Proposal to Encode the Masaram Gondi Script in Unicode

Anshuman Pandey  
Department of Linguistics  
University of California, Berkeley  
Berkeley, California, U.S.A.  
anshuman.pandey@berkeley.edu

April 15, 2015

1 Introduction

This is a proposal to encode the Gondi script created by Mangal Singh Masaram in 1918 in Unicode. It replaces and supersedes the following documents:

- L2/10-207 “Preliminary Proposal to Encode the Gondi Script in the UCS”
- L2/12-235 “Revised Preliminary Proposal to Encode the Gondi Script”
- L2/15-005 “Proposal to Encode the Gondi Script”

This document provides a description of the writing system, a code chart and names list, character properties, and specimens that illustrate letterforms and usage. It is a revision of L2/15-005 and contains several changes to the encoding of Gondi proposed in that document. The major changes are as follows:

- The block name has been changed from ‘Gondi’ to ‘Masaram Gondi’
- Separate encoding of repha and ra-kara for cluster-initial and cluster-final forms of ra
- Redefinition of virama specifically as a control character used only for producing conjuncts
- Addition of halanta as a combining sign used solely for silencing the inherent vowel
- Addition of candra for transcribing foreign vowel sounds

These changes were introduced as a result of discussions with experts such as Mukund Gokhale, members of the user community and with Unicode engineers at Google and Microsoft.

The ‘Masaram Gondi’ script is graphically and structurally distinct from another Gondi script known as ‘Gunjala Gondi’. This latter script has been proposed for encoding in Unicode (see Pandey 2015b).

The symbol ☩ appears upon several sources containing Masaram’s script and in various Gond religious contexts, in general. This symbol represents persapen, or the supreme spirit, in the indigenous Gond philosophical system known as koya punem. This symbol has been proposed for encoding in the ‘Miscellaneous Symbols and Pictographs’ block in Unicode, where several religious symbols are encoded (see Pandey 2015c).
2 Background

The script described here was invented by Munshi Mangal Singh Masaram of Kochewada, Balaghat District, Madhya Pradesh, India in 1918. It has no genetic relationship to other writing systems, but it is based upon the Brahmi model. The script was designed for writing Gondi (ISO 639-3: gon), a Dravidian language spoken by 2.6 million people, primarily in Madhya Pradesh and Maharashtra, with some speakers in Andhra Pradesh and Chhattisgarh. The language is generally written in both Devanagari and Telugu. Manuscripts containing yet another script, which appears to have been graphically inspired by these Modi writing systems, was found in Gunjala in the Adilabad district of Telangana Masaram’s Gondi script is actively used today for hand-written and printed materials. Fonts have been developed for the production of books. In 2011, the Akhil Gondvana Gondi Sahitya Parishad (Chandagadh, Maharashtra) passed a resolution adopting Masaram’s script as the official script of the Gondi language.

Masaram’s script has been slightly expanded and revised over the years in order to meet the needs and preferences of modern users. Innovations include the addition of new consonant letters, vowel signs, modifier signs, and the adoption of a Devanagari-style halanta for indicating the absence of the inherent vowel. Some of these new characters are included in the proposed repertoire, while others are not (see section 4.11).

3 Script Details

3.1 Name

Earlier versions of this proposal referred to name of the script block as ‘Gondi’. While the script is certainly used for writing Gondi, it is one of many scripts used for the language. Moreover, the ‘Gunjala Gondi’ script is also associated with the language and culture of the Gonds. Given this, it is appropriate to assign an identifier for the script block that precisely defines which ‘Gondi’ script is contained within that block. A designation that includes the name of the script’s creator seems appropriate. For this reason, the name of the script block in Unicode is ‘Masaram Gondi’ and the names of characters reflect the block name. Users may refer to the script as ‘Gondi’ or by whatever name they prefer.

3.2 Structure

Masaram’s Gondi script is an alphasyllabic writing system that is written from left to right. Consonant letters possess the inherent vowel $a$, which is graphically represented by a horizontal stroke that extends rightward from the right edge of each consonant letter. A bare consonant is represented by removing this stroke, particularly in the rendering of consonant clusters. Some modern users represent a word-final bare consonant by writing the halanta beneath the stroke of the consonant letter. Consonant clusters are represented as conjuncts, which are rendered as a linear sequence using bare forms for all consonants except for the final, which occurs in its regular form. There are some exceptions to this rule, namely the forms of ra when in occurs in initial and final positions in a cluster. Independent and initial vowels are written using vowel letters, while medial and final vowels are expressed using dependent signs. There is no mātrā reordering.

3.3 Character Repertoire

A total of 75 characters are proposed for encoding in the ‘Masaram Gondi’ script block. A code chart and names list are attached. Names for characters follow the UCS convention for Brahmi-based scripts and align with the Latin transliteration of Devanagari analogues for Gondi letters given by B. S. Masaram (1951).
3.4 Glyphic Representations

The glyphic representations of some consonant letters have changed since the invention of the script. The majority of changes are observed with consonant letters. These differences result from the simplification of glyphs for ease of writing, i.e. sets of independent circles being joined into a single-stroked loop. Representative glyphs are based upon forms shown in published script primers and reflect modern preferences.

4 Proposed Encoding

4.1 Vowel Letters

Ten vowel letters are proposed for encoding:

\[
\begin{align*}
\text{𑮐} & \quad \text{MASARAM GONDI LETTER A} \\
\text{𑮑} & \quad \text{MASARAM GONDI LETTER AA} \\
\text{𑮒} & \quad \text{MASARAM GONDI LETTER I} \\
\text{𑮓} & \quad \text{MASARAM GONDI LETTER II} \\
\text{𑮔} & \quad \text{MASARAM GONDI LETTER U} \\
\text{𑮕} & \quad \text{MASARAM GONDI LETTER UU} \\
\text{𑮖} & \quad \text{MASARAM GONDI LETTER E} \\
\text{𑮗} & \quad \text{MASARAM GONDI LETTER AI} \\
\text{𑮘} & \quad \text{MASARAM GONDI LETTER O} \\
\text{𑮙} & \quad \text{MASARAM GONDI LETTER AU}
\end{align*}
\]

Masaram’s script does not have independent letters or dependent signs for the Dravidian long vowels /eː/ and /oː/, which correspond to Telugu Ṇ ῦ and Ṇ ῦ.

4.2 Vowel Signs

Ten dependent vowel signs are proposed for encoding:

\[
\begin{align*}
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN AA} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN I} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN II} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN U} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN UU} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN VOCALIC R} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN E} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN AI} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN O} \\
\text{◌} & \quad \text{MASARAM GONDI VOWEL SIGN AU}
\end{align*}
\]

Vowel signs are written above and below the horizontal stroke of a consonant letter:

\[
\begin{align*}
\text{ka} & \quad \text{kā} & \quad \text{ki} & \quad \text{kī} & \quad \text{ku} & \quad \text{kū} & \quad \text{kr} & \quad \text{ke} & \quad \text{kai} & \quad \text{ko} & \quad \text{kau}
\end{align*}
\]

They are represented in encoded text as follows:

\[
\begin{align*}
kā & \quad \text{<>} <\text{骼 KA, ◌ SIGN AA}> \\
\text{ki} & \quad \text{<>} <\text{骼 KA, ◌ SIGN I}>
\end{align*}
\]
There is no independent analogue for vowel sign vocalic R. The independent form of this vowel is represented using a consonant-vowel combination composed with the letter RA:

\[ r \]  <\- KA, \+ VOWEL SIGN VOCALIC R>

4.3 Consonants

There are 34 consonant letters:

| \- | MASARAM GONDI LETTER KA | \- | MASARAM GONDI LETTER DA |
| \- | MASARAM GONDI LETTER KHA | \- | MASARAM GONDI LETTER DHA |
| \- | MASARAM GONDI LETTER GA | \- | MASARAM GONDI LETTER NA |
| \- | MASARAM GONDI LETTER GHA | \- | MASARAM GONDI LETTER PA |
| \- | MASARAM GONDI LETTER NGA | \- | MASARAM GONDI LETTER PHA |
| \- | MASARAM GONDI LETTER CA | \- | MASARAM GONDI LETTER BA |
| \- | MASARAM GONDI LETTER CHA | \- | MASARAM GONDI LETTER BHA |
| \- | MASARAM GONDI LETTER JA | \- | MASARAM GONDI LETTER MA |
| \- | MASARAM GONDI LETTER JHA | \- | MASARAM GONDI LETTER YA |
| \- | MASARAM GONDI LETTER NYA | \- | MASARAM GONDI LETTER RA |
| \- | MASARAM GONDI LETTER TTA | \- | MASARAM GONDI LETTER LA |
| \- | MASARAM GONDI LETTER TTHA | \- | MASARAM GONDI LETTER VA |
| \- | MASARAM GONDI LETTER DDA | \- | MASARAM GONDI LETTER SHA |
| \- | MASARAM GONDI LETTER DDHA | \- | MASARAM GONDI LETTER SSA |
| \- | MASARAM GONDI LETTER NNA | \- | MASARAM GONDI LETTER SA |
| \- | MASARAM GONDI LETTER TA | \- | MASARAM GONDI LETTER HA |
| \- | MASARAM GONDI LETTER THA | \- | MASARAM GONDI LETTER LLA |

The letter \= LL is not part of Masaram’s original script. It was introduced by modern users in order to properly represent Marathi \= la (see figure 19 for an example of LL in usage).
4.4 Vowel modifiers

Three vowel modifiers are proposed for encoding:

- ơ  MASARAM GONDI SIGN ANUSVARA
- ơ’  MASARAM GONDI SIGN VISARGA
- ơ”  MASARAM GONDI SIGN CANDRA

Anusvara The sign ơ’ ANUSVARA is used for marking nasalization. It is placed above the horizontal stroke of a consonant. Its position differs slightly in different sources. In some documents the position is altered by the presence of an accompanying above-base vowel sign:

As shown above, the sign ANUSVARA occurs in its normal position when there is no vowel sign or the vowel sign is positioned below the stroke. Its position is raised when ơ’ VOWEL SIGN AA is present. It is placed to the right of the following: ơ’ VOWEL SIGN I, ơ’ VOWEL SIGN II, ơ’ VOWEL SIGN U, ơ’ VOWEL SIGN UU, ơ’ VOWEL SIGN E, ơ’ VOWEL SIGN AI. It is placed to the right and raised higher when it occurs with ơ” VOWEL SIGN O, ơ” VOWEL SIGN AU. Some modern users prefer a more stationary position for ANUSVARA and place it above the body of the consonant:

These positional preferences are to be managed in the font. The sign ANUSVARA is used in encoded text as shown below. It is always placed after a vowel sign in the encoded sequence:

Visarga The ơ” VISARGA is used for the representation of Sanskrit words. It is written above the horizontal line of a consonant letter. When occurring with vowel signs its position is adjusted as follows:

Some modern sources show the ơ” VISARGA written as the glyphic variant ơ”. It is placed after the stroke:

The VISARGA is used in encoded text as follows:
Proposal to Encode the Masaram Gondi Script in Unicode

Anshuman Pandey

\[\text{kāḥ} \ 0^e <\text{O}^e \text{ KA, O” VISARGA}>\]

\[\text{kāḥ} \ 0^e <\text{O}^e \text{ KA, O” VOWEL SIGN AA, O” VISARGA}>\]

**Candra**  The sign \(\text{♀} \text{ CANDRA}\) is used for transcribing vowel sounds that do not occur in Gondi. Examples are given in figure 25. It is derived from the sign \(\text{♀} \text{ used in Devanagari orthography for Marathi}\) for representing the English vowel sounds /æ/ and /ɔ/. The CANDRA is written above the horizontal line of a consonant letter, and is used as follows:

\[
/\text{æ}/ \quad \text{♀} \quad \langle \text{L E, O” CANDRA}> \\
/\text{ɔ}/ \quad \text{♀} \quad \langle \text{♀ AA, O” CANDRA}> \\
/\text{kæ}/ \quad \text{♀} \quad \langle \text{O}^e \text{ KA, O” CANDRA}> \\
/\text{kɔ}/ \quad \text{♀} \quad \langle \text{O}^e \text{ KA, O” VOWEL SIGN AA, O” SIGN CANDRA}> \\
\]

In initial and independent contexts, CANDRA is used only with the vowel letters \(\text{L}\) and \(\text{♀}\) for representing /æ/ and /ɔ/, respectively. The combination \(\text{♀} \text{ L} \) corresponds to \(\text{♀} \ U+090D \text{ DEVANAGARI LETTER CANDRA E}\), while \(\text{♀} \text{ L} \) corresponds to \(\text{♀} \ U+0911 \text{ DEVANAGARI LETTER CANDRA O}\). In dependent contexts, the CANDRA combines with a consonant letter for the sound /æ/, and with the VOWEL SIGN AA for the sound /ɔ/. The \(\text{♀} \text{ CANDRA}\) corresponds to \(\text{♀} \ U+0945 \text{ DEVANAGARI VOWEL SIGN CANDRA E}\), while the sequence \(\langle\text{O} \text{ GONDI VOWEL SIGN AA, O” CANDRA}>\) corresponds to \(\text{♀} \ U+0949 \text{ DEVANAGARI VOWEL SIGN CANDRA O}\). Although the Gondi sequences correspond to atomic characters in Devanagari, there is no need to encode such precomposed letters and signs with CANDRA for Gondi.

Although the contexts in which CANDRA is used are limited, the sign technically may be used with any letter and may occur with any vowel sign. It would be positioned with other vowel signs as follows:

\[0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \]

**4.5 Nukta**

The \(\text{♀} \text{ MASARAM GONDI SIGN NUKTA}\) is used for representing sounds that are not native to the Gondi language. It is written beneath the horizontal stroke of a consonant:

\[0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \]

Some users prefer to position the NUKTA below the body of the consonant letter:

\[0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \ 0^e \]

These positional preferences are to be managed in the font. The NUKTA is used in encoded text as follows:
4.6 Halanta

The O मसराम गोंडी वायरम is proposed as a vowel silencer. It is used as follows:

\[
\begin{align*}
\text{k} & \text{ ṭ} <\text{क} \text{, घडा, मसराम गोंडी वायरम}> \\
k\text{h} & \text{ ṭ} <\text{क्ष्णा} \text{, घडा, मसराम गोंडी वायरम}>
\end{align*}
\]

The Gondi script as designed by Masaram does not have a native halanta, as the structure of the script does not require it. The embedding of the inherent vowel into the graphical structure of a consonant is a unique and innovative feature of the Gondi script. The horizontal stroke of each consonant letter represents the inherent vowel; removal of this stroke produces a bare consonant: O- ka → 0 k, etc.

In most Indic scripts the inherent vowel is not part of the graphical structure of a consonant letter. As a result, these scripts require a mechanism for indicating the absence of the inherent vowel. In Devanagari this mechanism is a sign called क U+094D विरमा: का + क → क, etc.

However, modern users have adopted the Devanagari halanta (or virama) for marking a bare consonant at the end of a word (see figure 24). The गोंडी वायरम has been included in the repertoire in order to provide this functionality in the proposed encoding.

The proposed encoding for Gondi separates the two functions of the virama character as used in the models for most Indic scripts in Unicode. In Devanagari, for example, the virama functions both as a vowel silencer (halanta) and a control character for forming conjuncts. The default representation of virama in Devanagari is as a visible sign beneath the consonant with which it combines. If a consonant is placed after the virama it causes a conjunct to be formed from the two consonants around it. In order to display a visible virama between adjacent consonants, it is necessary to break conjunct formation. For this purpose, the control character ए U+200C श्लोक जोड़ा is placed after virama. This approach was proposed for Gondi in L2/15-005. However, in the interest of developing a simple encoding model for Gondi, and to eliminate the need for usage of श्लोक जोड़ा or other invisible control characters, the encoding of the vowel silencing feature of the Indic virama has been encapsulated into the character O मसराम गोंडी वायरम. The conjunct forming function is retained in the गोंडी वायरम, described below.

4.7 Virama

The क मसराम गोंडी वायरम is a control character that is used specifically for producing the bare form of a consonant letter. It is represented in the code chart as क in order to indicate that it is a special character. Conceptually, virama produces a half-form by removing the horizontal stroke from the glyph of the letter after which it is placed.

\[
\begin{align*}
\text{k} & \text{ ०} <\text{क} \text{, घडा, मसराम गोंडी वायरम}> \\
k\text{h} & \text{ ०} <\text{क्ष्णा} \text{, घडा, मसराम गोंडी वायरम}>
\end{align*}
\]

It is used for producing conjuncts, similar to the control function of क U+094D विरमा. The Gondi virama, however, does not silence the inherent vowel; the O मसराम गोंडी वायरम is to be used for that purpose.
4.8 Consonant Conjuncts

Consonant clusters are represented as conjuncts and are rendered in a linear sequence using half-forms of all but the final letter in a cluster, which appears in its full form, eg. 00- kka, 0U- kta, 0UB- ktva, etc. Consonants are placed sequentially in the conjunct in the order that they occur in the cluster.

Conjuncts are represented in encoded text by placing the control character [𑯕] MASARAM GONDI VIRAMA after each non-initial consonant in a cluster. The sequence <(C, ¶ virama)>, Cf > produces a half-form of all C and the regular full-form of Cf:

ka 0- <0- ka>
kta 0U- <0- ka, ¶ virama, U- ta>
ktva 0UB- <0- ka, ¶ virama, U- ta, ¶ virama, 8- va>
ktvya 0UB-0- <0- ka, ¶ virama, U- ta, ¶ virama, 8- va, ¶ virama, 0- ya>

There are some exceptions to the rule of conjunct formation. The following character are proposed in order to properly represent all Gondi conjuncts:

𑯖 MASARAM GONDI REPHA
𑱀 MASARAM GONDI RA-KARA
𑱁 MASARAM GONDI LETTER KSSA
𑱂 MASARAM GONDI LETTER JNYA
𑱃 MASARAM GONDI LETTER TRA

The behavior of RA is described in section 4.8.1, and the use of atomic ligatures for three conjuncts is discussed in see section 4.8.2).

4.8.1 Forms of RA in conjuncts

Following the general rule of conjunct formation, the letter 0- RA would occur as the half-form 0 when it is initial or medial in a cluster and in its regular full-form when in cluster-final position. The Gondi RA, however, does not behave entirely as expected in conjuncts and is rendered in several ways. There are three ways of writing 0- RA in conjuncts. It occurs in its half-form 0 when cluster-initial or alternately as ¶ when cluster-initial and as ¶ when cluster-final. These are described below:

- **Half-form** The half-form of RA is used specifically for representing semantic distinctions of RA when the letter occurs at a morphological boundary. Such usage is influenced by Devanagari orthography for the Marathi language, in which र RA may occur as either the र regular repha or the ॥ ‘eyelash’ repha when it is the initial consonant in a cluster; the ‘eyelash’ repha marks plural suffixes (दर्भा daryā ‘valleys’ and दर्भा daryā ‘ocean’) and inflectional suffixes (आचायास ācāryās ‘to the cook’ and आचायास ācāryās ‘to the teacher’). The half-form 0 corresponds to the Devanagari ॥ ‘eyelash’ repha. It is also used when RA occurs in cluster-medial position.

- **Repha** When not used for marking morphological distinctions, cluster-initial RA is rendered as the ¶ repha. The logic of this character is based upon the Devanagari र regular repha. The Gondi repha
attaches after the last letter in a conjunct, above the horizontal line or an extension of it, depending upon the presence of an above-stroke vowel sign, eg. วก rka, วก rkā, etc.

Some modern sources show the repha represented using the form วก, eg. วก kra. This form is simply the regular sign for repha with the left stroke drawn past the horizontal bar and curving to the right. It is a glyphic variant and is to be handled by the font.

• Ra-kāra  In the current orthography, cluster-final ౎ is rendered as ౎ ra-kāra instead of as the full form ౎. The logic of the Gondi ra-kāra is based upon the Devanagari ౎ ra-kāra and ౎ vattu. The Gondi ra-kāra is positioned below the horizontal line of a consonant glyph or beneath an extension of the horizontal line:

Some modern sources show the ra-kāra represented using the form ౎, eg. ౎ kra. This form is a glyphic variant and is to be handled by the font.

• Full-form  The full-form of ౎ is rarely used at present when it is final in a cluster: ౎ kra. The preference is to use ౎ ra-kāra.

The representation of repha and ra-kāra requires an exception to the rule of conjunct formation in Gondi. The general rule states that the sequence <, VIRAMA, C> is rendered using the half-form ౎ of ra and the full-form of C. It also states that <C, VIRAMA, RA> would produce the half-form of C and the full-form ౎ of RA. For this reason another method is required for the encoded representation of repha and ra-kāra, for which the expected encoded sequences according to the general model of Indic scripts would also be <RA, VIRAMA, C> and <C, VIRAMA, RA>, respectively.

There are four possible models for accommodating encoded representations of RA in conjuncts. The first two are based upon the premise that the default behavior of RA in conjuncts is similar to that of all other consonants, ie. it is rendered using the half-form ౎ when cluster-initial and the full-form ౎ when cluster-final. These approaches treat repha and ra-kāra as exceptions. The third model deviates from this premise and establishes the repha and ra-kāra as default representations of RA in conjuncts, and the half-form and full-form as exceptions.

1. Use the Zero-Width Joiner

In L2/15-005, it was suggested that the generic control character วก U+200D ZERO WIDTH JOINER (ZWJ) be used for representing repha and ra-kāra in encoded text. The usage of ZWJ was chosen because the character is used in various Indic scripts for controlling different forms of letters in conjuncts. The same principle was applied to Gondi, such that various forms of RA in conjuncts would be produced as follows:

- half-form วก <ヴ RA, ฺ VIRAMA, C>
- repha วก <ヴ RA, ฺ ZERO WIDTH JOINER, ฺ VIRAMA, C>
- ra-kāra วก <C, ฺ VIRAMA, ฺ ZERO WIDTH JOINER,ヴ RA>
Representation of approach #1 in encoded sequences would be as follows:

\[
\begin{align*}
0\cdot rka & \langle 0\cdot \text{RA}, \heartsuit \text{VIRAMA}, 0\cdot \text{KA} \rangle \\
0\cdot kra & \langle 0\cdot \text{KA}, \heartsuit \text{VIRAMA}, 0\cdot \text{RA} \rangle \\
0\cdot rka & \langle 0\cdot \text{RA}, \text{ZERO WIDTH JOINER}, \heartsuit \text{VIRAMA}, 0\cdot \text{KA} \rangle \\
0\cdot kra & \langle 0\cdot \text{KA}, \heartsuit \text{VIRAMA}, \text{ZERO WIDTH JOINER}, 0\cdot \text{RA} \rangle 
\end{align*}
\]

The Script Ad-hoc Committee recommended against the usage of \(\text{zwj}\) for such cases in Gondi. In L2/15-045, the Subcommittee suggested that an alternative model be developed and that “[f]or repha and ra-kāra, encoding separate characters may be useful, similar to the Malayalam dot reph and the medial ra in Myanmar and Tai Tham.” The major concern is that usage of \(\text{zwj}\) introduces issues for both implementers and end users. For implementers, the use of \(\text{zwj}\) overloads the script-specific rules that must be tailored for each script. For end-users, \(\text{zwj}\) is problematic because it is an invisible control character and is neither readily available on keyboards or easily detectable in text sequence.

2. Redefine the rule for rendering \(<\text{RA}, \text{VIRAMA}>\)

The above model is based on the general rule of conjunct formation in Gondi: \(<C, \text{VIRAMA}>\) produces the half-form of \(C\) and therefore \(<\text{RA}, \text{VIRAMA}>\) should produce the half-form of \(\text{RA}\). The repha and ra-kāra are treated as exceptions to the rule. A third approach for handling the different forms of RA in conjuncts is to redefine the default rendering for RA. With this approach the various forms of RA in conjuncts would be produced in a fashion similar to other Indic scripts:

\[
\begin{align*}
\text{repha} & \circ \langle 0\cdot \text{RA}, \heartsuit \text{VIRAMA}, C \rangle \\
\text{ra-kāra} & \circ \langle C, \heartsuit \text{VIRAMA}, 0\cdot \text{RA} \rangle 
\end{align*}
\]

Representation of approach #2 in encoded sequences would be as follows:

\[
\begin{align*}
0\cdot rka & \langle 0\cdot \text{RA}, \heartsuit \text{VIRAMA}, 0\cdot \text{KA} \rangle \\
0\cdot kra & \langle 0\cdot \text{KA}, \heartsuit \text{VIRAMA}, 0\cdot \text{RA} \rangle 
\end{align*}
\]

This approach, however, does not provide a means for producing the half-form \(\emptyset\) of RA, or even the cluster-final full-form if ever needed. Some mechanism would be required to produce these forms. Inevitably, a control character such as \(\text{zwj}\) would be required:

\[
\begin{align*}
0\cdot rka & \langle 0\cdot \text{RA}, \text{ZERO WIDTH JOINER}, \heartsuit \text{VIRAMA}, 0\cdot \text{KA} \rangle \\
0\cdot kra & \langle 0\cdot \text{KA}, \heartsuit \text{VIRAMA}, \text{ZERO WIDTH JOINER}, 0\cdot \text{RA} \rangle 
\end{align*}
\]

This approach is essentially the converse of what was proposed in L2/15-005. It poses the same issues as it requires usage of \(\text{zwj}\).

3. Encode a ligating form of RA

Another option is to encode a dummy letter whose cluster-initial form is \(\circ \heartsuit \text{repha}\) and whose non-initial form is \(\circ \heartsuit \text{ra-kāra}\). This letter might be called \text{LIGATING RA} and it would be defined for usage
only in conjuncts. Another possible name might be rra, but this name is generally used in Indic scripts in Unicode for the Dravidian ra and as Gondi is a Dravidian language, this name is reserved for potentially encoding a such a character for the Gondi script. The representative glyph for this letter might be 𑮶, which is the letter 蘖 placed within a dashed box that indicates that its function as a control character. The sequence <LIGATING RA, VIRAMA, 𑮶> would produce karna, while <C, VIRAMA, LIGATING RA> would produce ra-kāra. By extension, it does not have a half-form. The letter would be rendered as 𑮶 when used in any other context. Using this character the various forms of 𑮶 would be produced as follows:

\[
\begin{align*}
\text{නා} & \quad \text{rka} \quad <\text{𑮶 RA}, \text{𑮶 VIRAMA}, \text{𑮶 KA}> \\
\text{නා} & \quad \text{kra} \quad <\text{𑮶 KA}, \text{𑮶 VIRAMA}, \text{𑮶 RA}> \\
\text{𑮶} & \quad \text{rka} \quad <\text{𑮶 LIGATING RA}, \text{𑮶 VIRAMA}, \text{𑮶 KA}> \\
\text{𑮶} & \quad \text{kra} \quad <\text{𑮶 KA}, \text{𑮶 VIRAMA}, \text{𑮶 LIGATING RA}>
\end{align*}
\]

Although LIGATING RA is an artificial letter and does not occur in the Gondi script, it facilitates the representation of karna and ra-kāra according to the general rules of conjunct formation, while also providing for the default rendering of ra as half-form and full-form in conjuncts:

\[
\begin{align*}
\text{𑮶} & \quad \text{rkra} \quad <\text{𑮶 LIGATING RA}, \text{𑮶 VIRAMA}, \text{𑮶 KA}, \text{𑮶 VIRAMA}, \text{𑮶 LIGATING RA}> \\
\text{𑮶} & \quad \text{rkra} \quad <\text{𑮶 RA}, \text{𑮶 VIRAMA}, \text{𑮶 KA}, \text{𑮶 VIRAMA}, \text{𑮶 RA}>
\end{align*}
\]

This approach also aligns with the concept that karna and ra-kāra in Devanagari and other scripts are special forms of a ra character. The approach that utilizes LIGATING RA offers users with a clear method for producing normative and special forms of ra in conjuncts using virama, which follows the general rule for the encoded representation of conjuncts in Gondi.

4. **Encode ‘repha’ and ‘ra-kāra’ as separate characters**

In L2/15-045, the Script Ad-hoc Committee recommended that ◀ repha and ◀ ra-kāra be encoded as separate characters. While this approach provides the easiest means for encoding different forms of ra, it also raises several questions. Where should a repha character occur in an encoded sequence? Should it be handled logically and placed at the position where ra would normally occur, with the expectation that the rendering engine would reorder it to the end of the conjunct? Or should it be handled visually, and be placed manually at the end of the conjunct in an encoded sequence? For example, would the encoded representation for the hypothetical conjunct ◀ rkra be ◀ repha, ◀ ra-kāra or ◀ ra-kāra, ◀ repha? Another issue concerns the classification of repha: is it to be considered a letter ( ◀ ) or a sign ( ◀ )? If repha is a letter, then is virama to be placed between it and the following consonant: ◀ repha, ◀ virama, ◀ KA, ◀ ra-kāra?

This approach requires the encoding of the following two characters:

\[
\begin{align*}
\text{𑮶} & \quad \text{REPHA} \\
\text{𑮶} & \quad \text{RA-KARA}
\end{align*}
\]

Furthermore, this approach requires the following considerations:
(a) The *repha* is defined a letter, not a combining sign. In the code chart the ^repha^ is placed within a dashed box as [𑯖] in order to indicate that it requires special rendering. The glyph ^repha^ is used in the output.

(b) The logical model for *repha* is preferred over the visual model, ie. typing the character at the end of a cluster and any accompanying vowel signs. This approach provides the ability to type syllables according to the underlying phonology. The logical model requires that *repha* be placed at the same position in the encoded sequence that it occurs in phonetic expression. The rendering engine will re-order the *repha* to the end of the conjunct after any accompanying vowel signs.

(c) The *ra-kāra* is a regular combining sign. It is to be placed, as expected, in its logical position before any accompanying vowel sign.

(d) The *virama* is not to be used after either the *repha* or *ra-kāra*. Its presence in such contexts would be ignored, for example:

\[
<\text{𑯖} \text{උൾඉඁൺ}, \text{𑮜} \text{඄lève}> \rightarrow \text{𑯖} \text{𑮜} \text{𑯕} \text{𑮜}
\]

The *repha* and *ra-kāra* would be used in encoded sequences as follows:

\[
\text{𑯖} rka <\text{𑯖} \text{REPHA}, \text{𑮜} \text{VIRAMA}, \text{𑮜} \text{KA}> \rightarrow \text{𑯖} \text{𑯕} \text{𑮜}
\]

The usage of *virama* after and before *ra* would produce the expected output:

\[
\text{𑯖} rka <\text{𑮜} \text{RA}, \text{𑯕} \text{VIRAMA}, \text{𑮜} \text{KA}> \rightarrow \text{𑮜} \text{𑯕} \text{𑯖}
\]

The most feasible of the above four is approaches #4. Representing cluster-specific forms of *ra* as separate characters offers a simpler implementation that does not require usage of control characters. For this reason, the suggestion made by the Script Ad-hoc Committee in In L2/15-045 has been adopted.

4.8.2 Conjunct letters

The clusters *kṣa*, *jña*, *tra* are represented not as regular conjuncts, but as distinctive letters. These are proposed for encoding as atomic letters:

\[
\text{𑮪} \text{MASARAM GONDI LETTER KSSA}
\]

\[
\text{𑮫} \text{MASARAM GONDI LETTER JNYA}
\]

\[
\text{𑮫} \text{MASARAM GONDI LETTER TRA}
\]

Following the rules of conjunct formation, the expected representation of these three conjuncts would be:
In the Gondi script, each of these three letters phonetically represent a consonant cluster, but they all have
the structure of an atomic letter. These forms are encoded as consonant letters because in all cases consonant
conjuncts are written as linear sequences of half-forms, not as ligatures. While in most Indic scripts the
written forms for kṣa, jña, tra have encoded representations as a character sequence, such an approach would
not be consistent with this script.

It is evident that these ligatures were developed because distinctive forms exist in Devanagari. These three
conjuncts are often shown at the end of Devanagari orthographies for various languages and are often inter-
preted by users as being distinctive letters that are fundamental elements of the script.

4.8.3 Rendering of conjuncts

As specified at the outset of this section, the sequence <(C, \textsc{virama})*, C f> produces a half-form of all
C and the regular full-form of C f. In order for the rendering of conjuncts to operate properly, the font must contain
a complete set of consonants half-forms (glyphs without the horizontal stroke). The font should substitute
each instance of <(C, \textsc{virama})> with the appropriate half-form for C. If this glyph is not available, then
the output should show the full form of C followed by \textsc{virama}.

4.9 Digits

The script has a full set of digits:

\begin{verbatim}
0  MASARAM GONDI DIGIT ZERO
L  MASARAM GONDI DIGIT ONE
U  MASARAM GONDI DIGIT TWO
N  MASARAM GONDI DIGIT THREE
\theta  MASARAM GONDI DIGIT FOUR
\B  MASARAM GONDI DIGIT FIVE
\L\textsc{lj}  MASARAM GONDI DIGIT SIX
J  MASARAM GONDI DIGIT SEVEN
\textsc{lj}  MASARAM GONDI DIGIT EIGHT
\textsc{l}  MASARAM GONDI DIGIT NINE
\end{verbatim}

Variant forms of digits are shown in figure 18.

4.10 Punctuation

Script-specific punctuation is not attested in Masaram’s Gondi script. The \textit{danda} and double \textit{danda} are
commonly used. These signs of punctuation are not proposed for separate inclusion in the ‘Masaram Gondi’
block, but are to be unified with \textsc{l} U+0964 DEVANAGARI DANDA and \textsc{ll} U+0965 DEVANAGARI DOUBLE DANDA.
Latin marks of punctuation, such as periods, are also used.

4.11 Characters Not Proposed for Encoding

The following are newly-invented characters. Actual usage of these characters, apart from their inclusion in
new charts of the script, is unknown. For this reason, they are not proposed for encoding at present.
Additional vocalic vowel signs  The original script provided for the writing of only one Sanskrit vocalic vowel sign: \( \text{◌} \) VOWEL SIGN VOCALIC R. In order to accommodate anticipated usage of the script for linguistics research, Mukund Gokhale designed signs for the other three vocalic sounds:

\[ \text{◌} \] GONDI VOWEL SIGN VOCALIC RR  
\[ \text{◌} \] GONDI VOWEL SIGN VOCALIC L  
\[ \text{◌} \] GONDI VOWEL SIGN VOCALIC LL

The sounds represented by these four vowel signs are not found natively in Gondi. They occur only in Sanskrit words, and only \( \text{◌} \) VOWEL SIGN VOCALIC R is attested in usage. There is no corresponding independent letter for these signs, and in such contexts they are represented using a consonant-vowel combination composed with the letters RA and LA and the dependent sign for the vocalic letter:

\[ r \] \( \text{◌} \) RA, \( \text{◌} \) VOWEL SIGN VOCALIC R>  
\[ *r \] \( \text{◌} \) RA, \( \text{◌} \) VOWEL SIGN VOCALIC RR>  
\[ l \] \( \text{◌} \) LA, \( \text{◌} \) VOWEL SIGN VOCALIC L>  
\[ *l \] \( \text{◌} \) LA, \( \text{◌} \) VOWEL SIGN VOCALIC LL>

The VOWEL SIGN VOCALIC R is included in the proposed repertoire, but the other three signs are not proposed for inclusion at present because their actual usage is not attested. Space has been reserved for these characters in the block in the event that a justifiable case to encode them is made in the future.

5 Character Data

5.1 Character Properties

The properties for ‘Masaram Gondi’ in the Unicode Character Database format are:

11B90;MASARAM GONDI LETTER A;Lo;0;L;;;;;N;;;;;  
11B91;MASARAM GONDI LETTER AA;Lo;0;L;;;;;N;;;;;  
11B92;MASARAM GONDI LETTER I;Lo;0;L;;;;;N;;;;;  
11B93;MASARAM GONDI LETTER II;Lo;0;L;;;;;N;;;;;  
11B94;MASARAM GONDI LETTER U;Lo;0;L;;;;;N;;;;;  
11B95;MASARAM GONDI LETTER UU;Lo;0;L;;;;;N;;;;;  
11B96;MASARAM GONDI LETTER E;Lo;0;L;;;;;N;;;;;  
11B98;MASARAM GONDI LETTER AE;Lo;0;L;;;;;N;;;;;  
11B99;MASARAM GONDI LETTER O;Lo;0;L;;;;;N;;;;;  
11B9B;MASARAM GONDI LETTER OU;Lo;0;L;;;;;N;;;;;  
11B9C;MASARAM GONDI LETTER KA;Lo;0;L;;;;;N;;;;;  
11B9D;MASARAM GONDI LETTER KHA;Lo;0;L;;;;;N;;;;;  
11B9E;MASARAM GONDI LETTER GA;Lo;0;L;;;;;N;;;;;  
11B9F;MASARAM GONDI LETTER GHA;Lo;0;L;;;;;N;;;;;  
11BA0;MASARAM GONDI LETTER NGA;Lo;0;L;;;;;N;;;;;  
11BA1;MASARAM GONDI LETTER CA;Lo;0;L;;;;;N;;;;;  
11BA2;MASARAM GONDI LETTER CHA;Lo;0;L;;;;;N;;;;;  
11BA3;MASARAM GONDI LETTER JA;Lo;0;L;;;;;N;;;;;  
11BA4;MASARAM GONDI LETTER JHA;Lo;0;L;;;;;N;;;;;  
11BA5;MASARAM GONDI LETTER NYA;Lo;0;L;;;;;N;;;;;  
11BA6;MASARAM GONDI LETTER TTA;Lo;0;L;;;;;N;;;;;  
11BA7;MASARAM GONDI LETTER TTHA;Lo;0;L;;;;;N;;;;;
Proposal to Encode the Masaram Gondi Script in Unicode

Anshuman Pandey

5.2 Linebreaking

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

11B90..11BC0; AL  # MASARAM GONDI LETTER A .. MASARAM GONDI LETTER TRA
11BC1..11BD4; CM  # MASARAM GONDI SIGN AA .. MASARAM GONDI SIGN HALANTA

11BA8;MASARAM GONDI LETTER DDA;Lo;0;L;;;;;N;;;;;
11BA9;MASARAM GONDI LETTER DDHA;Lo;0;L;;;;;N;;;;;
11BAA;MASARAM GONDI LETTER NNA;Lo;0;L;;;;;N;;;;;
11BAB;MASARAM GONDI LETTER TA;Lo;0;L;;;;;N;;;;;
11BAC;MASARAM GONDI LETTER THA;Lo;0;L;;;;;N;;;;;
11BAD;MASARAM GONDI LETTER DA;Lo;0;L;;;;;N;;;;;
11BAE;MASARAM GONDI LETTER DHA;Lo;0;L;;;;;N;;;;;
11BAF;MASARAM GONDI LETTER NA;Lo;0;L;;;;;N;;;;;
11BB0;MASARAM GONDI LETTER PA;Lo;0;L;;;;;N;;;;;
11BB1;MASARAM GONDI LETTER PHA;Lo;0;L;;;;;N;;;;;
11BB2;MASARAM GONDI LETTER BA;Lo;0;L;;;;;N;;;;;
11BB3;MASARAM GONDI LETTER BHA;Lo;0;L;;;;;N;;;;;
11BB4;MASARAM GONDI LETTER MA;Lo;0;L;;;;;N;;;;;
11BB5;MASARAM GONDI LETTER YA;Lo;0;L;;;;;N;;;;;
11BB6;MASARAM GONDI LETTER RA;Lo;0;L;;;;;N;;;;;
11BB7;MASARAM GONDI LETTER LA;Lo;0;L;;;;;N;;;;;
11BB8;MASARAM GONDI LETTER VA;Lo;0;L;;;;;N;;;;;
11BB9;MASARAM GONDI LETTER SHA;Lo;0;L;;;;;N;;;;;
11BBA;MASARAM GONDI LETTER SSA;Lo;0;L;;;;;N;;;;;
11BBB;MASARAM GONDI LETTER SA;Lo;0;L;;;;;N;;;;;
11BBC;MASARAM GONDI LETTER HA;Lo;0;L;;;;;N;;;;;
11BBD;MASARAM GONDI LETTER LLA;Lo;0;L;;;;;N;;;;;
11BBE;MASARAM GONDI LETTER KSSA;Lo;0;L;;;;;N;;;;;
11BBF;MASARAM GONDI LETTER JNYA;Lo;0;L;;;;;N;;;;;
11BC0;MASARAM GONDI LETTER TRA;Lo;0;L;;;;;N;;;;;
11BC1;MASARAM GONDI VOWEL SIGN AA;Mn;0;NSM;;;;;N;;;;;
11BC2;MASARAM GONDI VOWEL SIGN I;Mn;0;NSM;;;;;N;;;;;
11BC3;MASARAM GONDI VOWEL SIGN II;Mn;0;NSM;;;;;N;;;;;
11BC4;MASARAM GONDI VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
11BC5;MASARAM GONDI VOWEL SIGN UU;Mn;0;NSM;;;;;N;;;;;
11BC6;MASARAM GONDI VOWEL SIGN VOCALIC R;Mn;0;NSM;;;;;N;;;;;
11BCA;MASARAM GONDI VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
11BCC;MASARAM GONDI VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
11BCD;MASARAM GONDI VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;;
11BCF;MASARAM GONDI VOWEL SIGN AU;Mn;0;NSM;;;;;N;;;;;
11BD0;MASARAM GONDI SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11BD1;MASARAM GONDI SIGN VISARGA;Mn;0;NSM;;;;;N;;;;;
11BD2;MASARAM GONDI SIGN NUKTA;Mn;0;NSM;;;;;N;;;;;
11BD3;MASARAM GONDI SIGN CANDRA;Mn;0;NSM;;;;;N;;;;;
11BD4;MASARAM GONDI SIGN HALANTA;Mn;0;NSM;;;;;N;;;;;
11BD5;MASARAM GONDI SIGN VIRAMA;Mn;0;NSM;;;;;N;;;;;
11BD6;MASARAM GONDI SIGN REPHA;Lo;0;L;;;;;N;;;;;
11BD7;MASARAM GONDI SIGN RA-KARA;Mn;0;NSM;;;;;N;;;;;
11BE0;MASARAM GONDI DIGIT ZERO;Nd;0;L;;;;;N;;;;;
11BE1;MASARAM GONDI DIGIT ONE;Nd;0;L;;;;;N;;;;;
11BE2;MASARAM GONDI DIGIT TWO;Nd;0;L;;;;;N;;;;;
11BE3;MASARAM GONDI DIGIT THREE;Nd;0;L;;;;;N;;;;;
11BE4;MASARAM GONDI DIGIT FOUR;Nd;0;L;;;;;N;;;;;
11BE5;MASARAM GONDI DIGIT FIVE;Nd;0;L;;;;;N;;;;;
11BE6;MASARAM GONDI DIGIT SIX;Nd;0;L;;;;;N;;;;;
11BE7;MASARAM GONDI DIGIT SEVEN;Nd;0;L;;;;;N;;;;;
11BE8;MASARAM GONDI DIGIT EIGHT;Nd;0;L;;;;;N;;;;;
11BE9;MASARAM GONDI DIGIT NINE;Nd;0;L;;;;;N;;;;;
5.3 Syllabic Categories

Syllabic categories given in the format of IndicSyllabicCategory.txt:

```plaintext
# Indic_Syllabic_Category=Bindu
11BD0 ; Bindu # Mn MASARAM GONDI SIGN ANUSVARA

# Indic_Syllabic_Category=Visarga
11BD1 ; Visarga # Mc MASARAM GONDI SIGN VISARGA

# Indic_Syllabic_Category=Nukta
11BD2 ; Nukta # Mn MASARAM GONDI SIGN NUKTA

# Indic_Syllabic_Category=Virama
11BD5 ; Virama # Mn MASARAM GONDI VIRAMA

# Indic_Syllabic_Category=Pure_Killer
11BD4 ; Pure_Killer # Mn MASARAM GONDI SIGN HALANTA

# Indic_Syllabic_Category=Vowel_Independent
11B90..11B9B ; Vowel_Independent # Lo [10] MASARAM GONDI LETTER A .. AU

# Indic_Syllabic_Category=Vowel_Dependent
11BC1..11BC6 ; Vowel_Dependent # Mn [6] MASARAM GONDI VOWEL SIGN AA .. VOCALIC R
11BCA ; Vowel_Dependent # Mn MASARAM GONDI VOWEL SIGN E
11BCC..11BCD ; Vowel_Dependent # Mn [2] MASARAM GONDI VOWEL SIGN AI .. O
11BCF ; Vowel_Dependent # Mn MASARAM GONDI VOWEL SIGN AU
11BD3 ; Vowel_Dependent # Mn MASARAM GONDI VOWEL SIGN CANDRA

# Indic_Syllabic_Category=Consonant
11B9C..11BC0 ; Consonant # Lo [40] MASARAM GONDI LETTER KA .. TRA

# Indic_Syllabic_Category=Consonant_Succeeding_Repha
11BD6 ; Consonant_Succeeding_Repha # Lo MASARAM GONDI REPHA

# Indic_Syllabic_Category=Consonant_Subjoined
11BD7 ; Consonant_Subjoined # Mn MASARAM GONDI RA-KARA
```

5.4 Positional Categories

Positional data for Gondi combining signs in the format of IndicPositionalCategory.txt:

```plaintext
# Indic_Positional_Category=Top
11BC1..11BC5 ; Top # Mn [5] MASARAM GONDI VOWEL SIGN AA .. UU
11BCA ; Top # Mn MASARAM GONDI VOWEL SIGN E
11BCC..11BCD ; Top # Mn [2] MASARAM GONDI VOWEL SIGN AI .. O
11BCF ; Top # Mn MASARAM GONDI VOWEL SIGN AU
11BD0 ; Top # Mn MASARAM GONDI SIGN ANUSVARA
11BD1 ; Top # Mn MASARAM GONDI SIGN VISARGA
11BD3 ; Top # Mn MASARAM GONDI SIGN CANDRA

# Indic_Positional_Category=Bottom
11BC6 ; Bottom # Mn MASARAM GONDI VOWEL SIGN VOCALIC R
```
5.5 ‘Confusable’ Characters

Gondi characters that bear resemblances to those of other scripts are listed below:

- \texttt{11BC1} MASARAM GONDI VOWEL SIGN AA ; 0304 COMBINING MACRON
- \texttt{11BB1} MASARAM GONDI LETTER PHA ; 1109D KAITHI LETTER NNA
- \texttt{11BBA} MASARAM GONDI LETTER SSA ; 0398 GREEK CAPITAL LETTER THETA
- \texttt{11BE2} MASARAM GONDI DIGIT TWO ; 0055 LATIN CAPITAL LETTER U

6 References


गुरूजी, मनीराम दुगा [Gurūjī, Manīrāma Durgā]. गौंडी लाम्क पुंदान [Goṃḍī lamka puṃdāna].


Masaram, Bhava Singh. 1951. “गौंडी लिपि” [Goṃḍī lipi]. Central Institute of Indian Languages, Multimedia library, photograph no. 64.


रामानन्द [Rāmānanda]. गौंडी अक्षर ज्ञान [Goṃḍī Akṣara Jñāna].

7 Acknowledgments

I am deeply indebted to Dr. Mukund Gokhale (Script Research Institute, Pune) for providing me with numerous materials in Masaram’s Gondi script and for patiently answering my numerous questions about the script over the past few years. Dr. Gokhale’s efforts in developing typefaces for this Gondi script have helped to address issues regarding the interaction of multiple glyphs with a base letter. I also appreciate him for contacting Dr. Motiravan Kangle (Akhil-Gondwana Gondi Sahitya Parishad, Nagpur) with my questions. I thank Dr. Kangle for indulging my inquiries and for providing examples of current orthography of ra in conjuncts, and the usage of halanta and candra. The Gondi chart shown in figure ?? was provided by B. A. Sharada and Suman Kumari of the Central Institute of Indian Languages (Mysore). Mark Penny provided the chart shown in figure 6. I am also thankful to Roozbeh Pournader (Google) and Andrew Glass (Microsoft) for sharing their insights regarding effective models for representing the Gondi RA in conjuncts.

This project was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative, and a grant from the United States National Endowment for the Humanities (PR-50205-15), which funds 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 Google or the National Endowment for the Humanities.

Copyright © 2015 Anshuman Pandey. All rights reserved.
### Vowels

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11B0</td>
<td>MASARAM GONDI LETTER A</td>
</tr>
<tr>
<td>11B1</td>
<td>MASARAM GONDI LETTER AA</td>
</tr>
<tr>
<td>11B2</td>
<td>MASARAM GONDI LETTER I</td>
</tr>
<tr>
<td>11B3</td>
<td>MASARAM GONDI LETTER II</td>
</tr>
<tr>
<td>11B4</td>
<td>MASARAM GONDI LETTER U</td>
</tr>
<tr>
<td>11B5</td>
<td>MASARAM GONDI LETTER UU</td>
</tr>
<tr>
<td>11B6</td>
<td>MASARAM GONDI LETTER E</td>
</tr>
<tr>
<td>11B7</td>
<td>MASARM GONDI LETTER AI</td>
</tr>
<tr>
<td>11B8</td>
<td>MASARM GONDI LETTER O</td>
</tr>
<tr>
<td>11B9</td>
<td>MASARM GONDI LETTER AU</td>
</tr>
<tr>
<td>11BE8</td>
<td>MASARM GONDI DIGIT NINE</td>
</tr>
<tr>
<td>11BE9</td>
<td>MASARM GONDI DIGIT NINE</td>
</tr>
</tbody>
</table>

### Consonants

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BC</td>
<td>MASARM GONDI LETTER KA</td>
</tr>
<tr>
<td>11BD</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BE</td>
<td>MASARM GONDI LETTER GA</td>
</tr>
<tr>
<td>11BF</td>
<td>MASARM GONDI LETTER GHA</td>
</tr>
<tr>
<td>11BA</td>
<td>MASARM GONDI LETTER GA</td>
</tr>
<tr>
<td>11BC</td>
<td>MASARM GONDI LETTER JA</td>
</tr>
<tr>
<td>11BD</td>
<td>MASARM GONDI LETTER JHA</td>
</tr>
<tr>
<td>11BE</td>
<td>MASARM GONDI LETTER JHA</td>
</tr>
<tr>
<td>11BF</td>
<td>MASARM GONDI LETTER JHA</td>
</tr>
<tr>
<td>11BA</td>
<td>MASARM GONDI LETTER JHA</td>
</tr>
<tr>
<td>11BC</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BD</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BE</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BF</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BA</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BC</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BD</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BE</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BF</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BA</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BC</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BD</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BE</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BF</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
<tr>
<td>11BA</td>
<td>MASARM GONDI LETTER KHA</td>
</tr>
</tbody>
</table>

### Consonant clusters

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BE</td>
<td>MASARM GONDI LETTER KSSA</td>
</tr>
<tr>
<td>11BF</td>
<td>MASARM GONDI LETTER JNYA</td>
</tr>
<tr>
<td>11BC</td>
<td>MASARM GONDI LETTER TRA</td>
</tr>
</tbody>
</table>

### Vowels

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BE7</td>
<td>MASARM GONDI VOWEL SIGN AI</td>
</tr>
<tr>
<td>11BE8</td>
<td>MASARM GONDI VOWEL SIGN O</td>
</tr>
<tr>
<td>11BE9</td>
<td>MASARM GONDI VOWEL SIGN AU</td>
</tr>
</tbody>
</table>

### Various signs

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BD0</td>
<td>MASARM GONDI SIGN ANUSVARA</td>
</tr>
<tr>
<td>11BD1</td>
<td>MASARM GONDI SIGN VISARGA</td>
</tr>
<tr>
<td>11BD2</td>
<td>MASARM GONDI SIGN NUKTA</td>
</tr>
<tr>
<td>11BD3</td>
<td>MASARM GONDI SIGN CANDRA</td>
</tr>
</tbody>
</table>

### Virama

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BD5</td>
<td>MASARM GONDI VIRAMA</td>
</tr>
</tbody>
</table>

### Cluster-specific consonant forms

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BD6</td>
<td>MASARM GONDI REPHA</td>
</tr>
</tbody>
</table>

### Digits

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BE0</td>
<td>MASARM GONDI DIGIT ZERO</td>
</tr>
<tr>
<td>11BE1</td>
<td>MASARM GONDI DIGIT ONE</td>
</tr>
<tr>
<td>11BE2</td>
<td>MASARM GONDI DIGIT TWO</td>
</tr>
<tr>
<td>11BE3</td>
<td>MASARM GONDI DIGIT THREE</td>
</tr>
<tr>
<td>11BE4</td>
<td>MASARM GONDI DIGIT FOUR</td>
</tr>
<tr>
<td>11BE5</td>
<td>MASARM GONDI DIGIT FIVE</td>
</tr>
<tr>
<td>11BE6</td>
<td>MASARM GONDI DIGIT SIX</td>
</tr>
<tr>
<td>11BE7</td>
<td>MASARM GONDI DIGIT SEVEN</td>
</tr>
<tr>
<td>11BE8</td>
<td>MASARM GONDI DIGIT EIGHT</td>
</tr>
<tr>
<td>11BE9</td>
<td>MASARM GONDI DIGIT NINE</td>
</tr>
</tbody>
</table>

### Dependent vowel signs

<table>
<thead>
<tr>
<th>Unicode</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>11BC1</td>
<td>MASARM GONDI VOWEL SIGN AA</td>
</tr>
<tr>
<td>11BC2</td>
<td>MASARM GONDI VOWEL SIGN I</td>
</tr>
<tr>
<td>11BC3</td>
<td>MASARM GONDI VOWEL SIGN II</td>
</tr>
<tr>
<td>11BC4</td>
<td>MASARM GONDI VOWEL SIGN U</td>
</tr>
<tr>
<td>11BC5</td>
<td>MASARM GONDI VOWEL SIGN UU</td>
</tr>
<tr>
<td>11BC6</td>
<td>MASARM GONDI VOWEL SIGN VOCALIC R</td>
</tr>
<tr>
<td>11BC7</td>
<td>&lt;reserved&gt;</td>
</tr>
<tr>
<td>11BC8</td>
<td>&lt;reserved&gt;</td>
</tr>
<tr>
<td>11BC9</td>
<td>&lt;reserved&gt;</td>
</tr>
<tr>
<td>11BCA</td>
<td>&lt;reserved&gt;</td>
</tr>
<tr>
<td>11BCB</td>
<td>&lt;reserved&gt;</td>
</tr>
</tbody>
</table>
Figure 1: A document illustrating the characters and basic principles of Mangal Singh Masaram’s Gondi script. It was created by Mangal Singh’s son Bhava Singh in 1951. Image courtesy of the Central Institute of Indian Languages, Mysore.
Figure 2: A schematic designed by Mukund Gokhale showing the position of all combining signs on a base letter. The diagram shows a sign labelled “RR-vocalic sign”, which is a length variant of ऐ VOWEL SIGN VOCALIC R introduced by Gokhale (see section 4.11).
Proposal to Encode the Masaram Gondi Script in Unicode

Figure 3: Cover of Gondwana Darshan (March-April 1990, vol. 5).
Figure 4: Page from *Gomḍī Akṣara Jñāna* showing consonants and vowel signs, and consonant-vowel combinations (from Rāmānanda: 8).
Figure 5: Page from *Goṃḍī Lamk Pundan* showing consonants and vowel signs, and consonant-vowel combinations (from Guruji: 11).
## Consonants व्यङ्जन

<table>
<thead>
<tr>
<th>Cons. only</th>
<th>− [voice]</th>
<th>+ [voice]</th>
<th>Nasals</th>
<th>Cons. with word</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aspiration</strong></td>
<td>− [asp]</td>
<td>+ [asp]</td>
<td>− [asp]</td>
<td>+ [asp]</td>
</tr>
<tr>
<td>Velar</td>
<td>０</td>
<td>８</td>
<td>१</td>
<td>४</td>
</tr>
<tr>
<td>Palatal</td>
<td>४</td>
<td>५</td>
<td>६</td>
<td>७</td>
</tr>
<tr>
<td>Retroflex</td>
<td>७</td>
<td>८</td>
<td>९</td>
<td>०</td>
</tr>
<tr>
<td>Dental</td>
<td>२</td>
<td>३</td>
<td>४</td>
<td>५</td>
</tr>
<tr>
<td>Bilabial</td>
<td>६</td>
<td>७</td>
<td>८</td>
<td>९</td>
</tr>
<tr>
<td>Liquids and Semi-Vowels</td>
<td>८</td>
<td>९</td>
<td>०</td>
<td>१</td>
</tr>
<tr>
<td>Fricatives</td>
<td>२</td>
<td>३</td>
<td>४</td>
<td>५</td>
</tr>
<tr>
<td>Affricates</td>
<td>७</td>
<td>८</td>
<td>९</td>
<td>०</td>
</tr>
</tbody>
</table>

## Vowels स्वर

![Vowel Chart]

Figure 6: A handwritten chart of the Gondi script. Source: Ramesh Gedam and Mark Penny (2001).
Figure 7: A handwritten chart of the Gondi script (Maṇḍāle 2008: 8).
Figure 8: Page from Goṃḍī Aksara Jñāna showing vowel letters (from Rāmānanda: 1–4).
Figure 9: Page from *Gomdi Aksara Jhana* showing the letters TA .. VA (from Rāmānanda: 5–7).
Figure 10: Page from Gōndī Lamk Pandan showing vowel letters (from Guruji: 1–4). The first page (top right corner) shows the persapen symbol.
Figure 11: Page from *Gomdi Lamk Pandan* showing vowel letters (from Guruji: 5–8).
Figure 12: Page from *Gomdi Lamk Pandan* showing vowel letters (from Guruji: 9–10).
Figure 13: Page from *Goṃḍī Aksara Jñāna* showing consonant-vowel combinations for का .. दा (from Rāmānanda: 9, 10).
Figure 14: Page from Gomḍi Akṣara Jñāna showing consonant-vowel combinations for DHA .. TRA (from Rāmānanda: 11, 12).
Figure 15: Page from Goṃḍī Akṣara Jāna explaining conjunct formation (from Rāmānanda: 13). Bottom half describes the usage of Latin marks of punctuation.
Figure 16: Page from *Goṇḍī Akṣara Jiṇāna* showing Gondi digits (from Rāmānanda: 14).

Figure 17: Page from *Goṇḍī Lamk Pundan* showing Gondi digits (from Guruji: 14).
### Figure 18: Comparison of regional variants of Gondi digits (from Vahia and Halkare 2013: 33).

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Style in Maharashtra</th>
<th>Style in Madhya Pradesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Undi</td>
<td>U</td>
<td>L</td>
</tr>
<tr>
<td>2</td>
<td>Rand</td>
<td>R</td>
<td>U</td>
</tr>
<tr>
<td>3</td>
<td>Munda</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>Nalung</td>
<td>N</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>Sayyung</td>
<td>S</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>Sarung</td>
<td>S</td>
<td>L</td>
</tr>
<tr>
<td>7</td>
<td>Yerung</td>
<td>J</td>
<td>J</td>
</tr>
<tr>
<td>8</td>
<td>Arung</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>Narung</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>Pad</td>
<td>P</td>
<td>L0</td>
</tr>
</tbody>
</table>
Figure 19: A document showing usage of ०- l.la.
Proposal to Encode the Masaram Gondi Script in Unicode

Anshuman Pandey

Figure 20: A letter written in Gondi.

Figure 21: An invitation card written in the Devanagari and Gondi scripts. The persapen symbol appears in the top center of the card.
Figure 22: A Christian prayer typeset using digitized Gondi and Devanagari fonts. The Gondi font used in this specimen was designed by Mukund Gokhale.
Figure 23: The lyrics to “Vande Mataram” transliterated into Gondi.

Text is complete “Vande Mataram”
This ‘Gondi’ Script font is designed by : Prof Mukund Gokhale
Script Research Institute
m wgokhale_scriptresearch@yahoo.co.in
gokhale.mukund1@gmail.com

Dedicated to the inventor of the script ‘Munshi Mangal Singh Masaram’
Figure 24: Usage of the sign HALANTA with regular forms of consonants for representation of bare consonants. Contents of letter described in section 4.6.
Proposal to Encode the Masaram Gondi Script in Unicode

Anshuman Pandey

Figure 25: Usage of the sign CANDRA. Contents of letter described in section 4.4.
Figure 26: A calendar showing names for days of the week in Gondi and dates in Gondi digits.
Figure 27: Screen-shot of adivasiSwara.org showing content in Gondi script. The text in the banner is valid Gondi content. The text in the main frame of the site, however, is invalid: the script representing meaningless sequences of Gondi letters as the underlying text is Latin-script content in English. The Gondi text is represented using a server-side font based upon the Latin encoding that contains Gondi glyphs. The site appears to be a work-in-progress and is awaiting proper support for the Gondi script in Unicode in order to deliver proper content.
ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

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
for latest Roadmaps.

See also
http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html

A. Administrative

<table>
<thead>
<tr>
<th>1. Title:</th>
<th>Proposal to Encode the Masaram Gondi Script in Unicode</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Requester's name:</td>
<td>Script Encoding Initiative (SEI) / Anshuman Pandey (<a href="mailto:anshuman.pandey@berkeley.edu">anshuman.pandey@berkeley.edu</a>)</td>
</tr>
<tr>
<td>3. Requester type (Member body/Liaison/Individual contribution):</td>
<td>Liaison contribution</td>
</tr>
<tr>
<td>4. Submission date:</td>
<td>2015-04-15</td>
</tr>
<tr>
<td>5. Requester's reference (if applicable):</td>
<td></td>
</tr>
<tr>
<td>6. Choose one of the following:</td>
<td></td>
</tr>
<tr>
<td>(or) More information will be provided later:</td>
<td>Yes</td>
</tr>
</tbody>
</table>

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
      Proposed name of script: Masaram Gondi
      Name of the existing block: |
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block: |

2. Number of characters in proposal: 75

3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary X B.1-Specialized (small 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 (anshuman.pandey@berkeley.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.

---

### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?
   - Yes
   - If YES explain: Replaces L2/10-207, L2/12-235, L2/15-005

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: Mukund Gokhale (Script Research Institute, Pune)  
   - Motiravan Kangle (Akhil Gondwana Gondi Sahitya Parishad, Nagpur)

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)
   - Common
   - Reference: Used for writing the Gondi language in India

5. Are the proposed characters in current use by the user community?
   - Yes
   - 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?
   - 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?
   - 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?
   - 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?
    - 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?
    - Reference: Combining signs
    - Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    - 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)
    - Reference: 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?
    - Reference: