of Bhaiksuki is similar to that of Devanagari. Some dependent vowel signs consist of two or three parts and behave like those in

Bengali. Several vowel signs have contextual forms when they combine with certain consonants, and certain consonant-vowel sequences are written as ligatures. In some cases, consonant + *virāma* pairs are rendered using both a visible *virāma* and a special ligature. Consonant clusters are represented as conjuncts.

The structure of a Bhaiksuki consonantal syllable may be described as follows:

consonant [*consonant*]* [*vowel sign*] [CANDRABINDU | ANUSVARA] [VISARGA]

where there is one base consonant, which may occur in a conjunct with one or more consonants. The sources show conjuncts containing at least three consonants, but theoretically the number may be greater. According to the rules of the script, only one vowel sign may be used with a base consonant or conjunct. One of either the CANDRABINDU or ANUSVARA may occur with a consonant or vowel sign. The VISARGA may follow last.

3.3 Virāma

The ◌𑰿 VIRAMA is used for indicating the absence of the inherent vowel in a consonant letter. It is identical in function to the VIRAMA in Devanagari. Certain pairs of consonant + VIRAMA are rendered as both with visible VIRAMA and as a special ligature (see Section 3.7).

3.4 Vowels

There are 13 vowel letters:

𑰀	A	𑰄	U	𑰉	VOCALIC L	𑰏	AU
𑰁	AA	𑰅	UU	𑰊	E		
𑰂	I	𑰆	VOCALIC R	𑰋	AI		
𑰃	II	𑰇	VOCALIC RR	𑰌	O		

The vowel letter *VOCALIC LL is not attested, but space has been reserved for it.

3.5 Vowel Signs

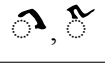
There are 12 dependent vowel signs:

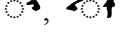
◌𑰸	VOWEL SIGN AA	◌𑰹	VOWEL SIGN UU	◌𑰺	VOWEL SIGN E
◌𑰻	VOWEL SIGN I	◌𑰼	VOWEL SIGN VOCALIC R	◌𑰽	VOWEL SIGN AI
◌𑰾	VOWEL SIGN II	◌𑰿	VOWEL SIGN VOCALIC RR	◌𑱀	VOWEL SIGN O
◌𑰶	VOWEL SIGN U	◌𑰷	VOWEL SIGN VOCALIC L	◌𑱁	VOWEL SIGN AU

The *VOWEL SIGN VOCALIC LL is not attested, but space has been reserved for it.

3.5.1 Contextual Alternates of Vowel Signs

Some vowel signs are written using alternate forms when they occur with certain consonants (see Section 3.8).

	Regular	Alternate
VOWEL SIGN AA		
VOWEL SIGN U		
VOWEL SIGN UU		
VOWEL SIGN E		

	Regular	Alternate
VOWEL SIGN AI		
VOWEL SIGN O		
VOWEL SIGN AU		

3.5.2 Recommendation Against the Decomposition of Two-Part Vowel Signs

The  VOWEL SIGN E,  VOWEL SIGN AI,  VOWEL SIGN O,  VOWEL SIGN AU are two-part vowel signs. Each sign contains the common element on the left and a distinguishing element on the right. In other scripts in the UCS, such two-part signs are decomposed into their constituent left and right parts; however, such an approach is not recommended for Bhaiksuki.

The primary reason is that the common left-side element  is not a meaningful sign in the script. Also, the right-hand elements of the four signs —  ,  ,  ,  — do not possess any inherent semantic value; moreover, the right-side element  of  VOWEL SIGN AI is the regular contextual alternate for  VOWEL SIGN E (see section 3.5.1).

Decomposition of these signs raises two major issues. First, it would entail adding five additional characters —  ,  ,  ,  , — to the proposed repertoire. Second, there are contextual alternates for these vowel signs, such as and for VOWEL SIGN O, and and for VOWEL SIGN AU, which consist of two-parts. Moreover, in certain contexts, the two-part VOWEL SIGN E is presented as the singular element . Are decompositions required for these contextual forms, which are the standard for certain consonants? If so, how to manage decompositions? Would is be necessary to encode these graphical components as separate characters?

While decomposition may be technically advantageous for some implementations, it is not recommended that artificial characters be encoded for the purpose of rendering two-part vowel signs.

3.6 Consonants

There are 33 consonant letters:

𑰎	KA	𑰞	NYA	𑰦	DHA	𑰭	LA
𑰏	KHA	𑰟	TTA	𑰧	NA	𑰮	VA
𑰐	GA	𑰠	TTHA	𑰨	PA	𑰯	SHA
𑰑	GHA	𑰡	DDA	𑰩	PHA	𑰰	SSA
𑰒	NGA	𑰢	DDHA	𑰪	BA	𑰱	SA
𑰓	CA	𑰣	NNA	𑰫	BHA	𑰲	HA
𑰔	CHA	𑰤	TA	𑰬	MA		
𑰕	JA	𑰥	THA	𑰴	YA		
𑰖	JHA	𑰦	DA	𑰵	RA		

Each consonant bears the inherent vowel /a/, which is silenced using ◌𑰿 VIRAMA. Consonant clusters are written as conjuncts (see Section 3.9).

3.6.1 Contextual forms of consonants

The letters PA, YA, RA are slightly modified when they occur with specific vowel signs.

PA The letter 𑰨 PA takes the contextual form 𑰹 when written with certain vowel signs. The regular form is used in

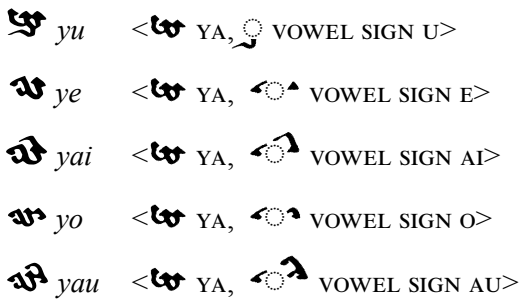
- 𑰨𑰸 *pu* <𑰨 PA, ◌𑰸 VOWEL SIGN U>
- 𑰨𑰹 *pū* <𑰨 PA, ◌𑰹 VOWEL SIGN UU>
- 𑰨𑰺 *pṛ* <𑰨 PA, ◌𑰺 VOWEL SIGN VOCALIC R>
- 𑰨𑰻 *pe* <𑰨 PA, ◌𑰻 VOWEL SIGN E>
- 𑰨𑰼 *pai* <𑰨 PA, ◌𑰼 VOWEL SIGN AI>
- 𑰨𑰽 *po* <𑰨 PA, ◌𑰽 VOWEL SIGN O>
- 𑰨𑰾 *pau* <𑰨 PA, ◌𑰾 VOWEL SIGN O>

while the form 𑰹 is used in

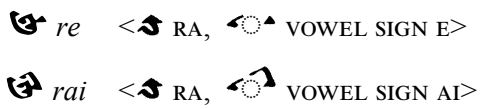
- 𑰨𑰺 *pā* <𑰨 PA, ◌𑰺 VOWEL SIGN AA>
- 𑰨𑰷 *pi* <𑰨 PA, ◌𑰷 VOWEL SIGN I>
- 𑰨𑰶 *pī* <𑰨 PA, ◌𑰶 VOWEL SIGN II>

This alternate shape also occurs in consonant conjuncts when PA is a non-initial consonant.

YA The letter 𑰬 YA takes a contextual form when it combines with VOWEL SIGN E, VOWEL SIGN AI, VOWEL SIGN O, VOWEL SIGN AU:



RA The letter 𑰢 RA takes a contextual form when it combines with all vowel signs:



3.7 Consonant-Virama Ligatures

Sequences of <*consonant*, VIRAMA> are rendered by default using a visible VIRAMA. However, three combinations are also represented using a special ligature, which is referred to here as a "*khaṇḍa*" form:

	combining *virāma*	*khaṇḍa* ligature
𑰢 TA + 𑰿 VIRAMA	𑰢𑰿	𑰣
𑰤 NA + 𑰿 VIRAMA	𑰤𑰿	𑰥
𑰫 MA + 𑰿 VIRAMA	𑰫𑰿	𑰪

An analysis of the available manuscripts indicates that there is no semantic distinction between the visible *virāma* forms and *khaṇḍa* ligatures. The two forms of <TA, VIRAMA> and <NA, VIRAMA> are used alternately in the same context. The *khaṇḍa* form of TA is used inconsistently and when it does occur, its use is functionally identical to ৎ U+09CE BENGALI LETTER KHANDA TA. The combination <MA, VIRAMA> occurs only as the *khaṇḍa* ligature in the available sources.

It is necessary to represent both the *virāma* and *khaṇḍa* forms because they occur simultaneously. Moreover, they must be distinguished for analytical purposes. For instance, in a pedagogical text on Bhaiksuki palaeography that describes the formation of the conjunct 𑰥 *mpa*, it is necessary to render <MA, VIRAMA> using a visible VIRAMA, eg. the statement "𑰫𑰿 *m* + 𑰬 *pa* → 𑰥 *mpa*" properly expresses the rule that a conjoining form of PA is written beneath the regular form of MA in order to produce the conjunct *mpa*, while the statement "𑰪 *m* + 𑰬 *pa* → 𑰥 *mpa*" does not correctly do so.

Given that there are only three attested *khaṇḍa* forms, it may be possible to encode each as independent characters. The limitation of this approach is that the discovery of other *khaṇḍa* forms would require encoding each as a distinct character. Another option is to encode a combining character with VIRAMA-like behavior and properties, which would control the representation of *khaṇḍa* forms. Such a control character would

allow for a generic way of representing both existing and other potential *khaṇḍa* forms; however, encoding such a character may increase the complexities of implementing support for the script in rendering engines.

As there is no known semantic distinction between the *khaṇḍa* forms and their visible *virāma* representations, the three ligatures are to be considered contextual variants and will not be represented in plain text. The display of *khaṇḍa* forms is to be controlled using smart font ligature features, such as those available in Graphite and OpenType. If a requirement to represent these *khaṇḍa* ligatures at the character level arises as a result of new information, then the matter may be discussed again at that time.

3.8 Consonant-Vowel Combinations

Several vowel signs are written using alternate forms when they combine with certain consonants. This is standard behavior for Bhaiksuki. These alternate forms are contextual variants and are not proposed for separate encoding. The font is responsible for producing the approprate alternate form of a vowel sign based upon the presence of the base consonant.

3.8.1 VOWEL SIGN AA

The shape of ◌𑰮 VOWEL SIGN AA differs based upon the consonant with which is it written. For letters with a single arrow-head, the regular sign ◌𑰮 is attache to the right of the main arrow. This sign is used with KA, CA, CHA, JHA, DDA, DDHA, TA, DA, NA, PHA, BHA, RA, VA, HA:

𑰎𑰮 *kā* <𑰎 KA, ◌𑰮 VOWEL SIGN AA>

𑰏𑰮 *cā* <𑰏 CA, ◌𑰮 VOWEL SIGN AA>

𑰓𑰮 *ḍā* <𑰓 DA, ◌𑰮 VOWEL SIGN AA>

𑰔𑰮 *tā* <𑰔 TA, ◌𑰮 VOWEL SIGN AA>

𑰓𑰮 *dā* <𑰓 DA, ◌𑰮 VOWEL SIGN AA>

𑰤𑰮 *hā* <𑰤 HA, ◌𑰮 VOWEL SIGN AA>

For letters with two arrow-heads or whose arrow head is positioned to the right, the short stroke ◌𑰯 is added to the existing arrow. This occurs with GA, GHA, NGA, JA, TTA, TTHA, NNA, THA, DHA, BA, MA, YA, LA, SHA, SSA, SA:

𑰐𑰯 *ghā* <𑰐 GHA, ◌𑰮 VOWEL SIGN AA>

𑰑𑰯 *ṅā* <𑰑 NGA, ◌𑰮 VOWEL SIGN AA>

𑰒𑰯 *jā* <𑰒 JA, ◌𑰮 VOWEL SIGN AA>

𑰖𑰯 *ṭhā* <𑰖 TTHA, ◌𑰮 VOWEL SIGN AA>

𑰗𑰯 *ṇā* <𑰗 NNA, ◌𑰮 VOWEL SIGN AA>

𑰘𑰯 *thā* <𑰘 THA, ◌𑰮 VOWEL SIGN AA>

𑰙𑰯 *dhā* <𑰙 DHA, ◌𑰮 VOWEL SIGN AA>

 *pā* <PA, ◌ VOWEL SIGN AA>

 *bā* <BA, ◌ VOWEL SIGN AA>

 *yā* <YA, ◌ VOWEL SIGN AA>

 śā <SHA, ◌ VOWEL SIGN AA>

 *sā* <SA, ◌ VOWEL SIGN AA>

The VOWEL SIGN AA takes the shape ◌ when it attaches to the letter KHA, which has no arrow-head:

 *khā* <KHA, ◌ VOWEL SIGN AA>

3.8.2 VOWEL SIGN I

When ◌ VOWEL SIGN I occurs with a single consonant, it replaces the arrow-head of the letter to which it is attached:

 *ki* <KA, ◌ VOWEL SIGN I>

 *gi* <GA, ◌ VOWEL SIGN I>

 ñi <NYA, ◌ VOWEL SIGN I>

 *di* <DA, ◌ VOWEL SIGN I>

 *ti* <TA, ◌ VOWEL SIGN I>

If the letter to which VOWEL SIGN I attaches has two arrow heads, then the sign is written in place of the right-hand arrow head:

 *ṅi* <NGA, ◌ VOWEL SIGN I>

 *dhi* <DHA, ◌ VOWEL SIGN I>

 mi <MA, ◌ VOWEL SIGN I>

 *yi* <YA, ◌ VOWEL SIGN I>

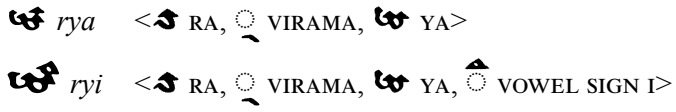
In the case of KHA, which has no arrow head, the sign attaches to the upper right side of the letter, eg. .

 *khi* <KHA, ◌ VOWEL SIGN I>

When VOWEL SIGN I occurs in a conjunct in which RA is the initial consonant (*repha*), the sign is written in place of the *repha*:

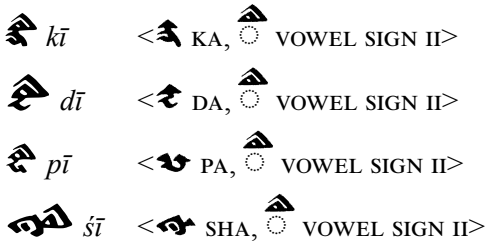
 *rtta* <RA, ◌ VIRAMA, TA, ◌ VIRAMA, TA>

 *rtti* <RA, ◌ VIRAMA, TA, ◌ VIRAMA, TA, ◌ VOWEL SIGN I>

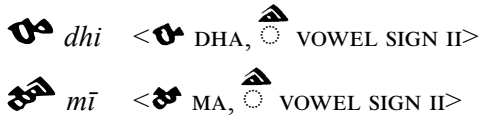


3.8.3 VOWEL SIGN II

The ◌ VOWEL SIGN II behaves identically to ◌ VOWEL SIGN I in that it replaces the arrow-head of the letter with which it occurs:



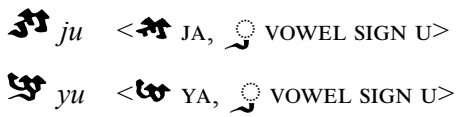
If the letter to which VOWEL SIGN II attaches has two arrow heads, then the sign is written in place of the right-hand arrow head:



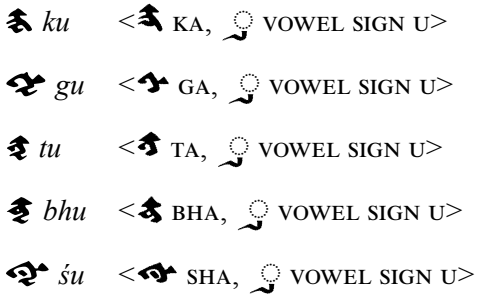
In the case of KHA, which has no arrow head, the sign attaches to the upper right side of the letter, eg. .

3.8.4 VOWEL SIGN U

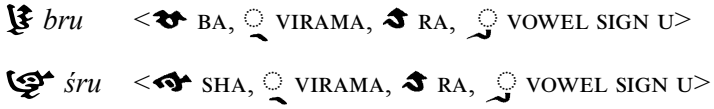
The ◌ VOWEL SIGN U is attached beneath the consonant letter:



It takes the shape ◌ when written with KA, GA, TA, BHA, SHA:



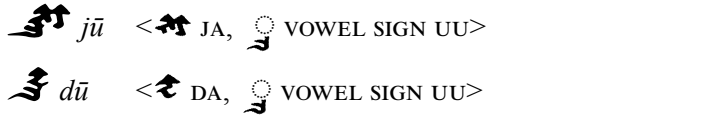
It takes the same shape with non-initial RA:



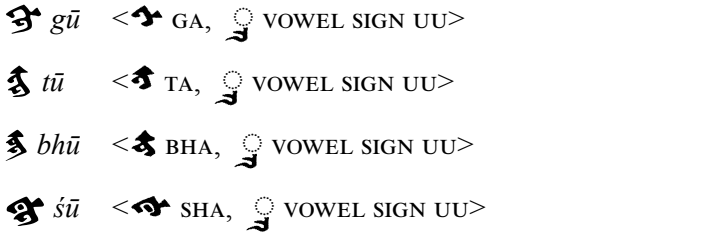
However, when VOWEL SIGN U occurs with an independent RA the combination is rendered as a special ligature:

3.8.5 VOWEL SIGN UU

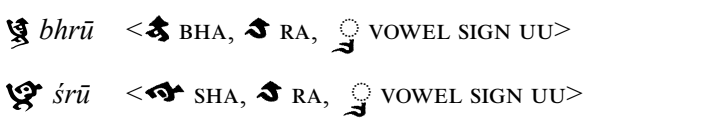
The VOWEL SIGN UU is written below the letter:



It takes the alternate form when it occurs with KA, GA, TA, BHA, SHA:



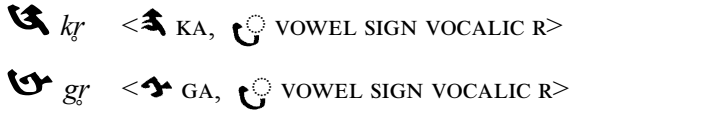
It takes the same shape with non-initial RA:



However, when VOWEL SIGN UU occurs with an independent RA the combination is written as a special ligature:

3.8.6 VOWEL SIGN VOCALIC R

The VOWEL SIGN VOCALIC R is written as an extension of the final downward stroke of the letter with which it occurs:



It could be noticed here that the combination of a consonant with the VOWEL SIGN VOCALIC R is clearly differentiated from the combination with the consonant RA:

 *tra* <TA, VIRAMA, RA>

 *tṛ* <TA, VOWEL SIGN VOCALIC R>

 *sra* <SA, VIRAMA, RA>

 *sṛ* <SA, VOWEL SIGN VOCALIC R>

3.8.7 VOWEL SIGN VOCALIC RR

The VOWEL SIGN VOCALIC RR is written by adding another stroke beneath VOWEL SIGN VOCALIC R:

 *kṝ* <KA,  VOWEL SIGN VOCALIC RR>

 *gṝ* <GA,  VOWEL SIGN VOCALIC RR>

3.8.8 VOWEL SIGN VOCALIC L

The  VOWEL SIGN VOCALIC L attaches below the letter:

 *dḷ* <DA,   VOWEL SIGN VOCALIC L>

3.8.9 VOWEL SIGN E

The  VOWEL SIGN E consists of two parts and is written to the right and left of a consonant letter:

 *ke* <KA,  VOWEL SIGN E>

 *ne* <NA,  VOWEL SIGN E>

It takes the alternate shape  with the following letters, whose arrow-head is placed to the right of the letter body: GA, NGA, NYA, TTA, TTHA, NNA, THA, DHA, SHA, eg.

 *ge* <GA,  VOWEL SIGN E>

 *ṭe* <TTA,  VOWEL SIGN E>

 *ṭhe* <TTHA,  VOWEL SIGN E>

 *ṇe* <NNA,  VOWEL SIGN E>

 *the* <THA,  VOWEL SIGN E>

 *śe* <SHA,  VOWEL SIGN E>

In the case of KHA, which has no arrow-head, the sign takes the shape  :

  khe <KHA,   VOWEL SIGN E>

3.8.10 VOWEL SIGN AI

The ◌𑰯 VOWEL SIGN AI consists of two parts and is written to the right and left of a consonant letter, eg.

𑰎𑰯 *kai* <𑰎 KA, ◌𑰯 VOWEL SIGN AI>

𑰛𑰯 *jai* <𑰛 JA, ◌𑰯 VOWEL SIGN AI>

𑰢𑰯 *nai* <𑰢 NA, ◌𑰯 VOWEL SIGN AI>

𑰩𑰯 *pai* <𑰩 PA, ◌𑰯 VOWEL SIGN AI>

It takes the shape ◌𑰻 when it occurs with GA, NGA, NYA, TTA, TTHA, NNA, THA, DHA, SHA, eg.

𑰟𑰻 *ṇai* <𑰟 NNA, ◌𑰯 VOWEL SIGN AI>

𑰪𑰻 *śai* <𑰪 SHA, ◌𑰯 VOWEL SIGN AI>

For the letter KHA, which does not have an arrow-headed top, the VOWEL SIGN AI is written as ◌𑰰:

𑰑𑰰 *khyai* <𑰑 KHA, ◌𑰯 VOWEL SIGN AI>

3.8.11 VOWEL SIGN O

The ◌𑰱 VOWEL SIGN O consists of two parts and attaches to the right and left of a consonant letter, eg.

𑰎𑰱 *ko* <𑰎 KA, ◌𑰱 VOWEL SIGN O>

𑰘𑰱 *co* <𑰘 CA, ◌𑰱 VOWEL SIGN O>

𑰣𑰱 *to* <𑰣 TA, ◌𑰱 VOWEL SIGN O>

It takes the alternate shape ◌𑰺 with the letters GA, NGA, NYA, TTA, TTHA, NNA, THA, DHA, SHA:

𑰏𑰺 *go* <𑰏 GA, ◌𑰱 VOWEL SIGN O>

𑰐𑰺 *ṅo* <𑰐 NGA, ◌𑰱 VOWEL SIGN O>

𑰑𑰺 *ño* <𑰑 NYA, ◌𑰱 VOWEL SIGN O>

𑰔𑰺 *ṭo* <𑰔 TTA, ◌𑰱 VOWEL SIGN O>

𑰕𑰺 *ṭho* <𑰕 TTHA, ◌𑰱 VOWEL SIGN O>

𑰟𑰺 *ṇo* <𑰟 NNA, ◌𑰱 VOWEL SIGN O>

𑰤𑰺 *tho* <𑰤 THA, ◌𑰱 VOWEL SIGN O>

𑰥𑰺 *dho* <𑰥 DHA, ◌𑰱 VOWEL SIGN O>

𑰪𑰺 *śo* <𑰪 SHA, ◌𑰱 VOWEL SIGN O>

When the sign is written with DDA, NA, and RA it takes the shape ◌ॱ:

 *no* <NA, VOWEL SIGN O>

 *ro* <RA, VOWEL SIGN O>

In the case of KHA, which has no arrow-head, the sign takes the shape ◌ॲ, similar to the contextual form of VOWEL SIGN E used with the letter:

 *kho* <KHA, VOWEL SIGN O>

The letters KHA, NYA, DDA, NA, YA, RA are slightly modified when they occur with VOWEL SIGN O. These modifications of consonant letters are described in section 3.9.1. In the case of NYA, the vertical stroke on the right-hand side is replaced with the right-hand element of ◌ॱ:

 *no* <NYA, VOWEL SIGN O>

 *yo* <YA, VOWEL SIGN O>

 *ro* <RA, VOWEL SIGN O>

3.8.12 VOWEL SIGN AU

The ◌ॻ VOWEL SIGN AU consists of two parts and is written to the right and left of a consonant, eg.

 *cau* <CA, VOWEL SIGN AU>

 *ḍhau* <DDHA, VOWEL SIGN AU>

 *dau* <DA, VOWEL SIGN AU>

 pau <PA, VOWEL SIGN AU>

 bhau <BHA, VOWEL SIGN AU>

It takes the alternate shape ◌ॴ when it is written with GA, NGA, NYA, TTA, TTHA, NNA, THA, DHA, SHA:

 *gau* <GA, VOWEL SIGN AU>

 *ṇau* <NNA, VOWEL SIGN AU>

 *dhau* <DHA, VOWEL SIGN AU>

 *śau* <SHA, VOWEL SIGN AU>

It takes the shape ◌ॵ when it occurs with NA and RA, with which the right-hand vertical stroke is elongated.

 nau <NA, VOWEL SIGN AU>

Although combinations of ॺ KHA, ॼ NYA, ॾ DDA with VOWEL SIGN AU are not attested, such pairs would

be represented using the contextual form  . This rationale is based upon the form of VOWEL SIGN O that is used with these letters.

3.9 Consonant Conjuncts

Consonant clusters are written as conjuncts, which are generally rendered as vertically stacked ligatures, with non-initial consonants joined sequentially beneath the initial letter. In some cases, conjuncts may be rendered as independent ligatures. The encoded representation for conjuncts is

<*consonant*, ◌ VIRAMA, [*consonant*, ◌ VIRAMA,]* *consonant*>

Generally, the arrow-heads of non-initial consonants are removed when they are subjoined, eg. see *cca*, *tta*, *dda* below. In some cases, however, the consonant retains its arrow-head, eg. see *ṅga*, *ṅśa*, *jña*, *ṣṭa* below. Examples of conjuncts are:

 *kṣa* <KA, ◌ VIRAMA, SSA>

 *ggra* <GA, ◌ VIRAMA, GA, ◌ VIRAMA, RA>

 *gna* <GA, ◌ VIRAMA, NA>

 *ṅga* <NGA, ◌ VIRAMA, GA>

 *ṅśa* <NGA, ◌ VIRAMA, SHA>

 *cca* <CA, ◌ VIRAMA, CA>

 *jña* <JA, ◌ VIRAMA, NYA>

 *ṇḍa* <NNA, ◌ VIRAMA, DDA>

 *ṇṇa* <NNA, ◌ VIRAMA, NNA>

 *tka* <TA, ◌ VIRAMA, KA>

 *tta* <TA, ◌ VIRAMA, TA>

 *tpa* <TA, ◌ VIRAMA, PA>

 *dda* <DA, ◌ VIRAMA, DA>

 *ddha* <DA, ◌ VIRAMA, DHA>

 *ntha* <NA, ◌ VIRAMA, THA>

 *ndha* <NA, ◌ VIRAMA, DHA>

 nna <NA, ◌ VIRAMA, NA>

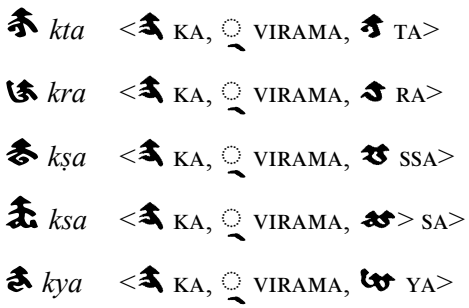
 bdha <BA, ◌ VIRAMA, DHA>

 ṣka <SSA, ◌ VIRAMA, KA>

3.9.1 Contextual forms of consonants in conjuncts

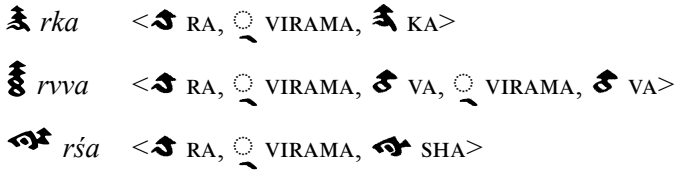
Several letters take special forms when they occur in conjuncts:

KA When initial, KA is slightly modified to :



PA The letter PA takes the contextual form when non-initial in a cluster:

RA When cluster initial, RA is written as an arrow-head *repha* above the following consonant in the cluster:



When *repha* occurs with a letter that has two arrow-heads, eg. SHA, SA, YA, etc., it is written on top of the right-hand arrow head:

When RA occurs final in a cluster it is written using the subjoined form (which is to be differentiated from the VOWEL SIGN VOCALIC R):

When VOWEL SIGN VOCALIC R occurs with RA, the latter is written as *repha* and placed above the arrow-head of the vowel letter:

ṛṛ <RA, VOWEL SIGN VOCALIC R>

YA When in conjunct final position, YA is written as the subjoined form:

kya <KA, VIRAMA, YA>

jya <JA, VIRAMA, YA>

yya <YA, VIRAMA, YA>

3.9.2 Conjuncts and Vowel Signs

Above-base and spacing signs attach to the glyph of the initial consonant, while below-base signs attach to the glyph of the final consonant:

kta <KA, VIRAMA, TA>

ktā <KA, VIRAMA, TA, VOWEL SIGN AA>

kti <KA, VIRAMA, TA, VOWEL SIGN I>

ktu <KA, VIRAMA, TA, VOWEL SIGN U>

ktau <KA, VIRAMA, TA, VOWEL SIGN AU>

ktyā <KA, VIRAMA, TA, VIRAMA, YA, VOWEL SIGN AA>

The shaping of vowel signs is dependent upon the letter to which they attach (see section 3.8).

3.10 Various Signs

Candrabindu The SIGN CANDRABINDU is used for nasalization. The variant form is written with a dot instead of a ring. This form is a glyphic variant and is semantically identical to the regular form.

Anusvara The SIGN ANUSVARA is used for nasalization. It has the variant form, written as a dot instead of as a ring. The dotted form is a glyphic variant and is semantically identical to the regular form.

Visarga The SIGN VISARGA represents post-vocalic aspiration (/h/) in Sanskrit.

Avagraha The SIGN AVAGRAHA marks the elision of word-initial *a* in Sanskrit as a result of sandhi.

3.11 Punctuation

***Daṇḍā*-s** The DANDA and DOUBLE DANDA are marks of general punctuation.

Word Separator The WORD SEPARATOR is used for demarcating lexical boundaries. It is written at the head-height. It also appears as a dot instead of a vertical bar.

Gap Filler The GAP FILLER is used generally as a spacing or completion mark, especially for justifying a text block, both at the end of the line before the binding area of the palm-leaf, as well as at the absolute end of a line. In other cases, the mark is used for indicating a gap after a deletion or to indicate a lacuna. The GAP FILLER is commonly written as , but the shape is also attested. In some cases, this latter shape is used as mark of deletion and is written over the text to be erased. These two characters may be considered glyphic variants of a single gap, completion, or deletion mark.

3.12 Digits

The following decimal digits are attested: ONE, TWO, FOUR, FIVE, SIX, SEVEN, EIGHT, NINE. The digits *ZERO and *THREE are not attested, but space has been reserved for them.

3.13 Numbers

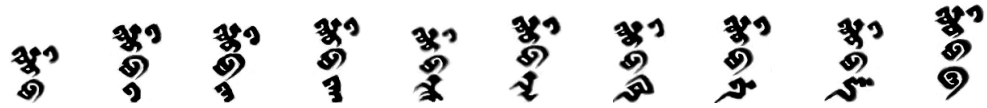
There is a positional number system used in addition to digits. It contains numbers (referred to as 'letter-numerals' in the scholarly literature) for the primary and tens units and a generic unit mark for the hundreds:

NUMBER ONE	NUMBER EIGHT	NUMBER SIXTY
NUMBER TWO	NUMBER NINE	NUMBER SEVENTY
NUMBER THREE	NUMBER TEN	NUMBER EIGHTY
NUMBER FOUR	NUMBER TWENTY	NUMBER NINETY
NUMBER FIVE	NUMBER THIRTY	NUMBER HUNDREDS
NUMBER SIX	NUMBER FORTY	
NUMBER SEVEN	NUMBER FIFTY	

Numbers are written vertically with each unit occupying a separate line. The largest unit is written first and all other units are written below. For instance, in writing the number 11, the NUMBER TEN is written first and NUMBER ONE is written below it. The same pattern is followed for all numbers 10–99. Shown below are the numbers 10–19:

10 11 12 13 14 15 16 17 18 19

The hundreds are written using the unit mark NUMBER HUNDREDS followed by a number that represents the primary numbers 1–9. For instance, 100 is <NUMBER HUNDREDS, NUMBER ONE>; 200 is <NUMBER HUNDREDS, NUMBER TWO>; and so forth. Given below are the numbers 110–119:



110 111 112 113 114 115 116 117 118 119

The system theoretically provides for all numbers in the range 1–999 to be written. However, numbers beyond 264 are not attested in the available manuscripts.

The vertical display of Bhaiksuki numbers is beyond the scope of character encoding and is to be managed at the presentation level.

3.14 Editorial marks

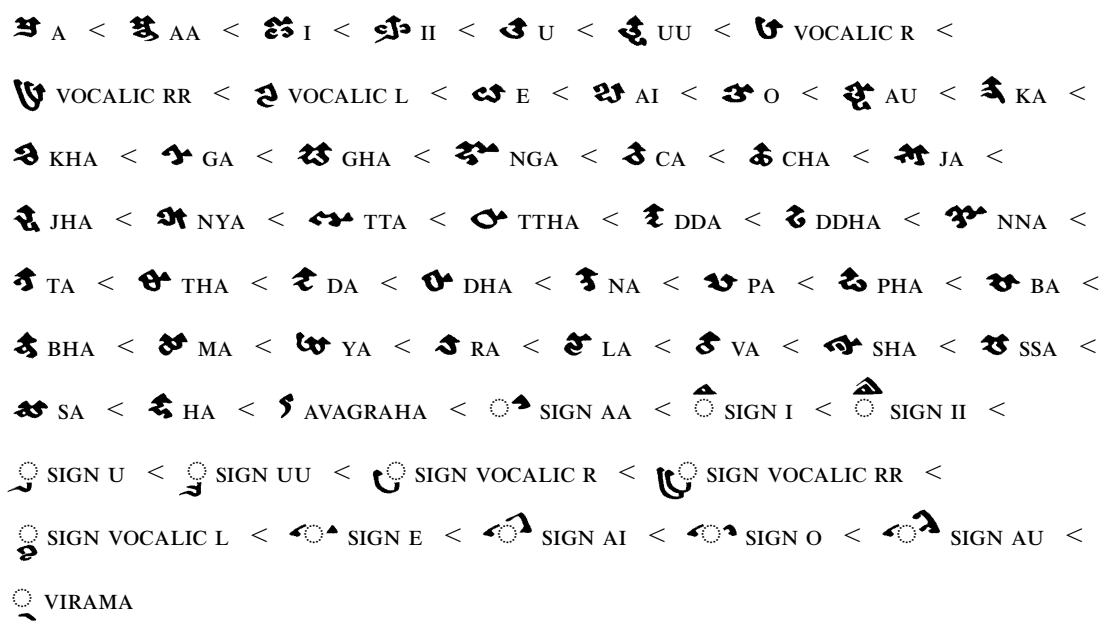
Editorial marks, as the insertion marks  and  , are used commonly in Bhaiksuki manuscripts. These are presently not proposed for encoding. The deletion and insertion marks resemble editorial marks used in other Indic scripts. It may be practical to unify these characters in a separate block of pan-Indic editorial marks.

3.15 Ornaments

In the Bhaiksuki manuscript shown recently in a Chinese documentary two elaborately ornamented *cakra*-s in different colors are visible. These may may be classified as proper ornaments. However, analysis of these ornaments requires access to the newly revealed manuscript. In the *Maṇicūḍajātaka* manuscript the sign  is used in one instance apparently as an ornamental device. It is graphically identical to NUMBER NINE. Until additional information is available, this sign is to be represented using NUMBER NINE.

3.16 Collation

The primary collating order for Bhaiksuki is as follows:



The following characters have secondary weights:

4 Character Data

4.1 Character Properties

The properties for Bhaiksuki characters in the Unicode Character Database format are:

```
11C00;BHAIKSUKI LETTER A;Lo;0;L;;;;;N;;;;;
11C01;BHAIKSUKI LETTER AA;Lo;0;L;;;;;N;;;;;
11C02;BHAIKSUKI LETTER I;Lo;0;L;;;;;N;;;;;
11C03;BHAIKSUKI LETTER II;Lo;0;L;;;;;N;;;;;
11C04;BHAIKSUKI LETTER U;Lo;0;L;;;;;N;;;;;
11C05;BHAIKSUKI LETTER UU;Lo;0;L;;;;;N;;;;;
11C06;BHAIKSUKI LETTER VOCALIC R;Lo;0;L;;;;;N;;;;;
11C07;BHAIKSUKI LETTER VOCALIC RR;Lo;0;L;;;;;N;;;;;
11C08;BHAIKSUKI LETTER VOCALIC L;Lo;0;L;;;;;N;;;;;
11C0A;BHAIKSUKI LETTER E;Lo;0;L;;;;;N;;;;;
11C0B;BHAIKSUKI LETTER AI;Lo;0;L;;;;;N;;;;;
11C0C;BHAIKSUKI LETTER O;Lo;0;L;;;;;N;;;;;
11C0D;BHAIKSUKI LETTER AU;Lo;0;L;;;;;N;;;;;
11C0E;BHAIKSUKI LETTER KA;Lo;0;L;;;;;N;;;;;
11C0F;BHAIKSUKI LETTER KHA;Lo;0;L;;;;;N;;;;;
11C10;BHAIKSUKI LETTER GA;Lo;0;L;;;;;N;;;;;
11C11;BHAIKSUKI LETTER GHA;Lo;0;L;;;;;N;;;;;
11C12;BHAIKSUKI LETTER NGA;Lo;0;L;;;;;N;;;;;
11C13;BHAIKSUKI LETTER CA;Lo;0;L;;;;;N;;;;;
11C14;BHAIKSUKI LETTER CHA;Lo;0;L;;;;;N;;;;;
11C15;BHAIKSUKI LETTER JA;Lo;0;L;;;;;N;;;;;
11C16;BHAIKSUKI LETTER JHA;Lo;0;L;;;;;N;;;;;
11C17;BHAIKSUKI LETTER NYA;Lo;0;L;;;;;N;;;;;
11C18;BHAIKSUKI LETTER TTA;Lo;0;L;;;;;N;;;;;
11C19;BHAIKSUKI LETTER TTHA;Lo;0;L;;;;;N;;;;;
11C1A;BHAIKSUKI LETTER DDA;Lo;0;L;;;;;N;;;;;
11C1B;BHAIKSUKI LETTER DDHA;Lo;0;L;;;;;N;;;;;
11C1C;BHAIKSUKI LETTER NNA;Lo;0;L;;;;;N;;;;;
11C1D;BHAIKSUKI LETTER TA;Lo;0;L;;;;;N;;;;;
11C1E;BHAIKSUKI LETTER THA;Lo;0;L;;;;;N;;;;;
11C1F;BHAIKSUKI LETTER DA;Lo;0;L;;;;;N;;;;;
11C20;BHAIKSUKI LETTER DHA;Lo;0;L;;;;;N;;;;;
11C21;BHAIKSUKI LETTER NA;Lo;0;L;;;;;N;;;;;
11C22;BHAIKSUKI LETTER PA;Lo;0;L;;;;;N;;;;;
11C23;BHAIKSUKI LETTER PHA;Lo;0;L;;;;;N;;;;;
11C24;BHAIKSUKI LETTER BA;Lo;0;L;;;;;N;;;;;
11C25;BHAIKSUKI LETTER BHA;Lo;0;L;;;;;N;;;;;
11C26;BHAIKSUKI LETTER MA;Lo;0;L;;;;;N;;;;;
11C27;BHAIKSUKI LETTER YA;Lo;0;L;;;;;N;;;;;
11C28;BHAIKSUKI LETTER RA;Lo;0;L;;;;;N;;;;;
11C29;BHAIKSUKI LETTER LA;Lo;0;L;;;;;N;;;;;
11C2A;BHAIKSUKI LETTER VA;Lo;0;L;;;;;N;;;;;
11C2B;BHAIKSUKI LETTER SHA;Lo;0;L;;;;;N;;;;;
11C2C;BHAIKSUKI LETTER SSA;Lo;0;L;;;;;N;;;;;
11C2D;BHAIKSUKI LETTER SA;Lo;0;L;;;;;N;;;;;
11C2E;BHAIKSUKI LETTER HA;Lo;0;L;;;;;N;;;;;
11C2F;BHAIKSUKI VOWEL SIGN AA;Mc;0;L;;;;;N;;;;;
11C30;BHAIKSUKI VOWEL SIGN I;Mn;0;NSM;;;;;N;;;;;
11C31;BHAIKSUKI VOWEL SIGN II;Mn;0;NSM;;;;;N;;;;;
11C32;BHAIKSUKI VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
11C33;BHAIKSUKI VOWEL SIGN UU;Mn;0;NSM;;;;;N;;;;;
11C34;BHAIKSUKI VOWEL SIGN VOCALIC R;Mn;0;NSM;;;;;N;;;;;
11C35;BHAIKSUKI VOWEL SIGN VOCALIC RR;Mn;0;NSM;;;;;N;;;;;
11C36;BHAIKSUKI VOWEL SIGN VOCALIC L;Mn;0;NSM;;;;;N;;;;;
```

```
11C38;BHAIKSUKI VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
11C39;BHAIKSUKI VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
11C3A;BHAIKSUKI VOWEL SIGN O;Mc;0;L;;;;;N;;;;;
11C3B;BHAIKSUKI VOWEL SIGN AU;Mc;0;L;;;;;N;;;;;
11C3C;BHAIKSUKI SIGN CANDRABINDU;Mn;0;NSM;;;;;N;;;;;
11C3D;BHAIKSUKI SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11C3E;BHAIKSUKI SIGN VISARGA;Mc;0;L;;;;;N;;;;;
11C3F;BHAIKSUKI SIGN VIRAMA;Mn;9;L;;;;;N;;;;;
11C40;BHAIKSUKI SIGN AVAGRAHA;Lo;0;L;;;;;N;;;;;
11C41;BHAIKSUKI DANDA;Po;0;L;;;;;N;;;;;
11C42;BHAIKSUKI DOUBLE DANDA;Po;0;L;;;;;N;;;;;
11C43;BHAIKSUKI WORD SEPARATOR;Po;0;L;;;;;N;;;;;
11C44;BHAIKSUKI GAP FILLER;Po;0;L;;;;;N;;;;;
11C51;BHAIKSUKI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
11C52;BHAIKSUKI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
11C54;BHAIKSUKI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
11C55;BHAIKSUKI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
11C56;BHAIKSUKI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
11C57;BHAIKSUKI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
11C58;BHAIKSUKI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
11C59;BHAIKSUKI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
11C5A;BHAIKSUKI NUMBER ONE;No;0;L;;;;1;N;;;;;
11C5B;BHAIKSUKI NUMBER TWO;No;0;L;;;;2;N;;;;;
11C5C;BHAIKSUKI NUMBER THREE;No;0;L;;;;3;N;;;;;
11C5D;BHAIKSUKI NUMBER FOUR;No;0;L;;;;4;N;;;;;
11C5E;BHAIKSUKI NUMBER FIVE;No;0;L;;;;5;N;;;;;
11C5F;BHAIKSUKI NUMBER SIX;No;0;L;;;;6;N;;;;;
11C60;BHAIKSUKI NUMBER SEVEN;No;0;L;;;;7;N;;;;;
11C61;BHAIKSUKI NUMBER EIGHT;No;0;L;;;;8;N;;;;;
11C62;BHAIKSUKI NUMBER NINE;No;0;L;;;;9;N;;;;;
11C63;BHAIKSUKI NUMBER TEN;No;0;L;;;;10;N;;;;;
11C64;BHAIKSUKI NUMBER TWENTY;No;0;L;;;;20;N;;;;;
11C65;BHAIKSUKI NUMBER THIRTY;No;0;L;;;;30;N;;;;;
11C66;BHAIKSUKI NUMBER FORTY;No;0;L;;;;40;N;;;;;
11C67;BHAIKSUKI NUMBER FIFTY;No;0;L;;;;50;N;;;;;
11C68;BHAIKSUKI NUMBER SIXTY;No;0;L;;;;60;N;;;;;
11C69;BHAIKSUKI NUMBER SEVENTY;No;0;L;;;;70;N;;;;;
11C6A;BHAIKSUKI NUMBER EIGHTY;No;0;L;;;;80;N;;;;;
11C6B;BHAIKSUKI NUMBER NINETY;No;0;L;;;;90;N;;;;;
11C6C;BHAIKSUKI NUMBER HUNDREDS;No;0;L;;;;100;N;;;;;
```

4.2 Linebreaking

Linebreaking properties of Bhaiksuki characters given in the data format of `LineBreak.txt`:

```
11C00..11C08; AL   # LETTER A .. LETTER VOCALIC L
11C0A..11C2E; AL   # LETTER E .. LETTER HA
11C2F..11C36; CM   # VOWEL SIGN AA .. VOWEL SIGN VOCALIC L
11C38..11C3F; CM   # VOWEL SIGN E .. VIRAMA
11C40; AL          # SIGN AVAGRAHA
11C41..11C43; BA   # DANDA .. GAP FILLER
11C51..11C52; NU   # DIGIT ONE .. DIGIT TWO
11C54..11C59; NU   # DIGIT FOUR .. DIGIT NINE
11C5A..11C6C; AL   # NUMBER ONE .. NUMBER HUNDREDS
```

4.3 Syllabic Categories

Syllabic categories given in the data format of `IndicSyllabicCategory.txt`:

```
# Indic_Syllabic_Category=Bindu
11C3C          ; Bindu              # Mn         SIGN CANDRABINDU
11C3D          ; Bindu              # Mn         SIGN ANUSVARA

# Indic_Syllabic_Category=Visarga
11C3E          ; Visarga            # Mc         SIGN VISARGA

# Indic_Syllabic_Category=Virama
11C3F          ; Virama             # Mn         SIGN VIRAMA

# Indic_Syllabic_Category=Vowel_Independent
11C00..11C0D   ; Vowel_Independent # Lo   [14] LETTER A .. LETTER AU

# Indic_Syllabic_Category=Vowel_Dependent
11C2F          ; Vowel_Dependent    # Mc    [3] VOWEL SIGN AA
11C30..11C36   ; Vowel_Dependent    # Mn    [8] VOWEL SIGN I .. VOWEL SIGN VOCALIC L
11C38..11C3B   ; Vowel_Dependent    # Mc    [2] VOWEL SIGN E .. VOWEL SIGN AU

# Indic_Syllabic_Category=Consonant
11C0E..11C2E   ; Consonant          # Lo   [34] LETTER KA .. LETTER HA
```

4.4 Matra Categories

Matra categories given in the data format of `IndicMatraCategory.txt`:

```
# Indic_Matra_Category=Right
11C2F          ; Right  # Mc       VOWEL SIGN AA

# Indic_Matra_Category=Left_And_Right
11C38          ; Left_And_Right # Mc   VOWEL SIGN E
11C3A          ; Left_And_Right # Mc   VOWEL SIGN O

# Indic_Matra_Category=Top
11C30..11C31   ; Top    # Mn    [2] VOWEL SIGN I .. VOWEL SIGN II

# Indic_Matra_Category=Top_And_Left_And_Right
11C39          ; Top_And_Left_And_Right # Mc   VOWEL SIGN AI
11C3B          ; Top_And_Left_And_Right # Mc   VOWEL SIGN AU

# Indic_Matra_Category=Bottom
11C32..11C36   ; Bottom # Mn   [5] VOWEL SIGN U .. VOWEL SIGN VOCALIC L
11C3F          ; Bottom # Mn       SIGN VIRAMA
```

5 References

Arrow-headed Script Project. `http://www.uni-marburg.de/fb10/iksl/indologie/forschung/drittmittelprojekte/bhaiksuki`

Dimitrov, Dragomir. 2010. *The Bhaikṣukī Manuscript of the Candrālaṃkāra: Study, Script Tables, and Facsimile Edition*. Harvard Oriental Series, 72. Cambridge: Department of Sanskrit and Indian Studies, Harvard University.

Hanisch, Albrecht. 2008. "Sarvarakṣita's *Maṇicūḍajātaka*. Reproduction of the Codex Unicus with Diplomatic Transcript and Palaeographic Introduction to the Bhaikṣukī Script." In *Sanskrit Texts from Giuseppe Tucci's Collection*, Part I. Manuscripta Buddhica, I, p. 195–342. Edited by Harunaga Isaacson and Francesco Sferra. Rome: Istituto Italiano per l'Africa e l'Oriente.

Pandey, Anshuman. 2011. "Preliminary Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646". N4121 L2/11-259 (July 11, 2011). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4121.pdf

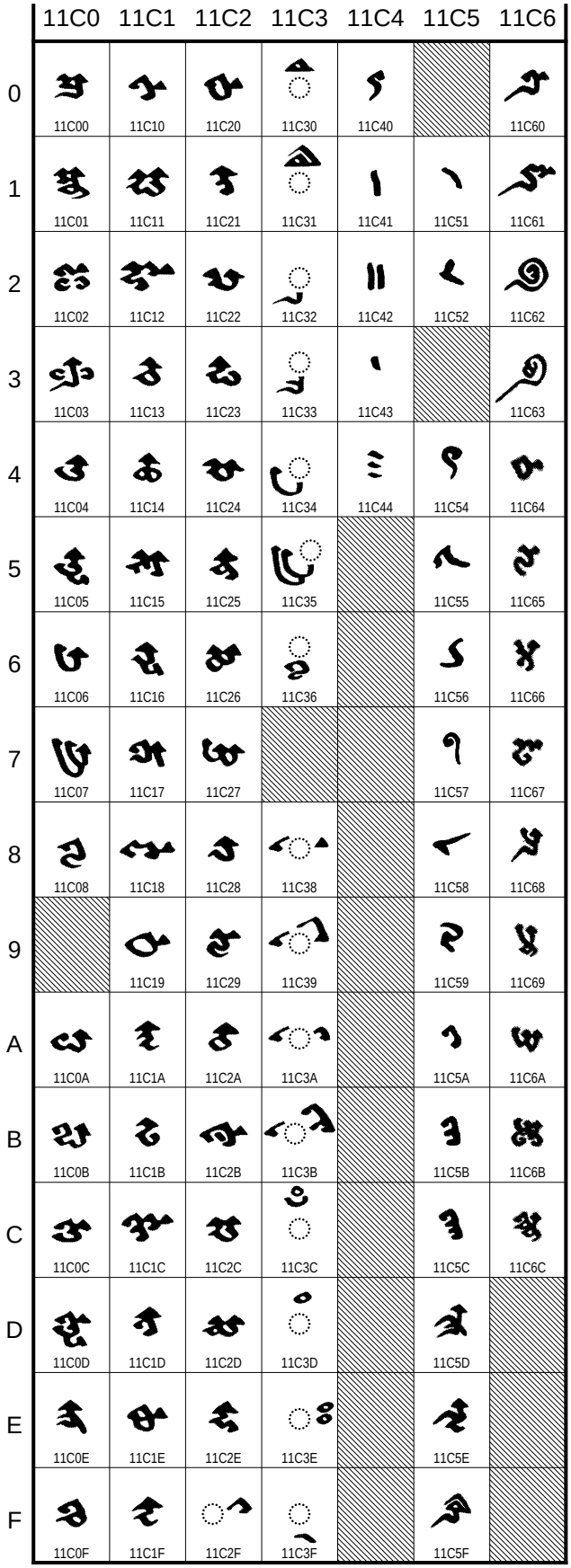
Pandey, Anshuman and Dragomir Dimitrov. 2013a. "Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646". N4469 L2/13-167 (July 22, 2013). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4469.pdf

———. 2013b. "Revised Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646". N4489 L2/13-194 (October 27, 2013). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4489.pdf

6 Acknowledgments

The authors would like to express their gratitude to Shriramana Sharma for discussions regarding the representation of consonant-virama ligatures.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.


11C0 11C1 11C2 11C3 11C4 11C5 11C6
0 11C00 11C10 11C20 11C30 11C40 11C60
1 11C01 11C11 11C21 11C31 11C41 11C51 11C61
2 11C02 11C12 11C22 11C32 11C42 11C52 11C62
3 11C03 11C13 11C23 11C33 11C43 11C63
4 11C04 11C14 11C24 11C34 11C44 11C54 11C64
5 11C05 11C15 11C25 11C35 11C55 11C65
6 11C06 11C16 11C26 11C36 11C56 11C66
7 11C07 11C17 11C27 11C57 11C67
8 11C08 11C18 11C28 11C38 11C58 11C68
9 11C19 11C29 11C39 11C59 11C69
A 11C0A 11C1A 11C2A 11C3A 11C5A 11C6A
B 11C0B 11C1B 11C2B 11C3B 11C5B 11C6B
C 11C0C 11C1C 11C2C 11C3C 11C5C 11C6C
D 11C0D 11C1D 11C2D 11C3D 11C5D
E 11C0E 11C1E 11C2E 11C3E 11C5E
F 11C0F 11C1F 11C2F 11C3F 11C5F

Independent vowels

11C00 BHAIKSUKI LETTER A
11C01 BHAIKSUKI LETTER AA
11C02 BHAIKSUKI LETTER I
11C03 BHAIKSUKI LETTER II
11C04 BHAIKSUKI LETTER U
11C05 BHAIKSUKI LETTER UU
11C06 BHAIKSUKI LETTER VOCALIC R
11C07 BHAIKSUKI LETTER VOCALIC RR
11C08 BHAIKSUKI LETTER VOCALIC L
11C09 <reserved>
11C0A BHAIKSUKI LETTER E
11C0B BHAIKSUKI LETTER AI
11C0C BHAIKSUKI LETTER O
11C0D BHAIKSUKI LETTER AU

Consonants

11C0E BHAIKSUKI LETTER KA
11C0F BHAIKSUKI LETTER KHA
11C10 BHAIKSUKI LETTER GA
11C11 BHAIKSUKI LETTER GHA
11C12 BHAIKSUKI LETTER NGA
11C13 BHAIKSUKI LETTER CA
11C14 BHAIKSUKI LETTER CHA
11C15 BHAIKSUKI LETTER JA
11C16 BHAIKSUKI LETTER JHA
11C17 BHAIKSUKI LETTER NYA
11C18 BHAIKSUKI LETTER TTA
11C19 BHAIKSUKI LETTER TTHA
11C1A BHAIKSUKI LETTER DDA
11C1B BHAIKSUKI LETTER DDHA
11C1C BHAIKSUKI LETTER NNA
11C1D BHAIKSUKI LETTER TA
11C1E BHAIKSUKI LETTER THA
11C1F BHAIKSUKI LETTER DA
11C20 BHAIKSUKI LETTER DHA
11C21 BHAIKSUKI LETTER NA
11C22 BHAIKSUKI LETTER PA
11C23 BHAIKSUKI LETTER PHA
11C24 BHAIKSUKI LETTER BA
11C25 BHAIKSUKI LETTER BHA
11C26 BHAIKSUKI LETTER MA
11C27 BHAIKSUKI LETTER YA
11C28 BHAIKSUKI LETTER RA
11C29 BHAIKSUKI LETTER LA
11C2A BHAIKSUKI LETTER VA
11C2B BHAIKSUKI LETTER SHA
11C2C BHAIKSUKI LETTER SSA
11C2D BHAIKSUKI LETTER SA
11C2E BHAIKSUKI LETTER HA

Dependent vowel signs

11C2F BHAIKSUKI VOWEL SIGN AA
11C30 BHAIKSUKI VOWEL SIGN I
11C31 BHAIKSUKI VOWEL SIGN II
11C32 BHAIKSUKI VOWEL SIGN U
11C33 BHAIKSUKI VOWEL SIGN UU
11C34 BHAIKSUKI SIGN VOCALIC R
11C35 BHAIKSUKI SIGN VOCALIC RR
11C36 BHAIKSUKI SIGN VOCALIC L
11C37 <reserved>
11C38 BHAIKSUKI VOWEL SIGN E
11C39 BHAIKSUKI VOWEL SIGN AI
11C3A BHAIKSUKI VOWEL SIGN O
11C3B BHAIKSUKI VOWEL SIGN AU

Various signs

11C3C BHAIKSUKI SIGN CANDRABINDU
11C3D BHAIKSUKI SIGN ANUSVARA
11C3E BHAIKSUKI SIGN VISARGA
11C3F BHAIKSUKI SIGN VIRAMA
11C40 BHAIKSUKI SIGN AVAGRAHA

Punctuation

11C41 BHAIKSUKI DANDA
11C42 BHAIKSUKI DOUBLE DANDA
11C43 BHAIKSUKI WORD SEPARATOR
11C44 BHAIKSUKI GAP FILLER

Digits

11C50 <reserved>
11C51 BHAIKSUKI DIGIT ONE
11C52 BHAIKSUKI DIGIT TWO
11C53 <reserved>
11C54 BHAIKSUKI DIGIT FOUR
11C55 BHAIKSUKI DIGIT FIVE
11C56 BHAIKSUKI DIGIT SIX
11C57 BHAIKSUKI DIGIT SEVEN
11C58 BHAIKSUKI DIGIT EIGHT
11C59 BHAIKSUKI DIGIT NINE

Numbers

Also known as letter-numerals

11C5A BHAIKSUKI NUMBER ONE
11C5B BHAIKSUKI NUMBER TWO
11C5C BHAIKSUKI NUMBER THREE
11C5D BHAIKSUKI NUMBER FOUR
11C5E BHAIKSUKI NUMBER FIVE
11C5F BHAIKSUKI NUMBER SIX
11C60 BHAIKSUKI NUMBER SEVEN
11C61 BHAIKSUKI NUMBER EIGHT
11C62 BHAIKSUKI NUMBER NINE
11C63 BHAIKSUKI NUMBER TEN
11C64 BHAIKSUKI NUMBER TWENTY
11C65 BHAIKSUKI NUMBER THIRTY
11C66 BHAIKSUKI NUMBER FORTY
11C67 BHAIKSUKI NUMBER FIFTY
11C68 BHAIKSUKI NUMBER SIXTY
11C69 BHAIKSUKI NUMBER SEVENTY
11C6A BHAIKSUKI NUMBER EIGHTY
11C6B BHAIKSUKI NUMBER NINETY
11C6C BHAIKSUKI NUMBER HUNDREDS

Figure 2: Folios 2a and 2b of the *Candrālaṃkāra* in Bhaiksuki (from Dimitrov 2010). Transliteration given in figure 3.

[fol. 2ˡa] śeṣā hiṃsyā **doṣā** rāgādayo yasya sa ○ **pradhvasta** iti *KARTTARI CĀRAMBHE KTAḤ* (1.2.68) ⟪|⟫ pradhvastavāṁś cāsau āśeṣadoṣaś ceti **pradhva**ː⟨2⟩**stāśeṣadoṣaḥ** | itas tataḥ kṣeptuṃ ː ○ prārabdhavān ˡ hiṃsyān ˡ rāgādīn ity arthaḥ | yadvā pradhvaṃsitum ārabdho aśeṣaː⟨3⟩doṣo yeneti karmmaṇi kte aśeṣaśa○bdaḥ sāvaśeṣārthaḥ ⟪|⟫ etenotpāditavodhicitta ity uktaṃ | nanv evaṃbhūtaḥ pṛthagja[no]⟨4⟩pi bhavatīti viśinaṣṭi | samasyante ː ○ satvasantāne prakṣipyante iti **samastāś** ca te **guṇāś** ca ⌈da⌋yāmudito pekṣā⟨5⟩kṣamādayaːːḥ | teṣāṃ **śalanaṃ** śālaḥ ○ | *śala gatau* (Dhātup. 1.572)[72] ghañ{a} | jñānaṃ śobhā vā prāptir vāsyāstīti | athavā taiḥ śali(tap)[i](tā) ⟨6⟩ jñātuṃ | śobhitum āptum vā śīlam asye○ti | etena svārthanirapekṣāvyāhataparārthakāritayā ārya tam uktaṃ | a[ta e]⟨7⟩vāha | **parasmai anuggrahas** tena **dakṣo** vaː○rddh[i]to bodhisatvatvena lokair jñāto vā | *dakṣa vṛddhau gatau ce*ti (Dhātup. 1.446, 1.518)[73] Dhā⌈tu⌋(pāṭh). .. ..[a]⟨8⟩ta evaitasmād guṇatrayayogād **Bu**ː○**ddhaḥ** | budhadhātor dantyauṣṭhavakārapakṣe vaśāditvābhāvād .(ṣa) .. .. + + + + ... ⟨9⟩ jñānaṃ tāṃ jihīte uttarottaram adhiga○cchati pūrvvapūrvvaṃ jahātīti | *JHALO JAŚ* iti (6.3.67) dhasya [da]ḥ (|) .. ..

[fol. 2ˡb] [kṛ]te Buddhaḥ | athavā *KARTTARI CĀRAMBHE KTAḤ* (1.2.68) | svayaṃ ○ boddhuṃ pravṛttaḥ parañ ca vedayitum iti | yadvā budhyate sa [tair iti b]u[d]dhaḥ | [evaṃ](bhūtā)⟨2⟩ya **⌈na⌋amaḥ sade**ti śeṣaḥ ⟪|⟫ bhinnakleśatvena *gṛhī* ː ○ *vā navako pi vā vandyo vratadharair* iti (*Gurupañcāśikā* 4bc)[74] vacanāt | dvāv eva vandyau .. + + .. (vṛddha)[ś] (cet)[i] (asarvva)⟨3⟩janaviṣayatvāc cācāryasya tannamaskāra evā○śaṃsyata ity **astu** ity uktaṃ | Bhagavatas tu trailokyagurutvāt svata e[va s]iddha[tvād] . .. . ⟨4⟩[t]enāpi tadviṣayaḥ śiṣyāṇām āśaṃsyate | a○[śr]utatvāt | **siddham ityādi** vākyasya tadarthatvāc ca | bhikṣutvena la⌈dhatvān n(ā)⌋śas(o) nāpi ⟨5⟩ sadā sarvvakālam astv iti vidhau | ā vodheḥ śara○ṇagamanena vidhyabhāvāt ⟪|⟫ ggranthasyādau namo stu pratipannaśāstraketarobhayeṣāː⟨6⟩m ity abhisandhinā *praiṣānujñāyor lloḍ* ity anye Su○gate namaskāram ādbhaḥ[75] ⟪|⟫ nanv ācāryasyānekaggranthakartṛtvāt ⌈k⌋im iti na Lokānandādau ː ⟨7⟩ ṭīkā kriyata iti stūyamānaśālitayā ○ śāstraṃ stauti **jayatī**tyādi | Candranāmnā praṇītaś **Candraḥ** prabhākaravata upa[ca]⟨8⟩ryamāṇasya dhvanayo bhavanti yathā *yaṣṭī*⟪*ḥ*⟫ *pra○veśaye*ti[76] puṃliṅgaḥ śāstravācī | ata eva pūrvvaśloke nāmānabhidhāne [pi] viś[e]⟨9⟩ṣa⌈par⌋iggrahaḥ | athavā sakalalakṣāvabhāsa○katvāc candayatīti candraḥ | *cadi āhlādane* | *dīptau ce*ti (*Dhātupradīpa* 1.55) Raki[77] | katham avabhāsaya

Figure 3: Transliteration of Folio 2 of the *Candrālaṃkāra* from figure 2.

Figure 4: A statue of Buddha from Gaya, Bihar with a Bhaiksuki inscription (enlarged) on the underside of the base.

Figure 5: Folios showing the 'Sindhura' script with Tibetan correspondences.

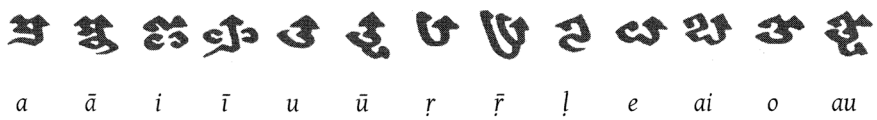


Figure 6: Bhaiksuki vowels (from Dimitrov 2010: 75).

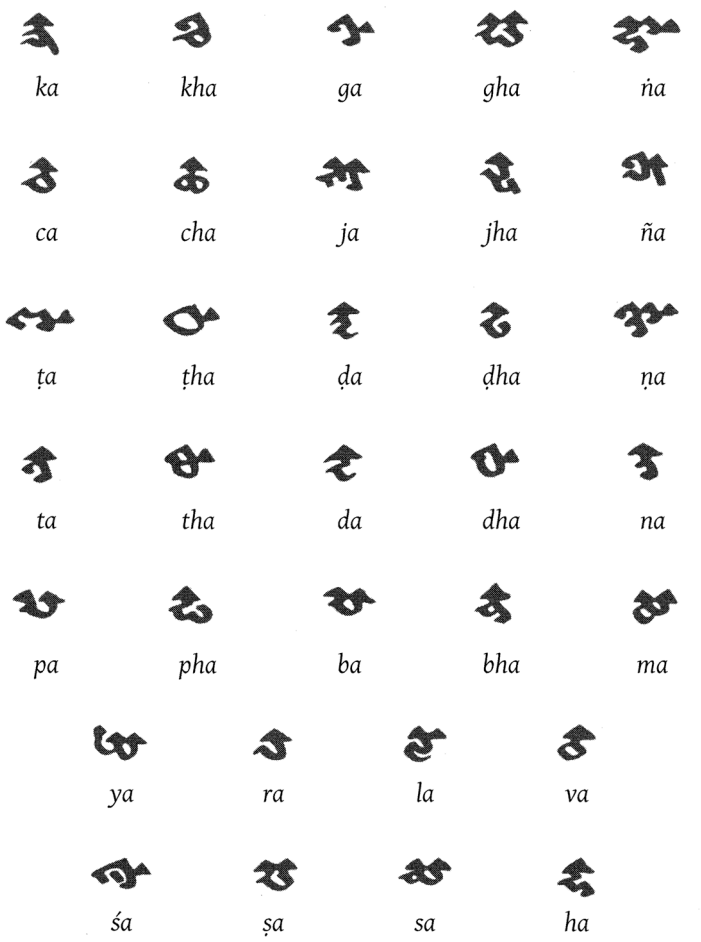


Figure 7: Bhaiksuki consonants (from Dimitrov 2010: 75).

Figure 8: Consonant-vowel combinations with VOWEL SIGN AA (from Dimitrov 2010: 85).

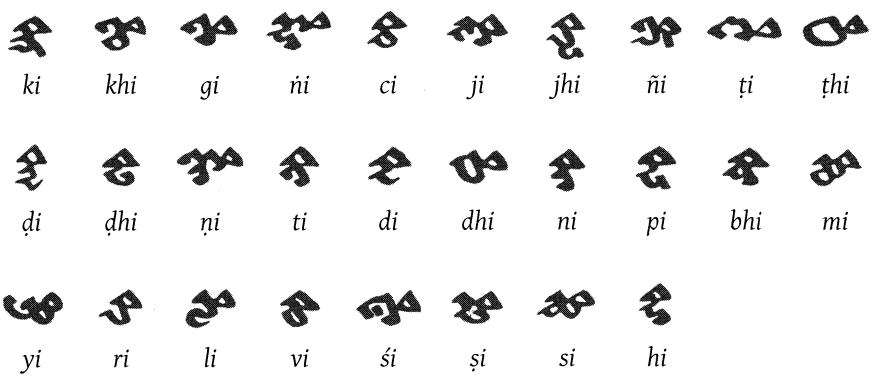


Figure 9: Consonant-vowel combinations with VOWEL SIGN I (from Dimitrov 2010: 86).

Figure 10: Consonant-vowel combinations with VOWEL SIGN II (from Dimitrov 2010: 86).

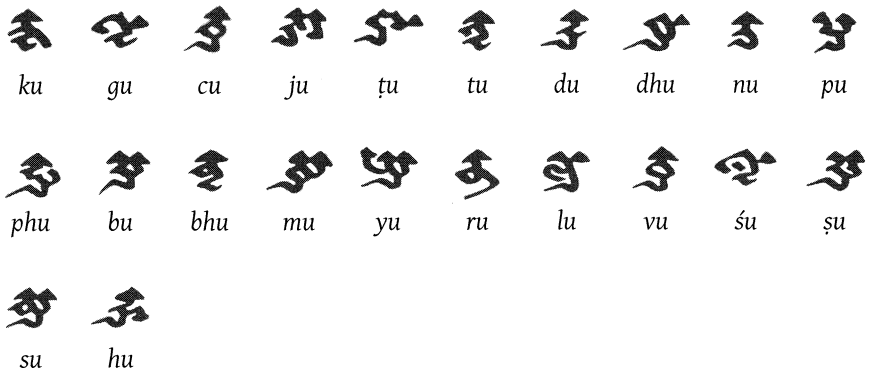


Figure 11: Consonant-vowel combinations with VOWEL SIGN U (from Dimitrov 2010: 86).

Figure 12: Consonant-vowel combinations with VOWEL SIGN UU (from Dimitrov 2010: 87).

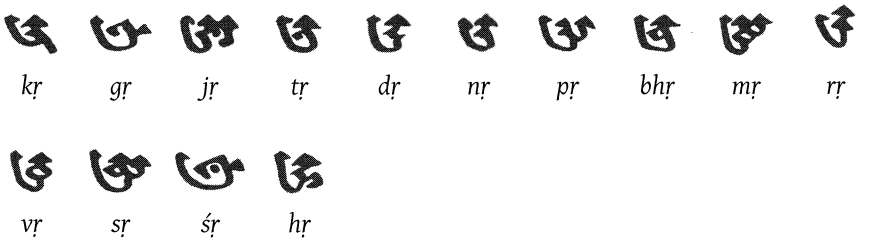


Figure 13: Consonant-vowel combinations with VOWEL SIGN VOCALIC R (from Dimitrov 2010: 87).

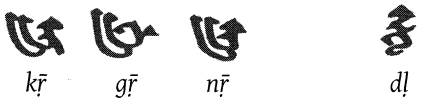


Figure 14: Consonant-vowel combinations with VOWEL SIGN VOCALIC RR and VOWEL SIGN VOCALIC L (from Dimitrov 2010: 87).

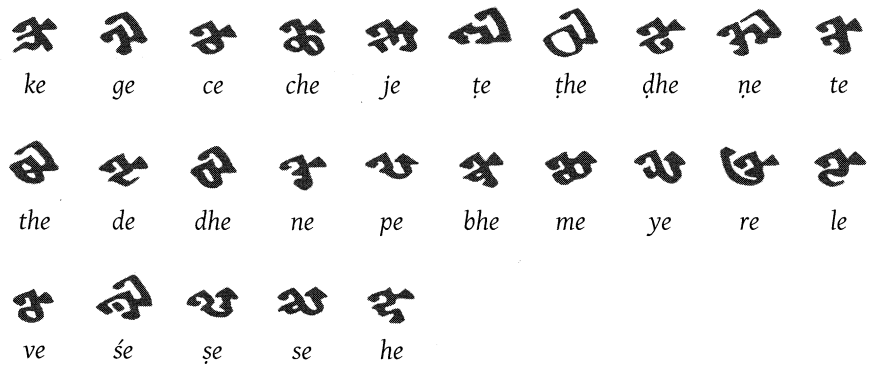


Figure 15: Consonant-vowel combinations with VOWEL SIGN E (from Dimitrov 2010: 87).

Figure 16: Consonant-vowel combinations with VOWEL SIGN AI (from Dimitrov 2010: 88).

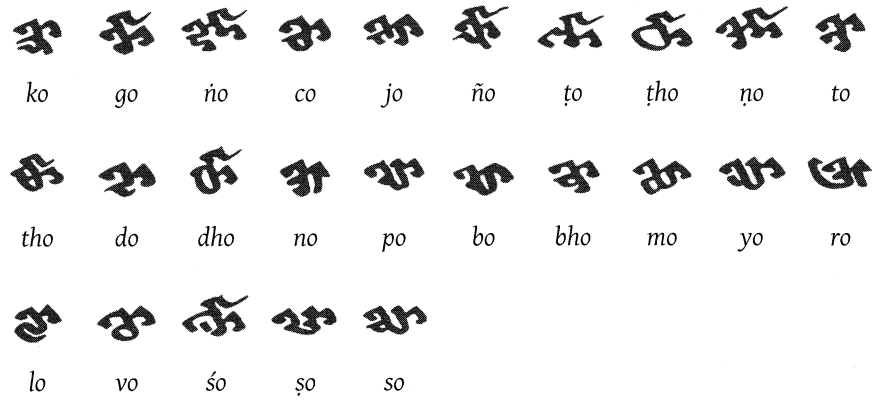


Figure 17: Consonant-vowel combinations with VOWEL SIGN O (from Dimitrov 2010: 88).

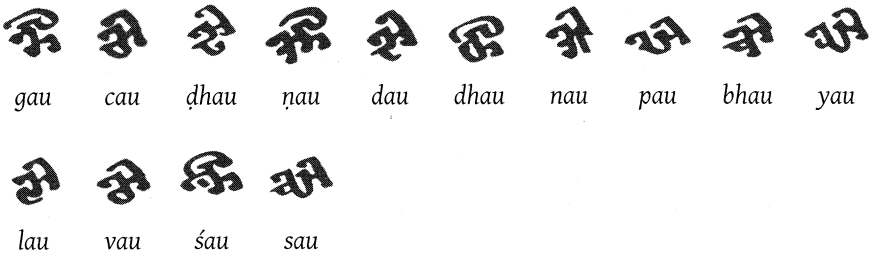


Figure 18: Consonant-vowel combinations with VOWEL SIGN AU (from Dimitrov 2010: 88).

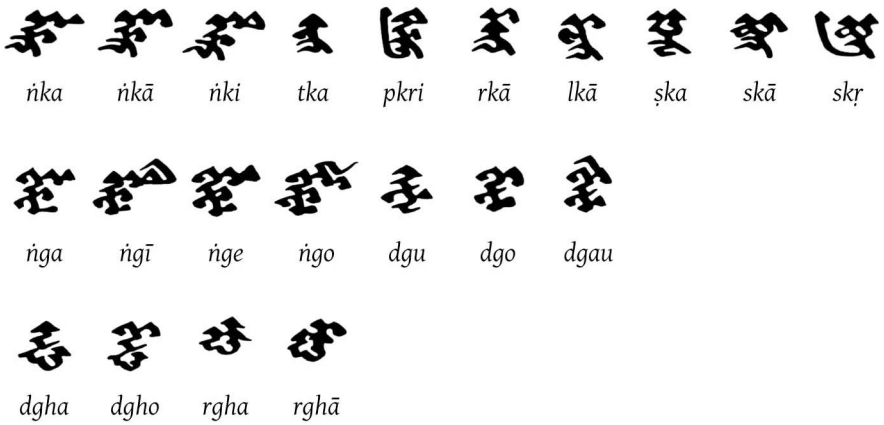


Figure 19: Examples of conjuncts with velar consonants (from Dimitrov 2010: 90).

Figure 20: Examples of conjuncts with palatal consonants (from Dimitrov 2010: 90–91).

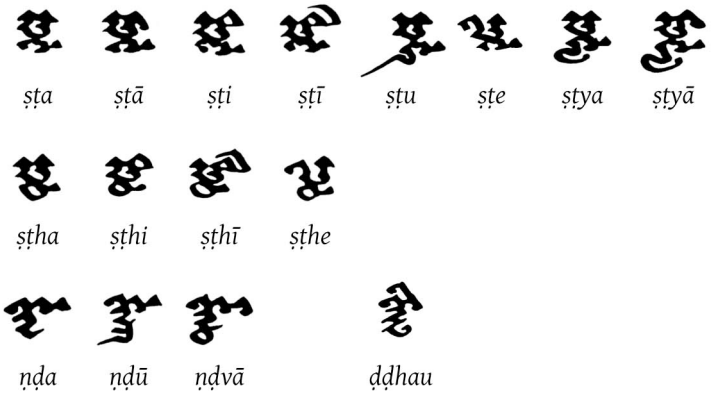


Figure 21: Examples of conjuncts with retroflex consonants (from Dimitrov 2010: 91).

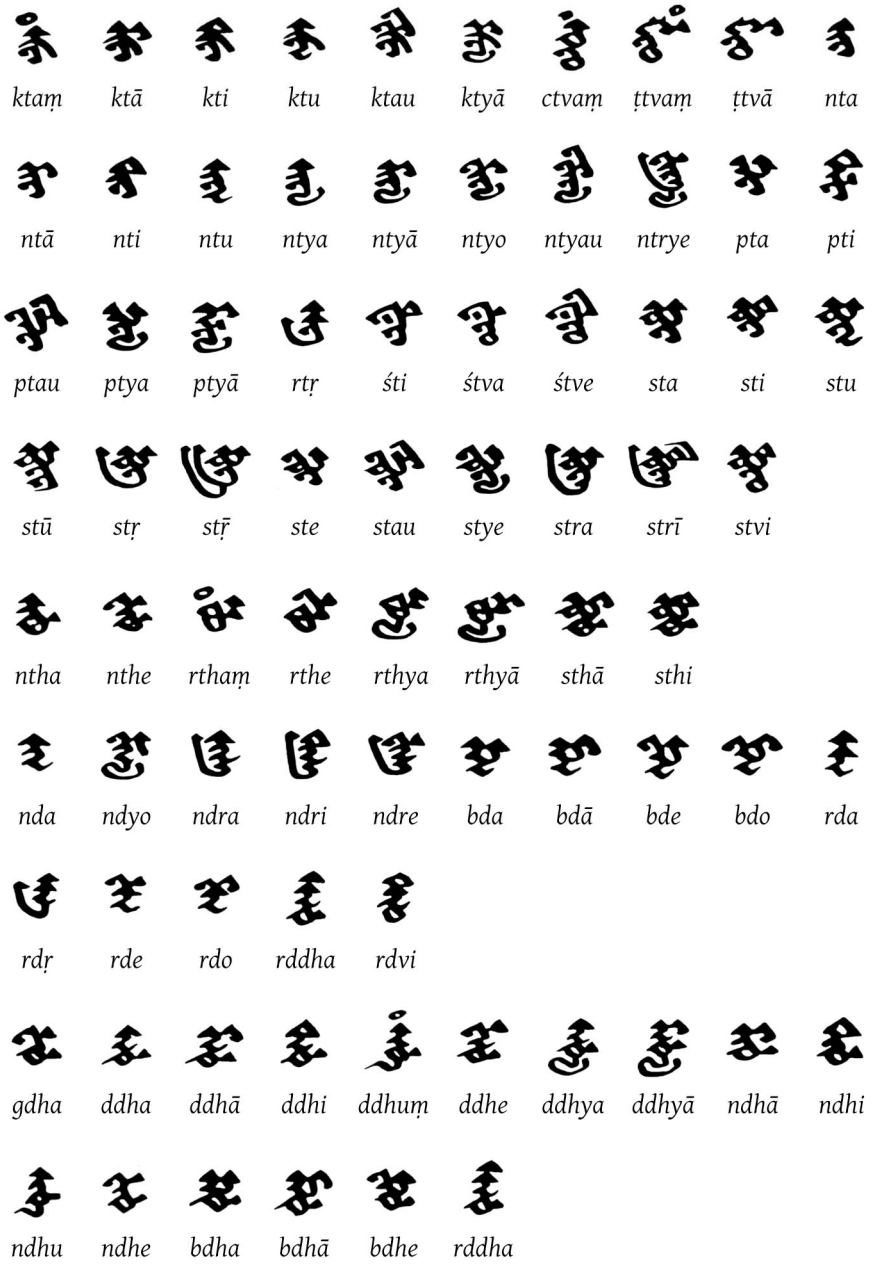


Figure 22: Examples of conjuncts with dental consonants (from Dimitrov 2010: 91–92).

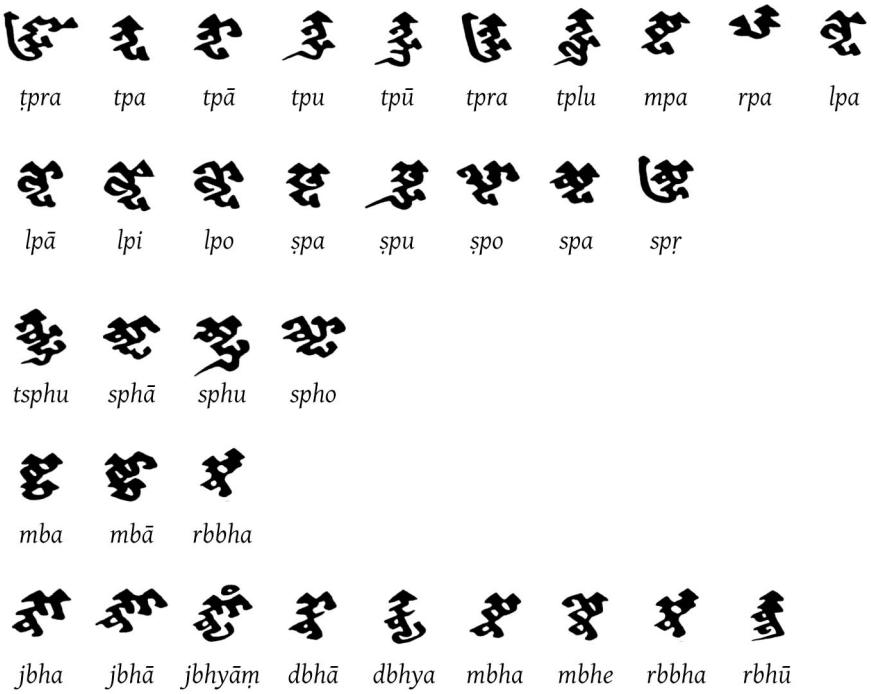


Figure 23: Examples of conjuncts with labial consonants (from Dimitrov 2010: 93).

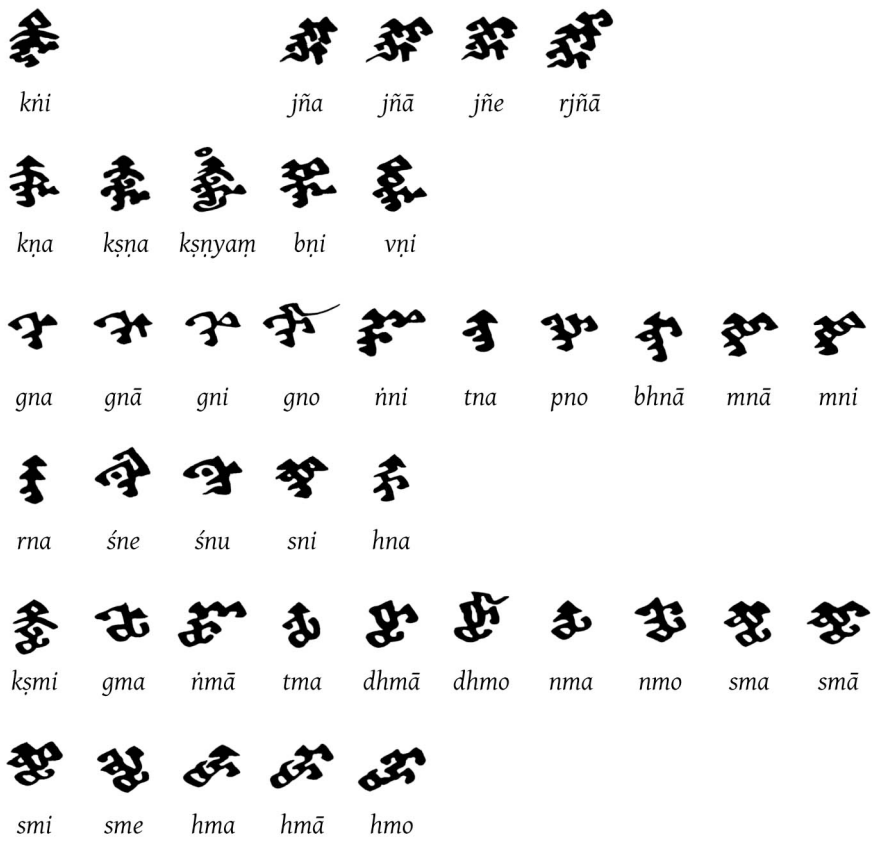


Figure 24: Examples of conjuncts with nasal consonants (from Dimitrov 2010: 94).

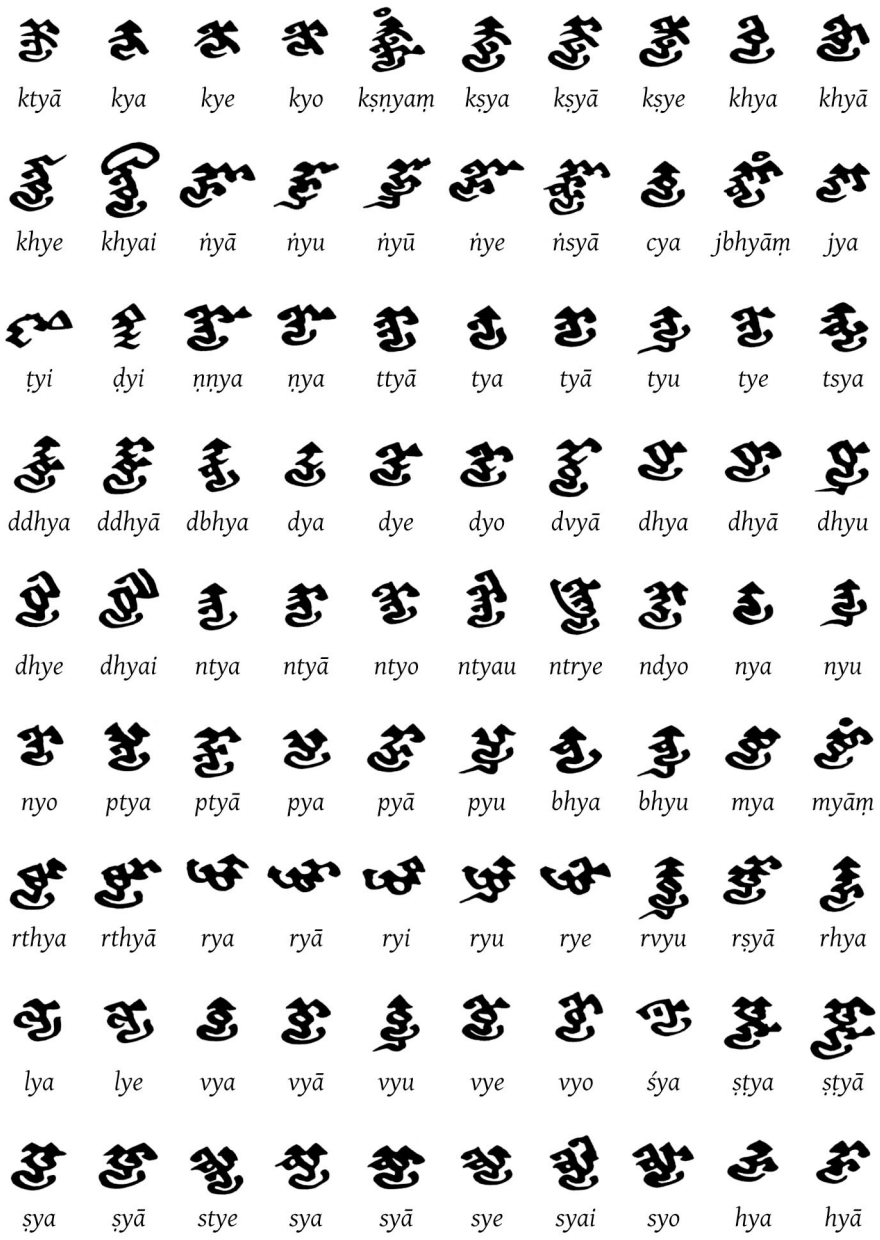


Figure 25: Examples of conjuncts with YA (from Dimitrov 2010: 95).

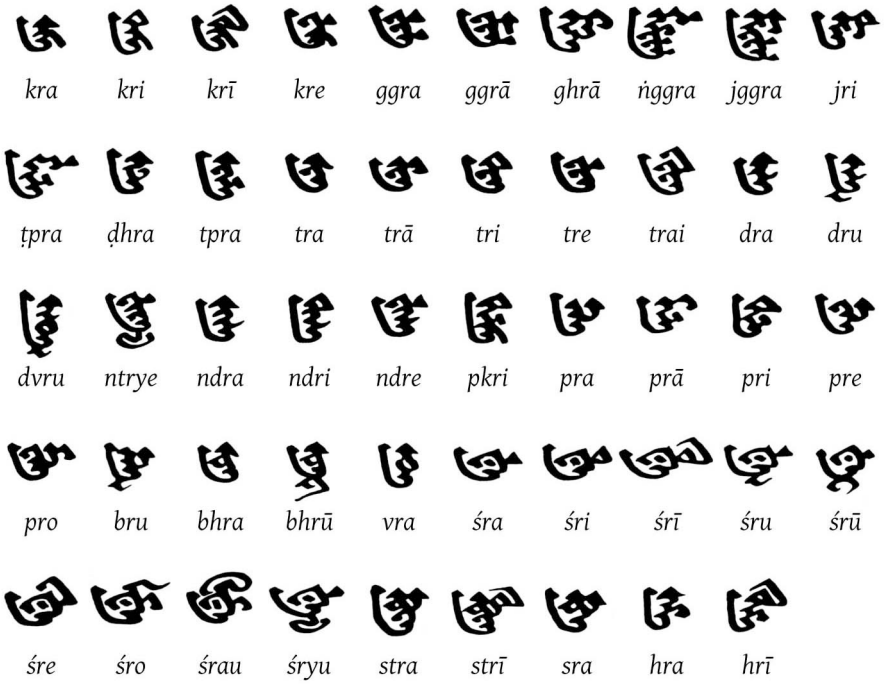


Figure 26: Examples of conjuncts with RA (from Dimitrov 2010: 96).

Figure 27: Examples of conjuncts with LA (from Dimitrov 2010: 96).

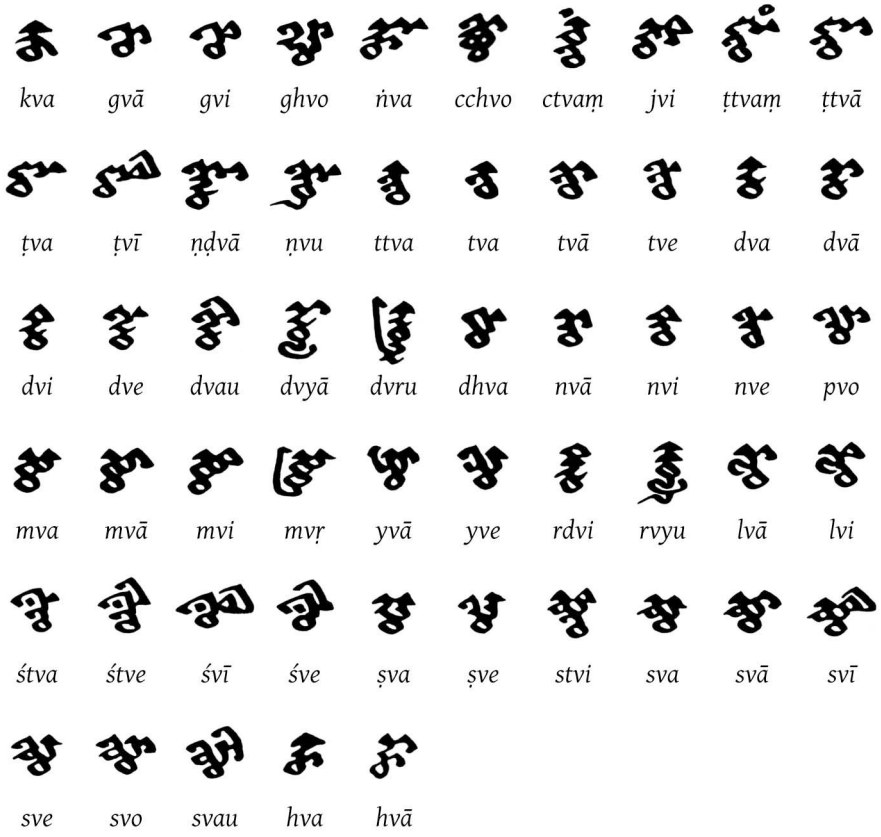


Figure 28: Examples of conjuncts with VA (from Dimitrov 2010: 97).

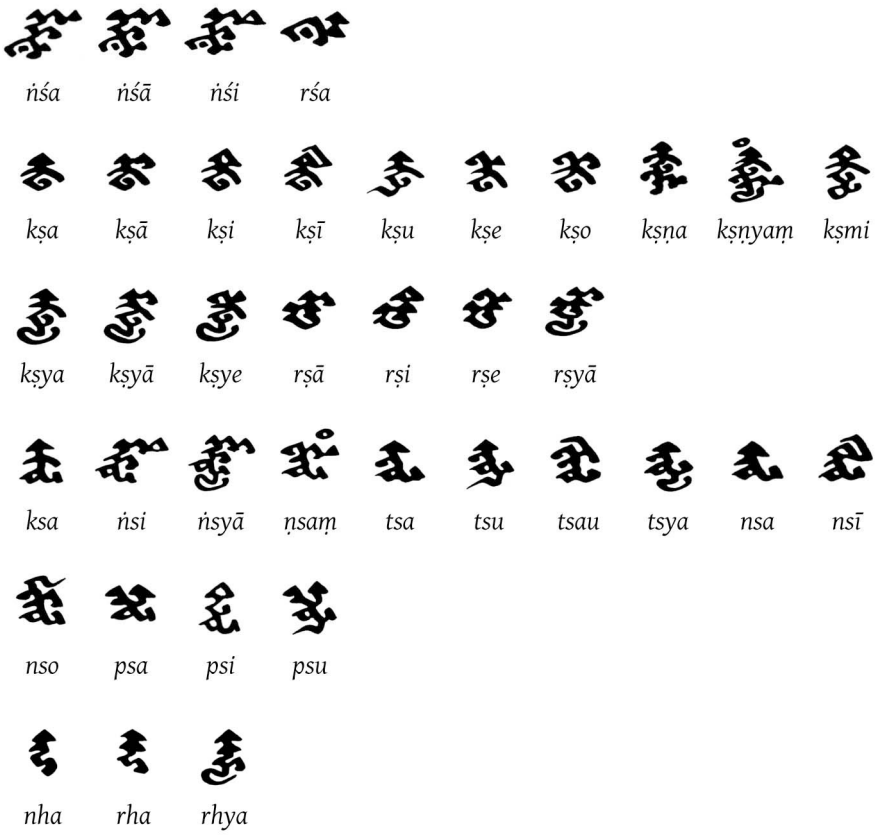


Figure 29: Examples of conjuncts with sibilant consonants (from Dimitrov 2010: 97–98).

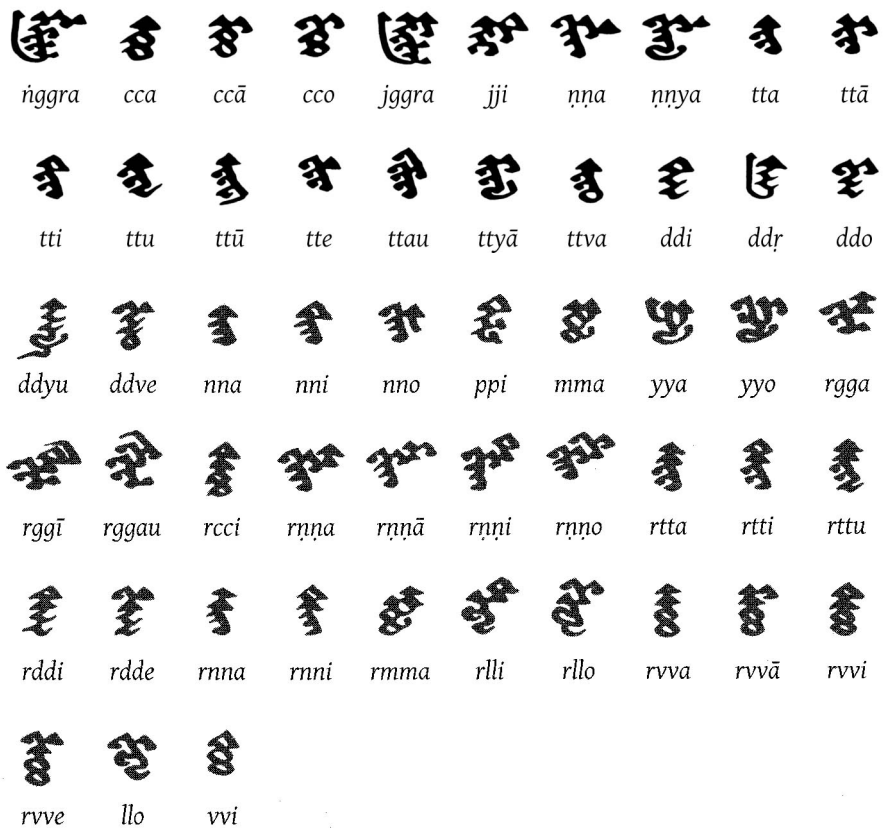


Figure 30: Examples of conjuncts with geminate consonants (from Dimitrov 2010: 98–99).

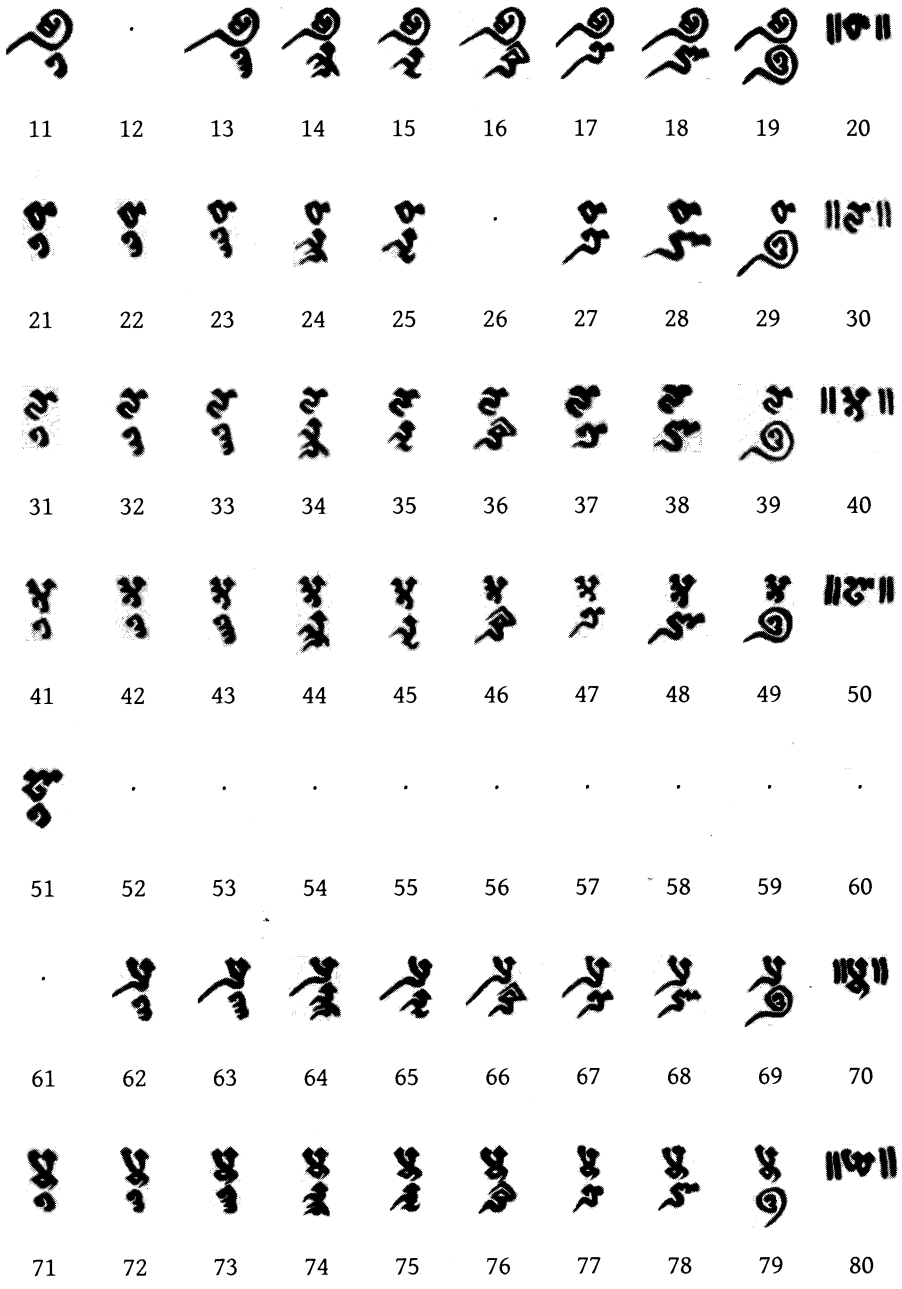


Figure 31: Some Bhaiksuki numbers (from Dimitrov 2010: 54).

	Bhai.	Sid.	Sha.	Dev.		Bhai.	Sid.	Sha.	Dev.
A				अ					
AA				आ	SIGN AA				
I				इ	SIGN I				
II				ई	SIGN II				
U				उ	SIGN U				
UU				ऊ	SIGN UU				
VOC.R				ऋ	SIGN VOC.R				
VOC.RR				ॠ	SIGN VOC.RR				
VOC.L				ऌ	SIGN VOC.L		—		
VOC.LL	—			ॡ	SIGN VOC.LL	—	—		
E				ए	SIGN E				
AI				ऐ	SIGN AI				
O				ओ	SIGN O				
AU				औ	SIGN AU				

Table 7: Comparison of vowel letters and signs of Bhaiksuki, Siddham, Sharada, and Devanagari.

	Bhai.	Sid.	Sha.	Dev.		Bhai.	Sid.	Sha.	Dev.
KA			क	क	DA			ऽ	द
KHA			ष	ख	DHA			ऽ	ध
GA			ग	ग	NA			न	न
GHA			ध	घ	PA			प	प
NGA			ए	ङ	PHA			ढ	फ
CA			ब	च	BA			ब	ब
CHA			ळ	छ	BHA			भ	भ
JA			ल	ज	MA			म	म
JHA			ञ	झ	YA			य	य
NYA			ह	ञ	RA			ा	र
TTA			ट	ट	LA			ल	ल
TTHA			ठ	ठ	VA			व	व
DDA			ड	ड	SHA			श	श
DDHA			ढ	ढ	SSA			ष	ष
NNA			ण	ण	SA			स	स
TA			उ	त	HA			ऊ	ह
THA			घ	थ					

Table 8: Comparison of consonant letters of Bhaiksuki, Siddham, Sharada, and Devanagari.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** ***Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646***
2. Requester's name: *Script Encoding Initiative (SEI) / Anshuman Pandey (pandey@umich.edu) / Dragomir Dimitrov (dimitrov@staff.uni-marburg.de)*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2014-01-27*
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: *Bhaiksuki*
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block:
2. Number of characters in proposal: *94*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary ; B.1-Specialized (small collection) *X* ; B.2-Specialized (large collection) ;
   C-Major extinct ; D-Attested extinct ; E-Minor extinct ;
   F-Archaic Hieroglyphic or Ideographic ; G-Obscure or questionable usage symbols
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 (pandey@umich.edu)
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: N4102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *No*
   If YES explain
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? *Yes*
   If YES, with whom?
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare) *Rare*
   Reference:
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference:
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *N/A*
   If YES, is a rationale provided?
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *Yes*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? *No*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *Yes*
    If YES, is a rationale for such use provided? *Yes*
    If YES, reference: *Combining signs*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *Yes*
    If YES, describe in detail (include attachment if necessary) *Virama; see text of the proposal*
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: