

# Towards an Encoding for the Maithili Script in ISO/IEC 10646

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

30 September 2009

## 1 Introduction

Maithili is the traditional writing system for the Maithili language, which is spoken predominantly in the state of Bihar in India and in the Narayani and Janakpur zones of Nepal. The script is also known as Mithilākṣara and Tirahutā, names which refer to Mithila, or Tirhut, as these regions of India and Nepal are historically known. Maithili is a Brahmi-based script derived from Gauḍī, or ‘Proto-Bengali’, which evolved from the Kuṭila branch of Brahmi by the 10th century CE.<sup>1</sup> It is related to Bengali, Newari, and Oriya, which are also descended from Gauḍī, and became differentiated from these scripts by the 14th century. It remained the dominant writing system in Mithila from that period until the 20th century.

Maithili is associated with a scholarly tradition that spans six centuries, as attested by manuscripts and inscriptions records. Uniquely Maithili manuscripts consist of the vernacular songs of the poet Vidyāpati Ṭhākura (14th century) and Sanskrit treatises on Nyāya philosophy; moreover, Maithili is the traditional script for Sanskrit in the Mithila region. Inscriptions in Maithili are located in Bihar and Nepal. The Maithil Brahmin community has used the script for maintaining *pāñjī* or genealogical records since the 14th century. Maithili metal fonts were developed in Kolkata (Calcutta) in the 1920s and books were first printed in the script at that time. Devanagari replaced Maithili in the 20th century; however, Maithili is still used for producing genealogical records, religious texts, and letters. The script is used in signage in north Bihar and is permitted as an optional script for writing the Bihar civil services examination.

The Government of India recognized Maithili as a scheduled language in 2004, a status that ensures official support for the development of the language. The official status of Maithili has further revitalized interest in the traditional script for the language. In 2005, in a presentation to the Unicode Technical Committee, Dr. Om Vikas of the Department of Information Technology, Government of India noted the historical importance of Maithili (Tirhuta) and expressed the Department’s interest in establishing a Unicode standard for the script.<sup>2</sup> This development builds upon the activities of Maithili speakers in India and Nepal, who, in the past decade, have produced digitized Maithili fonts<sup>3</sup> and published dictionaries, books of poetry, novels in the Maithili script and language in print and digital media. Electronic journals on Maithili literature and culture are published in the Maithili script,<sup>4</sup> and there is an effort to develop resources such as Wikipedia in the Maithili language and script.<sup>5</sup> An encoding for Maithili in the UCS will support current activity and research on Maithili language and script, and will facilitate the development of new resources and technologies for Maithili.

<sup>1</sup> Salomon 1998: 41. <sup>2</sup> Vikas 2005: 31. <sup>3</sup> eCube Solutions & Research 2003. <sup>4</sup> Thakur 2009. <sup>5</sup> Wikimedia 2008.

## 2 The Issue of Representing Maithili in Unicode

Maithili is a writing system that is associated with a distinct linguistic community and a historically important Sanskritic scholarly tradition. It is a script that despite being nearly obsolete, has been preserved and has received renewed attention through revival efforts. An encoding for Maithili, therefore, may be viewed by its users as a means for preserving not only the script and the materials written in it, but also as the preservation of a cultural identity associated with the Maithili language and script. While such a perspective is certainly positive, the preservation of linguistic identity is not a directive of the UCS. The representation of scripts in the UCS is driven not by socio-cultural factors, but by the typological features and technical requirements of a writing system and by the design principles of character encoding.

The primary issue facing the representation of Maithili in the UCS is its close affinity to the Bengali script. Maithili has several characters that are similar or identical to those already encoded for Bengali. However, the two scripts also differ in terms of character repertoires, glyph shapes, orthographic features, and rendering behaviors. The correspondences between the characters of Maithili and Bengali requires that a method for representing Maithili be established, which meets the requirements of the user community, while adhering to the design principles that govern the structure of Unicode.

With regard to the encoding of Maithili, the most salient of the Unicode design principles is ‘plain text’. Plain text is the representation of character content that is independent of formatting. In other words, the display of Maithili in plain text means that the characters and orthographic features of the script are managed at the character level, not through font changes or alterations to text style. The user requirement to represent Maithili in plain text adheres to the structural principles of Unicode; however, the script must also conform to another design principle, namely, ‘unification’. Unification is the principle of preventing the duplicate encoding of characters within scripts across languages. According to this principle, while Maithili is distinct from Bengali in several respects, it is considered a candidate for unification with Bengali because of the number of correspondences between the two scripts.

Taking into consideration the principles of ‘plain text’ and ‘unification’, there are three models for representing Maithili in the UCS:

1. Encode Maithili as an independent script
2. Encode Maithili as a subset of Bengali
3. Completely unify Maithili with Bengali

## 3 Assessment of the Models for Representing Maithili

Based upon the number of Maithili characters that correspondence with those of Bengali, it might be argued that Maithili is a variant of Bengali, and is therefore a candidate for unification with Bengali. This position suggests that Maithili should be accommodated in the UCS as a subset of Bengali either by (a) managing it at the presentation level through fonts, or (b) encoding Maithili-specific characters as extensions to Bengali. Both of These approaches are problematic:

- *Complete unification of Maithili with Bengali* Unifying Maithili and Bengali is synonymous with defining Maithili as a variant of Bengali and establishing no standard for the script. Complete unification mandates that the representation and display of Maithili be bound to font control, ie. Maithili glyphs will be designed and assigned to the code points of corresponding Bengali characters in Bengali fonts. Thus, when users create or view Maithili content, the text will be represented in Bengali script; the user must chose a Bengali font designed to accommodate Maithili glyph in order to display Maithili

properly. Complete unification of Maithili with Bengali does not provide a means for representing the script in plain text.

- *Encoding Maithili as subset of Bengali* The Bengali script is used for languages other than Bengali, such as Assamese and Manipuri. With the exception of two characters, Assamese orthography is identical to that for Bengali. These two characters were added to the Bengali block in order to accommodate Assamese orthography in the UCS. Although these characters could be treated as glyphic variants of existing Bengali characters, a decision was made to encode them separately. The Assamese case offers a precedent for encoding scripts with large correspondences with an already-encoded script. The correspondences between Maithili and Bengali are much less than between Assamese and Bengali. If Assamese characters were added as atomic characters, then similarly, Maithili must at minimum be represented in the UCS as an extension to Bengali. This will require that at least 40 Maithili-specific characters be added for Bengali. Managing Maithili as an extension to Bengali will entail not only the addition of a large number of characters to an existing script, but require the development of new rendering rules for an already complex script.

Aside from the technical difficulties related to unification of complex scripts, an equally troublesome issue with unification resides in the definition of the unification principle itself. The principle prevents against the encoding of duplicate characters *within scripts across languages*. Although Maithili and Bengali possess corresponding characters, unification of the two suggests that they are one and the same script. Unifying characters used in the orthographies of languages based upon the Latin script is justifiable; however, despite the correspondences between Latin and Cyrillic, characters in the two scripts that have identical semantic and glyphical value are encoded independently. The rationale is that unification operates *within* scripts, not *across* them. The Maithili language is not associated with the Bengali script, and the Maithili script is not considered a variant of the Bengali script. Therefore, to encode Maithili by unifying it with Bengali would be to unify characters across scripts, which contradicts Unicode principles.

Presented below are technical considerations regarding unification, which is followed by a discussion of other factors, such as user perception, that affect the method of representing Maithili as a script unified with Bengali.

## 4 Technical Considerations

### 4.1 Character Repertoire

An analysis of the Maithili character repertoire indicates that it has several characters that are identical to Bengali, but it also has several character that are distinct from analogues in Bengali. Of the 85 characters proposed for Maithili (see tables 6 and 7), 39 characters are contrastive in some aspect from their Bengali analogues, while 6 characters are not found in Bengali (see tables 1, 2, and 3):

- **Vowel letters:** A, AA, I, II, U, UU, VOCALIC R, VOCALIC RR, VOCALIC L, VOCALIC LL.
- **Vowel signs:** II, U, UU, VOCALIC R, VOCALIC RR, VOCALIC L, VOCALIC LL, AI.
- **Consonants:** NGA, CHA, JHA, TTA, DDHA, RHA, NAA, PHA, BA, BHA, RA, LA, SHA, HA.
- **Digits:** THREE, FOUR, FIVE, SEVEN, EIGHT, NINE.
- **Various signs:** ANUSVARA.
- **No correspondence:** VOWEL SIGN SHORT E, LLA, VA, ANJI, OM, GVANG, ABBREVIATION SIGN.

Nearly all Maithili vowel letters contrast with Bengali characters. Only the independent forms of E, AI, O, and AU are identical. The majority of vowel signs are contrastive between the scripts. Maithili has signs

for SHORT E and SHORT O, which are not found in Bengali, although the latter corresponds glyphically to ৌ U+09D7 BENGALI AU LENGTH MARK, despite the name.

Although the correspondences between consonant letters are greater, Maithili has characters with no Bengali analogues and there are important distinctions between corresponding characters. Maithili has distinct characters for 𑒧 BA and 𑒧 VA, while Bengali does not have VA.

## 4.2 Size of Character Repertoire

Unifying Maithili with Bengali is a more complex task than unifying Assamese with Bengali. The latter required the encoding of only two characters: 𑒧 U+09F0 BENGALI LETTER RA WITH MIDDLE DIAGONAL and 𑒧 U+09F1 BENGALI LETTER RA WITH LOWER DIAGONAL. Such accommodations are acceptable when the difference between two scripts is a few characters. Representing Maithili as a subset of Bengali would require encoding at least 40 characters as extensions to Bengali.

## 4.3 Homoglyphic Characters

A unified encoding for Maithili with Bengali would prevent the encoding of duplicate characters, but it would not prevent the encoding of homoglyphic characters within a single script block. Maithili and Bengali have characters that are semantically contrastive, but glyphically identical, eg. Maithili 𑒧 RA and Bengali 𑒧 BA, Maithili 𑒧 VA and Bengali 𑒧 RA. Other glyphical correspondences are not identical, but still similar: Maithili 𑒧 LA and Bengali 𑒧 NNA, Maithili 𑒧 NNA and Bengali 𑒧 LA. This issue affects not only characters, but also character elements. The element 𑒧 is a contextual form of VOWEL SIGN U in Maithili, but a conjoining form of BA in Bengali. Without appropriate linguistic context, the semantic value of the form 𑒧 is ambiguous; it can represent either Maithili SA + VOWEL SIGN U or Bengali SA + VIRAMA + BA.

## 4.4 Variant Characters

Maithili has variant forms of characters and conjuncts. For instance, the letter 𑒧 RA has two forms: (a) 𑒧 and (b) 𑒧. These variants produce consonant-vowel ligatures: (a) 𑒧 and 𑒧, and (b) 𑒧 and 𑒧. An example of a variant conjunct form is that of JA + VIRAMA + NYA, which is represented by both 𑒧 and 𑒧. Some character elements have variant forms, eg. the conjunct-onset form of 𑒧 TA is represented as 𑒧 and 𑒧. All variant forms for Maithili would be managed at the presentation level, as is done for other Indic scripts.

## 4.5 Rendering Requirements

### 4.5.1 Representation of Corresponding Characters

Although several Maithili characters have glyph shapes identical to Bengali characters, the manner of representing combinations of corresponding characters often differ in the two scripts.

- The letter CA has the shape 𑒧 in both Maithili and Bengali, but the sequence CA + VIRAMA + CA is written in Maithili as 𑒧 and as 𑒧 in Bengali.
- The letter TA has the shape 𑒧 in both scripts. The sequence TA + VIRAMA + YA is written in Maithili as 𑒧 and in Bengali as 𑒧.
- The letter MA is identical in both scripts, but Maithili has a specific shaping requirement for MA + VIRAMA at a word boundary, where the sequence is represented as 𑒧 not as 𑒧.

Another issue arises with the possibility of character sequences consisting of a mixture of Maithili- and Bengali-specific characters:

- NGA + VIRAMA + GA is represented in Maithili as ङ्ग and in Bengali as ঙ্গ. The letter গ GA is common between Maithili and Bengali, but NGA is ঙ in Maithili and ঙ in Bengali.
- RA + VIRAMA + GA, which is represented in Maithili as the ligature ঙ্গ, while in Bengali this combination is rendering using *repha* as ঙ্গ.

Given the requirement to represent Maithili in plain-text, such rendering behaviors cannot be considered a matter of font control or through application of language-specific rules. If Maithili and Bengali are unified, it will be necessary to establish methods of rendering sequences of corresponding characters that are represented differently in the two scripts, as well as sequences of mixed characters. Such rendering requirements could potentially be met through the use of control characters such as ZWJ. However, this would conflict with existing rules for producing conjuncts. Moreover, such solutions impose an additional burden upon users.

#### 4.5.2 Contextual Behaviors of Vowel Signs

- **Contextual forms of VOWEL SIGN U** The vowel sign U ू has the contextual form ्र, which is used with certain consonants and is written with these letters as a ligature.

	<i>gu</i>	<i>ju</i>	<i>nu</i>	<i>du</i>	<i>nu</i>	<i>pu</i>	<i>mu</i>	<i>lu</i>	<i>śu</i>	<i>su</i>
Maithili	गु	जु	नु	दु	नु	पु	मु	लु	शु	सु
Bengali	গু	জু	নু	দু	নু	পু	মু	লু	শু	সু

In Bengali, this element represents *ba-phala* and is used only in the creation of conjuncts. For example, শু represents *su* in Maithili, but *sva* in Bengali; in Maithili *sva* is written as শু.

- **Ligatures formed with VOWEL SIGN U** Special ligatures are used to write certain consonant-vowel combinations with U:

	<i>ku</i>	<i>tu</i>	<i>dhu</i>	<i>bhu</i>	<i>yu</i>	<i>yu</i>	<i>ru</i>	<i>śu</i>	<i>hu</i>
Maithili	कु	तु	धु	भु	यु	यु	रु	शु	हु
Bengali	কু	তু	ধু	ভু	যু	যু	রু	শু	হু

- **Ligature forms of VOWEL SIGN UU** The form of the sign for the vowel UU is ू. Most consonant-vowel combinations involving UU are written with the dependent vowel form, except for a few that are written as special ligatures:

	<i>kū</i>	<i>dhū</i>	<i>rū</i>	<i>hū</i>
Maithili	कु	धु	रु	हु
Bengali	কু	ধু	রু	হু

- **Ligature forms of VOWEL SIGN VOCALIC R** The dependent sign for VOCALIC R is ্র. Certain consonant-vowel combinations with VOCALIC R are written as ligatures:

	<i>kr</i>	<i>tr</i>	<i>br</i>	<i>bhr</i>	<i>hr</i>
Maithili	𑂗	𑂘	𑂙	𑂚	𑂛
Bengali	ক	ত	ব	ভ	হ

### 4.5.3 Contextual Behaviors of Consonants

Some Maithili consonants change shape in certain conditions.

Maithili uses *phala* or primitive forms for producing consonant conjuncts with particular consonants. Both Maithili and Bengali use *phala* forms of YA and RA, when these consonants are the final element of a consonant conjunct. The *phala* form of য YA is 𑂛; that of ব RA is 𑂙. However, in Maithili, *phala* forms are also used for NA, LA, and VA when writing conjuncts. This practice is maintained in Oriya, another script derived from Proto-Bengali.

**LETTER TA** There are two contextual forms of ত TA:

1. *Conjunct onset* When TA is the first consonant in a cluster its glyph shape is determined by the second consonant. If the second consonant is TA, YA, RA, or VA, the conjunct is written as a ligature. With all other consonants, TA is represented as the two-part form 𑂛. The explicit VIRAMA is written beneath the second consonant, eg. *saratkāla* is written in Maithili as सबकौल and in Bengali as সরৎকাল. A variant form attested in recent printed materials is written as 𑂛, eg. সবতকাল. This behavior of TA is similar to Bengali ত KHANDA TA; however, unlike KHANDA TA, it is unnecessary to encode 𑂛 as an atomic character because the contexts for its use are predictable.
2. *Word-final* When TA + VIRAMA occurs in a word-final position it is written as 𑂛 not as 𑂛.

**LETTER NA** Some consonant conjuncts involving ন NA are produced using the *phala* form 𑂛 of NA: eg *khna* 𑂛 and *phna* 𑂛 are two examples.

**LETTER BA** The letter 𑂛 BA has a contextual form when it appears in conjuncts and certain consonant-vowel combinations:

1. *Consonant-vowel combinations* The letter 𑂛 BA take the conjunct contextual form when it is written with VOWEL SIGN VOCALIC R. This combination is written not as \*𑂛, but as 𑂛.
2. *Conjunct onset* When 𑂛 BA appears as the first element of a conjunct with certain consonants it takes the shape 𑂛, eg. *bda* 𑂛.
3. *Conjunct final* When 𑂛 BA appears as the final element of a conjunct it is represented by 𑂛 *ba-phala*, eg. *mba* 𑂛, cf. *mu* 𑂛 and *mva* 𑂛.

**LETTER MA** When MA + VIRAMA occurs at a word-final position, the shape of 𑂛 changes. The combination MA + VIRAMA is written as 𑂛 not as 𑂛.

**LETTER YA** In consonant clusters where YA is the second element, it takes the *phala* form. The use of *ya phala* in Maithili is identical to that in Bengali. However, in Maithili the *ya phala* joins with consonants to form a distinct ligature. This feature was historically present in Bengali, but it is rare in modern orthography.

	<i>tya</i>	<i>dya</i>	<i>nya</i>	<i>bhya</i>	<i>śya</i>	<i>sya</i>	<i>hya</i>
Maithili	𑒧	𑒨	𑒩	𑒪	𑒫	𑒬	𑒭
Bengali	ত	দ্য	ন্য	ভ্য	শ্য	স্য	হ্য

**LETTER RA** Maithili *repha* is similar semantically to its analogue in Bengali, but it is graphically distinct. It is written above the following consonant letter: *rpa* in Maithili 𑒧 and in Bengali ৰ্ণ. In some cases, *repha* merges with the consonant letter with which it is written to create a distinct ligature: *rga* in Maithili 𑒧 and Bengali ৰ্ণ.

**LETTER LA** Some consonant conjuncts involving LA are produced using the 𑒧 *phala* form of LA. Some examples are *kla* 𑒧 (Bengali 𑒧), *phla* 𑒧 (Bengali 𑒧), *'sla* 𑒧 (Bengali 𑒧),

**LETTER VA** Some consonant conjuncts involving VA are produced using the *phala* form of VA: *va* is represented by the 𑒧 *va-phala*.

#### 4.5.4 Conjunct Forms

The Maithili and Bengali scripts have retained a more conservative position on the use of conjuncts. Even so, there are efforts to simplify the writing of conjuncts in Bengali, as has been done for Devanagari in the orthographies for Hindi and other modern languages. Maithili displays a tendency towards the retention of complex conjunct forms. This may change through increased use of the script in digital media. A comparison of conjunct forms in Maithili and Bengali is given in Table 5.

## 5 User Perception

In addition to the technical considerations that guide a determination on whether or not to unify Maithili and Bengali, factors such as user perception of script distinctions and the support and development of fonts and a character-encoding standard must be also considered.

### 5.1 User Perception of Script Distinctiveness

User perception of a script and its distinctiveness is an important issue to consider when unifying scripts. When the printing of Maithili books was begun in the 1920s, metal fonts for Maithili were developed in Calcutta by modifying existing Bengali fonts. Bengali fonts were widely available and the characters were similar enough to Maithili, so that the cost of developing a set of Maithili fonts from existing sources was lesser than producing a set of fonts anew. Similarly, when Maithili digital fonts were developed in Nepal in 2003, the corresponding forms in Bengali fonts were replaced by newly drawn Maithili-specific characters. If Maithili were perceived by its users as a variant of Bengali, printers and publishers would have simply produced Maithili books using Bengali metal and digital fonts. The distinction between the two scripts is also understood in Bengal, where Maithili is known as *tirute*, or ‘belonging to the Tirhuta region’.<sup>6</sup>

The disunification of Greek and Coptic rested upon user perceptions of the distinctions between scripts. Michael Everson and Kamal Mansour advocated the disunification of Coptic and Greek in N2444 (2002) on the principle that “unification of scripts should imply that readers of a language are able to make sense, with relative ease, of a text written in a variant of the script.”<sup>7</sup> This rationale may be applied to the case for

<sup>6</sup> Chatterji 1926: 225. <sup>7</sup> Everson and Mansour 2002.

disunifying Maithili and Bengali. Consider the following: the Bengali term for the Bengali language and script, বাংলা, is transliterated into Maithili as बाँला. While the style of the two words are similar and certain character elements are identical, the representations of the word *bāmlā* are entirely different. The difference is not one of style, but of character identity. A speaker of Bengali would not immediately recognize the name of his language as बाँला. The perceived differences between Maithili and Bengali suggests that despite the similarities, readers of Maithili and Bengali view their scripts not as variations upon a common model, but distinct writing systems belonging to distinct languages and linguistic communities.

## 5.2 Support and Development

User perception is critical for a character-encoding standard for any given script because the effectiveness of a standard is dependent upon its acceptance by the target user community. User perception regarding a script and the characters that constitute it is essential not only for the adoption of an encoding standard for a writing system, but also for continued support and development of fonts and other implementations. The unification of Maithili and Bengali would require that at least 40 Maithili-specific characters be added as extensions to the Bengali script. Such a unification binds the support and development of Maithili to the motivations and perceptions of Bengali font developers.

Even if Maithili were added as a separate block named “Bengali Extensions”, the support for such a block is dependent upon Bengali and implementers of Bengali. Extended characters are not always viewed positively. An important example is Meetei Mayek. The Government of Manipur, India rejected a proposal for Meetei Mayek because it disagreed with the presence of historical characters in the code block, which are not considered part of the official modern orthography for the script.<sup>8</sup> The characters in question were removed from the proposal and, ultimately, a proposal for adding an extended block of historical Meetei Mayek characters was also tabled. While political considerations should not guide character encoding, the standard for Meetei Mayek reflects the requirements of its users; the Government of Manipur being a key stakeholder. A standard encoding for Meetei Mayek containing characters, even those termed as extended characters, that are not perceived as part of the script by users was likely to be unadopted and unsupported.

With specific reference to Bengali, questions have been raised regarding the presence of historical characters in the Bengali block and the necessity of maintaining them. Most recently, Bidyut Baran Chaudhuri of the Society for Natural Language Technology Research (Kolkata, India) suggested reformation of the Bengali code block. He stated that “[m]any code points are unnecessarily filled by some old (historical) character or modifier signs.”<sup>9</sup> The reference is to the vocalic vowel letters. Chaudhuri suggested that these historical characters be placed in the Private Use Area so that code points may be made available for other characters, presumably those used in representing modern Bengali. If native Bengali users do not perceive the importance of maintaining historical Bengali characters, then presumptively there will be equal hesitation to accommodate and support Maithili characters, which are not used in Bengali.

## 6 Conclusions

The most appropriate method of representing Maithili in the UCS is to encode it as an independent script. While this approach would contradict the principle of unification, it would facilitate the representation of Maithili in plain text. An independent encoding for Maithili in the UCS will benefit the Maithili-speaking community by enabling it to adapt its traditional script for general communication and literary development through the digital medium. A standard encoding for the script will also enable scholars engaged in research on Maithili language and literature to preserve and reproduce Maithili manuscripts through digital technologies. Moreover, organizations in South Asia are actively developing fonts and standards for representing

<sup>8</sup> India. Government of Manipur 2008. <sup>9</sup> Chaudhuri 2009.



the Maithili script. A standard encoding in Unicode would benefit technical groups that are implementing the Maithili language and script for use in information technology. Although Maithili is a historical script, the Maithili-speaking community has preserved it for traditional purposes and has recently revived its use through digital publishing. An encoding for Maithili in the UCS will enable users of the language and script to meet the needs of modern communication and computing technologies.

## 7 References

- Chaudhuri, Bidyut Baran. 2009. “Proposals for discussion on Bengali Script” (L2/09-274). <http://www.unicode.org/L2/L2009/09274-bangla-agenda.pdf>.
- Chatterji, Suniti Kumar. 1926. *The Origin and Development of the Bengali Language*. Part I: Introduction, Phonology. Calcutta: Calcutta University Press.
- eCube Solutions & Research. 2003. “Tirhuta Lipi.” <http://www.tirhotalipi.4t.com/>.
- Everson, Michael and Kamal Mansour. 2002. “Coptic supplementation in the BMP” (N2444). <http://www.dkuug.dk/jtc1/sc2/wg2/docs/n2444.pdf>.
- India. Government of Manipur. 2008. “Proceeding of Meeting of Meetei Mayek Expert Committee Held on 17-07-2008” (L2/08-291). <http://www.unicode.org/L2/L2008/08291-meetei.pdf>.
- Pandey, Anshuman. 2006. “Request to Allocate the Maithili Script in the Unicode Roadmap” (L2/06-226). <http://www.umich.edu/~pandey/06226-maithiliroadmap.pdf>.
- Salomon, Richard. 1998. *Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and the Other Indo-Aryan Languages*. New York: Oxford University Press.
- Thakur, Gajendra. 2009. “Videha”, a fortnightly Maithili e-journal. <http://www.videha.co.in/>.
- Vikas, Om. 2005. “Issues in Representation of Indic Scripts in Unicode” (L2/05-063). <http://www.unicode.org/L2/L2005/05063-vikas.pdf>.
- Wikimedia. 2008. “Requests for new languages/Wikipedia Maithili.” [http://meta.wikimedia.org/wiki/Requests\\_for\\_new\\_languages/Wikipedia\\_Maithili](http://meta.wikimedia.org/wiki/Requests_for_new_languages/Wikipedia_Maithili).

## 8 Acknowledgments

I owe much gratitude to Dr. Ramanand Jha ‘Raman’ and Pandit Govinda Jha, both of Patna, India for providing materials on the Maithili script and for their patience in answering the numerous questions that I posed to them over the years. They are an invaluable resource on all things Maithili.

Gajendra Thakur of New Delhi graciously met with me and corresponded at length about Maithili, offered valuable specimens of Maithili manuscripts, printed books, and other records, and provided feedback regarding requirements for the encoding of Maithili in the UCS.

The first proposal for Maithili (L2/06-226) owes much to the advice and generosity of Dr. Dragomir Dimitrov of the Nepal Research Center, Kathmandu, Nepal, who provided me with Maithili materials printed in Nepal.

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 of the Humanities.

	MAITHILI	BENGALI	MAITHILI	BENGALI
A	अ	অ	<i>no dependent form</i>	
AA	आ	আ	া	া
I	इ	ই	ি	ি
II	ई	ঈ	ী	ী
U	उ	উ	ূ	ূ
UU	ऊ	ঊ	্	্
VOCALIC R	ऋ	ঋ	্	্
VOCALIC RR	ॠ	ঠ	্	্
VOCALIC L	ऌ	ঌ	্	্
VOCALIC LL	ॡ	ড	্	্
E	ए	এ	ে	ে
SHORT E	<i>no independent form</i>		ে	—
AI	ऐ	ঐ	ে	ে
O	ओ	ও	ৌ	ৌ
SHORT O	<i>no independent form</i>		ৌ	ৌ
AU	औ	ঔ	ৌ	ৌ

Table 1: Comparison of Maithili and Bengali vowel letters and signs (differences highlighted).

	MAITHILI	BENGALI		MAITHILI	BENGALI
KA	क	ক	DA	দ	দ
KHA	ख	খ	DHA	ধ	ধ
GA	ग	গ	NA	ন	ন
GHA	घ	ঘ	PA	প	প
NGA	ङ	ঙ	PHA	फ	ফ
CA	च	চ	BA	ब	ব
CHA	छ	ছ	BHA	भ	ভ
JA	ज	জ	MA	म	ম
JHA	झ	ঝ	YA	य	য
NYA	ञ	ঞ	YYA	य्य	য়
TTA	ट	ট	RA	र	র
TTHA	ठ	ঠ	LA	ल	ল
DDA	ड	ড	LLA	ल्ल	—
DDDA	ड़	ड़	VA	व	—
DDHA	ढ	ঢ	SHA	श	শ
RHA	ऱ	ৱ	SSA	ष	ষ
NNA	ण	ণ	SA	स	স
TA	त	ত	HA	ह	হ
THA	थ	থ			

Table 2: Comparison of Maithili and Bengali consonant letters (differences highlighted).



	MAITHILI	BENGALI		MAITHILI	BENGALI
<i>kta</i>	क	ক	<i>ṅḍha</i>	ঙ	ঙ
<i>kra</i>	क्रे	ক্র	<i>tta</i>	ত	ত
<i>kṣa</i>	क़	ক্ষ	<i>tva</i>	ত্ব	ত্ব
<i>gga</i>	ग	গ	<i>ddha</i>	द	দ
<i>gdha</i>	द	দ	<i>ndha</i>	न	ন
<i>ṅka</i>	कं	কং	<i>nha</i>	न्ह	নহ
<i>ṅkha</i>	कं	কং	<i>bda</i>	द	দ
<i>ṅga</i>	गं	গং	<i>bdha</i>	द	দ
<i>ṅgha</i>	गं	গং	<i>ścha</i>	च	চ
<i>cca</i>	च	চ	<i>ṣṭa</i>	ष	ষ
<i>ccha</i>	च	চ	<i>ṣṭha</i>	ष	ষ
<i>jña</i>	ज	জ	<i>ṣṇa</i>	ष	ষ
<i>ñca</i>	च	চ	<i>rga</i>	र	র
<i>ñcha</i>	च	চ	<i>hṇa</i>	ह	হ
<i>ṭṭa</i>	ट	ট	<i>hna</i>	ह	হ
<i>ṭṭha</i>	ट	ট	<i>hma</i>	ह	হ
<i>ṇṭa</i>	ण	ণ	<i>hra</i>	ह	হ
<i>ṇṭha</i>	ण	ণ	<i>hla</i>	ह	হ
<i>ṇḍa</i>	ण	ণ	<i>hva</i>	ह	হ

Table 5: Comparison of conjunct forms used in Maithili and Bengali

	1148	1149	114A	114B	114C	114D
0	◌̇ 11480	◌̈ 11490	◌̉ 114A0	◌̊ 114B0	◌̋ 114C0	◌̌ 114D0
1	◌̍ 11481	◌̎ 11491	◌̏ 114A1	◌̐ 114B1	◌̑ 114C1	◌̒ 114D1
2	◌̓ 11482	◌̔ 11492	◌̕ 114A2	◌̖ 114B2	◌̗ 114C2	◌̘ 114D2
3	◌̙ 11483	◌̚ 11493	◌̛ 114A3	◌̜ 114B3	◌̝ 114C3	◌̞ 114D3
4	◌̟ 11484	◌̠ 11494	◌̡ 114A4	◌̢ 114B4	◌̣ 114C4	◌̤ 114D4
5	◌̥ 11485	◌̦ 11495	◌̧ 114A5	◌̨ 114B5	◌̩ 114C5	◌̪ 114D5
6	◌̫ 11486	◌̬ 11496	◌̭ 114A6	◌̮ 114B6	◌̯ 114C6	◌̰ 114D6
7	◌̱ 11487	◌̲ 11497	◌̳ 114A7	◌̴ 114B7	◌̵ 114C7	◌̶ 114D7
8	◌̷ 11488	◌̸ 11498	◌̹ 114A8	◌̺ 114B8	◌̻ 114C8	◌̼ 114D8
9	◌̽ 11489	◌̾ 11499	◌̿ 114A9	◌̺̇ 114B9	◌̻̇ 114C9	◌̼̇ 114D9
A	◌̽̇ 1148A	◌̾̇ 1149A	◌̿̇ 114AA	◌̺̈ 114BA	◌̻̈ 114CA	
B	◌̽̈ 1148B	◌̾̈ 1149B	◌̿̈ 114AB	◌̺̉ 114BB		
C	◌̽̉ 1148C	◌̾̉ 1149C	◌̿̉ 114AC	◌̺̊ 114BC		
D	◌̽̊ 1148D	◌̾̊ 1149D	◌̿̊ 114AD	◌̺̋ 114BD		
E	◌̽̋ 1148E	◌̾̋ 1149E	◌̿̋ 114AE	◌̺̌ 114BE		
F	◌̽̌ 1148F	◌̾̌ 1149F	◌̿̌ 114AF	◌̺̍ 114BF		

Table 6: Proposed code chart for Maithili

**Various signs**

11480	◌◌	MAITHILI SIGN CANDRABINDU
11481	◌◌	MAITHILI SIGN ANUSVARA
11482	◌ः	MAITHILI SIGN VISARGA

**Independent vowels**

11483	अ	MAITHILI LETTER A
11484	आ	MAITHILI LETTER AA
11485	इ	MAITHILI LETTER I
11486	उ	MAITHILI LETTER II
11487	ऋ	MAITHILI LETTER U
11488	ॠ	MAITHILI LETTER UU
11489	ऋ	MAITHILI LETTER VOCALIC R
1148A	ॠ	MAITHILI LETTER VOCALIC RR
1148B	ऌ	MAITHILI LETTER VOCALIC L
1148C	ॡ	MAITHILI LETTER VOCALIC LL
1148D	ए	MAITHILI LETTER E
1148E	ऐ	MAITHILI LETTER AI
1148F	ओ	MAITHILI LETTER O
11490	औ	MAITHILI LETTER AU

**Consonants**

11491	क	MAITHILI LETTER KA
11492	ख	MAITHILI LETTER KHA
11493	ग	MAITHILI LETTER GA
11494	घ	MAITHILI LETTER GHA
11495	ङ	MAITHILI LETTER NGA
11496	च	MAITHILI LETTER CA
11497	छ	MAITHILI LETTER CHA
11498	ज	MAITHILI LETTER JA
11499	झ	MAITHILI LETTER JHA
1149A	ञ	MAITHILI LETTER NYA
1149B	ट	MAITHILI LETTER TTA
1149C	ठ	MAITHILI LETTER TTHA
1149D	ड	MAITHILI LETTER DDA
1149E	ढ	MAITHILI LETTER DDDA
1149F	ण	MAITHILI LETTER DDHA
114A0	र	MAITHILI LETTER RHA
114A1	ल	MAITHILI LETTER NNA
114A2	व	MAITHILI LETTER TA
114A3	ष	MAITHILI LETTER THA
114A4	द	MAITHILI LETTER DA
114A5	ध	MAITHILI LETTER DHA
114A6	न	MAITHILI LETTER NA
114A7	प	MAITHILI LETTER PA
114A8	फ	MAITHILI LETTER PHA
114A9	ब	MAITHILI LETTER BA
114AA	भ	MAITHILI LETTER BHA
114AB	म	MAITHILI LETTER MA
114AC	य	MAITHILI LETTER YA
114AD	र	MAITHILI LETTER YYA
114AE	व	MAITHILI LETTER RA
114AF	ल	MAITHILI LETTER LA
114B0	ल	MAITHILI LETTER LLA
114B1	व	MAITHILI LETTER VA
114B2	श	MAITHILI LETTER SHA
114B3	ष	MAITHILI LETTER SSA
114B4	स	MAITHILI LETTER SA
114B5	ह	MAITHILI LETTER HA

**Dependent vowel signs**

114B6	ा	MAITHILI VOWEL SIGN AA
114B7	ि	MAITHILI VOWEL SIGN I
114B8	ी	MAITHILI VOWEL SIGN II
114B9	ु	MAITHILI VOWEL SIGN U
114BA	ू	MAITHILI VOWEL SIGN UU
114BB	ृ	MAITHILI VOWEL SIGN VOCALIC R
114BC	ॠ	MAITHILI VOWEL SIGN VOCALIC RR
114BD	ॡ	MAITHILI VOWEL SIGN VOCALIC L

114BE	ो	MAITHILI VOWEL SIGN VOCALIC LL
114BF	े	MAITHILI VOWEL SIGN E
114C0	े	MAITHILI VOWEL SIGN SHORT E
114C1	ै	MAITHILI VOWEL SIGN AI
114C2	ो	MAITHILI VOWEL SIGN O
114C3	ो	MAITHILI VOWEL SIGN SHORT O
114C4	ौ	MAITHILI VOWEL SIGN AU

**Various signs**

114C5	्	MAITHILI SIGN VIRAMA
114C6	ॲ	MAITHILI SIGN AVAGRAHA
114C7	◌◌	MAITHILI ABBREVIATION SIGN
114C8	।	MAITHILI ANJI
114C9	ॐ	MAITHILI OM
114CA	॰	MAITHILI GVANG

**Digits**

114D0	०	MAITHILI DIGIT ZERO
114D1	१	MAITHILI DIGIT ONE
114D2	२	MAITHILI DIGIT TWO
114D3	३	MAITHILI DIGIT THREE
114D4	४	MAITHILI DIGIT FOUR
114D5	५	MAITHILI DIGIT FIVE
114D6	६	MAITHILI DIGIT SIX
114D7	७	MAITHILI DIGIT SEVEN
114D8	८	MAITHILI DIGIT EIGHT
114D9	९	MAITHILI DIGIT NINE

Table 7: Proposed names list for Maithili