

SCRIPT GRAMMAR FOR MAITHILI LANGUAGE

Prepared by

Technology Development for Indian Languages (TDIL) Programme

Department of Information Technology, Government of India

in association with

Centre for Development of Advanced Computing (C-DAC)

Table of Contents

0. INTRODUCTION	3
1. OBJECTIVES OF SCRIPT GRAMMAR	4
2. END USERS FOR SCRIPT GRAMMAR	5
3. SCOPE	6
4. TERMINOLOGY	7
5. PHILOSOPHY AND UNDERLYING PRINCIPLES.....	11
6. SCRIPT GRAMMAR STRUCTURE.....	12
6.1. PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR.....	13
6.2. CONFORMITY TO THE SYLLABLE STRUCTURE	14
6.3 SCRIPT GRAMMAR PROPER.....	18
6.3.1. The Character Set of Maithili.	18
6.3.2. Consonant Mātrā Combinations.	25
6.3.3. The Ligature Set of Maithili.	38
6.3.4 The Collation Order of Maithili.....	47
7. REFERENCES	48
8. ANNEXURES	49
Annexure 1: Names of experts who have contributed to the script grammar.....	49
Annexure 2: Unicode Table of Devanāgarī	50

0. INTRODUCTION

The term **script grammar** refers to the behaviour pattern of the writing system of a given language. Languages which have written representations do not use a haphazard manner of storing the information within the system, but use a coherent pattern which is similar to the linguistic grammar of a given language. With the help of specialists (not necessarily linguist) who work in the area of the written representation of the language, the manner in which the shapes of the characters of the language and the representation of the conjunct forms is provided. In other words the Script Grammar deals with the surface structure of the language and tries to provide the best possible “fit” for shapes and their representation. Since this is a highly subjective issue, the shapes provided here are recommendations at the best and conform to the perception of the mandating body/evaluators who consensually arrive at the “best possible fit” which is acceptable to a majority of users. An example from the Devanāgarī script will make the above clear. Although Marathi and Hindi share the same script Devanāgarī, not only do they not share the same character inventory but in addition the representation of certain characters is different. Thus the Hindi /la/ is different from the Marathi /la/ in so far as the placement of the stem is concerned Hindi ल Marathi ल. This ensures that the Script Grammar conforms to the language in question and provides the character shapes acceptable to a given user community. It should be noted that this does not mean monotony. The Marathi and Hindi /la/ can have a variety of forms once the intrinsic structure of the character is determined.

Script Grammar is the term used to define:

- the writing system used to inscribe a given language
- the history of the script and language (wherever available)
- the syllabic structure of the writing system of the language
- the rule ordering of the characters within the syllable (akshar)
- description of the syllabic clusters
- collation order of the characters: lexical / dictionary sorting order

1. OBJECTIVES OF SCRIPT GRAMMAR

The Objectives of the script grammar for each language can be divided into two major parts:

Societal:

- Provide a visual representation of shapes that are deemed to be in conformity with the perception of a given community
- Ensure thereby that this perception is safe-guarded
- Through wide-spread dissemination and creation of appropriate tools ensure that within the given linguistic community, all media tries to adopt the given shape.

Technical:

- Classify the language in terms of its ISO and also whether it belongs to the Abjad, Akshar (Alphasyllabary) class.
- Provide an inventory of the characters pertinent to the language and classify the same in terms of their taxonomy.
- As a corollary determine whether the inventory is in conformity to the Syllable formalism as stipulated in ISCI'91 and subsequently adopted by Unicode.
- Since Brahmi is written from left to right, and since certain characters do not follow the linear L to R order, provide an inventory of displaced concatenators i.e. characters such as Mātrās that concatenate to the Consonant
- Propose the best shape representation of the individual characters as well as of the ligatures used within a given script. As a corollary request the expert(s) to identify the largest possible strings of such ligatures.
- Finally provide the collation order pertinent to that script/language, which would be of great utility to high-end NLP as well as to CLDR's in the pertinent language. The collation order for Nepali is different from Hindi although both languages share the same script. Thus in Nepali क्ष, ज्ञ are placed at the end of the consonant inventory i.e. after ह in the sort order. In Hindi क्ष is sorted along with क and ज्ञ with ज

2. END USERS FOR SCRIPT GRAMMAR

The script-grammar specific to a given language can be used by a large number of users.

- Most importantly it can be used by font developers desirous of developing a font which is compliant with the perception of the characters and ligatures of a language by its user community.
- Certain features of the script grammar such as the shapes can also be used for testing OCR and OHWR. Similarly information regarding Ligatures as well as collation order can help in high-end NLP work such as detecting invalid combinations, correct implementation of syllable structure, prediction routines to name a few. Information regarding collation and character sets can be also used for CLDR.
- They allow the font designer to design a font which is in compliance with the norms and standards of that particular script. A major problem which will be dealt with in the template is one of ligatures. The final list of ligatures defined by the script grammar allows the font designer to write specific rules for such glyphs.
- It permits the software developer to design and implement the keyboard and the input mechanism which will meet the requirement of the particular linguistic community.
- The collation or sort order as described in a Script Grammar permits the software developer to write software functions/ routines for sorting data in all applications.
- Script Grammars are equally important for keyboard design, especially when supplemented by frequency data from a corpus.

As can be seen the script grammar has a wide range of use and can be of utility to font developers, Indian language developers and linguists in the area of computation.

3. SCOPE

This script grammar document contains following information about the language and the script used for writing the language.

1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1 & 639.3 standard.
2. Script used to inscribe the given language
3. The structure of the script used for writing the language
 - Rule ordering of the characters within the syllable formation is a language
 - Description of the syllabic clusters of the script
 - Collation order of the characters: lexical / dictionary sorting order
 - Compliance of the script with Unicode.

These will be treated within the relevant sections of the script grammar

4. TERMINOLOGY¹

Abjad: A writing system in which each symbol always or usually stands for a consonant. The long vowels are indicated. However the short vowels are rarely marked and the reader needs to supply these. Example: Urdu written in Perso-Arabic Script is an example of this writing system.

Abugida: also called an alphasyllabary, is a segmental writing system in which consonant–vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel notation is obligatory but secondary²

Akshar: see **Abugida**

Allographs: Variants of the representation of a character. Thus æ and æ [U+00E6] in Latin alphabet are allographs.

Allo-Script: The term relates to languages which share a common script. Thus Devanāgarī is used to write 9 official languages. However these languages do not use the same set of characters. Thus Marathi uses the retroflex lla ऌ [U+ 0933] which Hindi does not use. Flaps used in Hindi ञ [U+095C] ण [U+095D] are not used in Konkani. These sub-sets of scripts based on a single “matricial” script are termed as allo-scripts.

Alphabet: A set of letters used in writing a language. Example: The English Alphabet.

Aspirated consonant: A consonant which is pronounced with an extra puff of air coming out at the time of release of the oral obstruction. This has a sound of an extra "h".

Basic alphabet: The minimal set of letters which can be used for uniquely encoding every word of a language. The basic alphabet for English consists of only the upper-case letters A-Z

Catenators: Also termed as Concatenators are characters which are concatenated to another character. In the Brahmi script these are the Mātrās or Vowel modifiers which are adjoined to the consonant and add a vocalic value to the consonant.

Conjunct: The Indic scripts are noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent letterforms. This abbreviation takes place only in the context of a consonant cluster....Under normal circumstances, a consonant cluster is depicted with a conjunct glyph if such a glyph is


¹ As in the case of the BIS Document, in order to make the terminology accessible for all readers, examples have been chosen from English/Latin scripts, wherever possible. Some definitions have been excerpted from the BIS ISCI91 document and suitably modified where necessary.

² Wikipedia definition

available in the current font. In the absence of a conjunct glyph, the one or more dead consonants that form part of the cluster are depicted using half-form glyphs. In the absence of half-form glyphs, the dead consonants are depicted using the nominal consonant forms combined with visible virama signs.³

Consonant: A letter representing a speech sound in which the breath is at least partly obstructed.

Diacritic: A mark added to a letter which distinguishes it from the same letter without a mark, usually having a different phonetic value or stress.

Displaced Catenator: (see Catenator) Within the Brahmi script, the writing system is linear and moves from left to right. However in the case of some catenators this rule is not observed and the catenator (wholly or partially) is placed to the right of the consonant to which it relates. The short vowel I // in Devanāgarī is an example of a displaced catenator.

Display composing: The process of organizing the basic shapes available in a font in order to display (or print) a word.

Display rendition: The process by which a string of characters is displayed (or printed). In this process several consecutive characters may combine with each other on the screen. The sequence of display of the characters may become different.

Eyebrow repha: (See **Eyesh ra**)

Eyesh ra: The eyesh ra is used in Konkani, Nepali and Marathi. It is treated as different from the र (repha) by certain linguists. While the former is treated as a flap, the latter is a continuant trill (*cf.*, Kalyan Kale and Anjali Soman. 1986).

Font: A set of symbols used for display or printing of a script in a particular style.

International numerals: The conventional 0 to 9 digits used in English for denoting numbers. These are also known as Indo-Arabic numerals (to differentiate them from the Roman numerals like IX for 9).

Latin alphabet: The alphabet used for writing the language of ancient Rome. Also known as the Roman alphabet. The alphabet is used today for writing English and European languages.

Letter: A character representing one or more of the simple or compound sounds used in speech. It can be any of the alphabetic symbols.

Ligature: (see **Conjunct**)

³ Unicode ver. 6.0 Chapter 9.0 pp 6-7

Nasal consonant: A consonant pronounced with the breath passing through the nose. Example *m n* in English.

Nasalized vowel: A vowel pronounced with the breath passing both through the nose and the mouth. In Indian scripts this is denoted by a Chandrabindu and gives the vowel/vowel modifier over which it placed a nasal value. Example: जाँच

Phonetic alphabet: An alphabet which has direct correspondence between letters and sounds Example: The International Phonetic Alphabet..

Pure consonant: A consonant which does not have any vowel implicitly associated with it.

Rafar: A special case of a ligature constituted by the adjunction of ra followed by a halanta to consonant. The resultant combination places the ra on top of the consonant to which it is adjoined. In case the consonant itself is adjoined to another consonant, the rafar is placed above the consonant e.g. र्+क कर्क , र्+घ्+य घ्यर्

Rakar: A special case of a ligature constituted by the adjunction of a consonant followed by a halanta to ra. In a large number of Brahmi scripts the ra is adjoined to the stem of consonant to which it relates. In the case of consonants which have no stem such as the dental retroflexes in Devanāgarī, the rakar is placed below the consonant to which it relates.

Repha: (see **Rafar**)

Roman script: The script based on the ancient Roman alphabet, with the letters A-Z and additional diacritic marks. Used for writing a language which is not usually written in the Roman alphabet.

Script: A distinctive and complete set of characters used for the written form of one or more languages.

Script numerals: The 0 to 9 digits in a script, which have shapes distinct from their international counterparts.

Syllable: A unit of pronunciation uttered without interruption, forming whole or part of a word, and usually having one vowel or diphthong sound optionally surrounded by one or more consonants

Transliteration: Representation of words with the closest corresponding letters in an alphabet of a different language.

Vowel: A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction

Vowel sign: A graphic character associated with a letter, to indicate a vowel to be associated with that character (Mātrā in Hindi).

5. PHILOSOPHY AND UNDERLYING PRINCIPLES

The script grammar is based on the following principles:

1. The Grammar aims to depict the surface grammar of the written language: the manner in which characters as well as conjuncts are depicted
2. Where a given script admits many languages, it is pre-suppose that such languages will prescribe different representations for a given shape or conjunct according to the perception of the native users of that language
3. Corollary to the above the result is a script and allo-scripts i.e. a given script shared by many languages is not uniformly deployed across all the languages but is subject to variations and modulations.
4. The term Grammar is used here in a non-normative sense: what is prescribed is in the form of recommendations provided by experts who visualize the shape of the given script in their mother tongue in a specific manner. Subjective variations may occur⁴
5. The Grammar is limited to its synchronic use i.e. the manner in which a given language as of today admits a character set within the script used to write it. It is not diachronic or historical in nature and does not study the evolution of the given script across centuries.

⁴ It is recommended that such variations be culled by placing the Grammars of different scripts in public review.

6. SCRIPT GRAMMAR STRUCTURE

The script grammar provided below has the following parts.

Part 6.1. deals with peripheral elements such as the ISO of the language, the writing system used: (Alphasyllabic) Abugida or Abjad.

Part 6.2. treats of the syllabic structure. It verifies whether the character set of the language complies with the ISCII syllabic structure and if not which cases are not compliant.

Part 6.3 is the script grammar proper and describes the character set as well as the conjunct shapes of the given script along with the collation order

6.1. PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR

These constitute the elements that are peripheral to the Script Grammar. The main parameters considered are the mnemonic and name of the language (needed for CLDR and also for language tags), the writing system used to inscribe the language and wherever possible a short history of the language.

6.1.1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1. & 639.3

Name of the Language: MAITHILI
ISO Mnemonics: *mai*

This refers to a one line description of the language and its mnemonic representation as per the ISO.

6.1.2. Identification of the writing system(s) used to inscribe the given language Maithili is written using the Devanāgarī script. It is an alphasyllabary with the akshar as its core.

This is a one line description of the script used to write the language. However in case the language uses more than one script, all the scripts in question are specified, provided these constitute the official language of the given state.

All scripts derived from Brahmi are Abugidas i.e. syllabary driven systems. The main features of Abugidas are as under:

- The consonant has an implicit vowel built-in which is normally the schwa.
- The inherent vowel can be modified by the addition of other vowels or muted by a diacritic termed as a Virama or Halanta
- Vowels can be handled as full vowels with a vocalic value
- When two or more consonants join together they form ligatures which can be recognized by their shape क्ल or alternatively form an entirely new shape क्ख = क्ष.

Abugidas/Alphasyllabaries because of their syllabic structure require a special description which is the subject of the discussion in 6.2. below.

6.1.3. Amendments needed in Unicode for Maithili language

None has been proposed by the experts who have mandated the script grammar.

6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

Maithili language complies with the syllable (akshar) structure described above. It can admit up to 3 consonant clusters.

Alphasyllabaries are determined by the notion of the syllable or the Akshar. The compositional grammar of the syllable determines its well-formedness. This is through a series of formal constraints based on a Backus-Naur Formalism which is given below. The syllable (akshar), first defined in the ISCII document (1991), identifies the following character ‘sub-sets’ for the purposes of identifying the syllable (akshar). In what follows the syllable analysis will be restricted to Maithili.

(C) Consonants

क	ख	ग	घ	ङ
च	छ	ज	झ	ञ
ट	ठ	ड	ड़	ढ
ढ़	ण	त	थ	द
ध	न	प	फ	ब
भ	म	य	र	ल
व	श	ष	स	ह
ज़ ⁵				

(V) Vowels

अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ	औ
---	---	---	---	---	---	---	---	---	---	---

(M) Mātrās or Vowel Modifiers

ा	ि	ी	ु	ू	ृ	े	ै	ो	ौ
---	---	---	---	---	---	---	---	---	---

(D) Diacritics

◌ं: Anuswar	Anuswara, a nasal, is denoted by a dot above the letter after which it is to be pronounced. This falls under Nasal category.
-------------	--

⁵ The ज़ is a new addition to Maithili and is attested both in the press as well as on blog sites and is used to render the English loan words with the sound /z/

ँ: Chandrabindu	Chandrabindu, a nasal, is denoted by a breve with a dot superposed above the letter after which it is to be pronounced. This falls under Nasal category.
ः Visarga	Visarga, denoted by two dots placed above the other.
ऌ: Avagraha	The avagraha sign, U+093D is used to indicate extra-long vowels in Maithili.
' <i>bikari kaamaa</i>	modifier letter apostrophe is used to indicate the prolongation of a short a and to indicate the truncation of words. This sign is called bikari kaamaa

(H):Halanta ् - Halant used in most writing systems to signify the lack of an inherent vowel.

(N)⁶ Nukta ् - used in Maithili

Each of these sub-types has its restrictions in terms of what can precede or follow it, within a syllable (akshar), as shown in the table below:

PRECEDED BY	SUBTYPE	FOLLOWED BY
-, H	C	N,M,D,H
C	N	C,M,D,H
-	V	D
C, N	M	D
C, N,V,M	D	-
C, N	H	C

C can be preceded by H or no subtype and followed by any one of the following: M,D,H

N can be preceded by C and followed by any one of the following: N,M,D,H

V can be preceded by no subtype and followed by D but not by another sub-type.

M can be preceded by C and followed by D.

D can be preceded by C, V, M and followed by no other subtype. It closes the syllable (akshar).

H can be preceded by C alone and followed only by C and no other sub-set.

Syllable (akshar) Types

The formalism defines the syllable (akshar) in terms of both what can constitute a syllable (akshar) and what cannot. A valid syllable (akshar) as per this definition can be of only two types:

A vowel syllable (akshar): a full vowel.

A consonant syllable (akshar): a full consonant (having a weak vowel or a mātṛā)

⁶ The nukta is a small dot placed under a character in certain scripts to show that they are flapped or for deriving consonants required for representing Urdu loan words क़, ख़, ग़, ज़, फ़

The four other subsets viz. Mātrās, Vowel Modifiers, Halanta and Nukta cannot constitute a syllable (akshar) by themselves or in combination among themselves.

1. The Vowel syllable (akshar) is of the following types:

1.1. A pure vowel all by itself: अ, /a/ आ /ā/ etc.

1.2. A vowel followed by a modifier i.e. either a nasal marker (anunasika or anuswara) or a visarga: ई /ī /, आः /āH/

2. The Consonant syllable (akshar) can be of the following types:

2.1. A full consonant (with or without Nukta) i.e. with the inherent vowel : क : /ka/

2.2. A consonant⁷ (with or without Nukta) followed by a mātrā i.e. the inherent vowel being substituted by another vowel: की /ki:/, की /qi:/

2.3. A consonant (with or without Nukta) followed by a modifier: कं /k̃/, हः /haH⁸/

2.4. A consonant (with or without Nukta) followed by a mātrā and a modifier: कुं /kū/, दुः /duH/.

2.5. A consonant cluster i.e. a dead or half consonant (Consonant+Halanta) followed by a full consonant followed optionally by a mātrā, a modifier or a combination of both. These result in a ligature or what is often termed as yuktakshara.

त्क /tka/ त्कं /tk̃/, त्कः /tkaH/ त्कुं /tkū/, त्दुः /tdu/.

The above permutations and combinations result in 7 major syllable (akshar) types. Of these the last type introduces the problem of the number of consonant clusters. ISCII (91, p.23) provides for up to three consonant clusters as the “worst case” i.e. the largest possible string. This is functional for modern prakrits where the largest consonantal cluster rarely exceeds three consonant.

This means that theoretically the following forms can be postulated:

Vowel Set: With the Vowel as the node.

V VD

Consonant set: With the Consonant as the node (an implicit or modified vowel is pre-implied).

Node	Mātrā	Modifier	Mātrā+Modifier
C ⁹	CM	CD	CMD
CHC	CHCM	CHCD	CHCMD
CHCHC	CHCHCM	CHCHCD	CHCHCMD
-	-	-	-

A total number of 12 theoretical syllables is therefore possible. It will be seen that the written syllable (akshar) is not very different in structure from the phonetic syllable and that the movement from the written to the spoken levels is made feasible by application of certain rules.

⁷ For purposes of Simplification, C here will automatically be treated as being also consonant+nukta C+ N

⁸ This character represents phonetically the weak implicit vowel, termed as schwa and often shown as /a/ also.

⁹ C here will automatically be treated as being also consonant+nukta, C+N to simplify the explanation

Since the formal structure script grammar of the syllable (akshar) is common to all Brahmi based scripts, it will not be treated in the sample template, but it will form the basis of an exhaustive description of the characters as well as their ligatural representations.

6.3 SCRIPT GRAMMAR PROPER

This section lays down in detail the different parameters of the Script Grammar for Maithili. These are:

- 6.3.1. The Character Set of Maithili.
- 6.3.2. The Consonant mātrā combinations of Maithili.
- 6.3.3. The Ligature Set of Maithili.
- 6.3.4. Collocation Order of Maithili

6.3.1. The Character Set of Maithili.

This section provides detailed information about the characters in the language and the list of the same and also more importantly shows the manner in which the character is to be written. Each subsection comprises therefore two parts: the basic character set and the shape each character should have, as mandated by the experts who have designed the script grammar of Maithili.

This comprises the following:

- 6.3.1.1. The Consonant Set
- 6.3.1.2. The Vowel Set
- 6.3.1.3. The Mātrā Set
- 6.3.1.4. Displaced Catenators
- 6.3.1.5. Shape of the combination of ra (rakar, repha)
- 6.3.1.6. The Set of Diacritics
- 6.3.1.7. Numerals
- 6.3.1.8. Punctuation marks
- 6.3.1.9. Other symbols

Each of these will be analysed in detail:

6.3.1.1. The Consonant Set

The Consonant set of Maithili comprises the following characters:

A basic Consonant inventory arranged as per their Vargas.

	STOPS				NASAL
	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	+voiced -aspirated
Velar	क	ख	ग	घ	ङ
Palatal	च	छ	ज	झ	ञ
Retroflex	ट	ठ	ड	ढ	ण
Flaps			ड़	ढ़	
Dental	त	थ	द	ध	न
B-labial	प	फ	ब	भ	म

Other consonants

य	र	ल	व	श
ष	स	ह	ज़ ¹⁰	

The exact shapes as desired by the experts are provided in the table below:

	STOPS				NASAL
	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	+voiced -aspirated
Velar	क	ख	ग	घ	ङ
Palatal	च	छ	ज	झ	ञ
Retroflex	ट	ठ	ड	ढ	ण
Flaps			ड़	ढ़	
Dental	त	थ	द	ध	न
B-labial	प	फ	ब	भ	म

Other consonants

य	र	ऌ	ल	व
श	ष	स	ह	

6.3.1.2. The Vowel Set

The Vowel set of Maithili is as under:

अ	DEVANAGARI LETTER A
आ	DEVANAGARI LETTER AA
इ	DEVANAGARI LETTER I

¹⁰ The ज़ is a new addition to Maithili and is attested both in the press as well as on blog sites and is used to render the English loan words with the sound /z/. See footnote 5 above

ई	DEVANAGARI LETTER II
उ	DEVANAGARI LETTER U
ऊ	DEVANAGARI LETTER UU
ऋ	DEVANAGARI LETTER VOCALIC R
ए	DEVANAGARI LETTER E
ऐ	DEVANAGARI LETTER AI
ओ	DEVANAGARI LETTER O
औ	DEVANAGARI LETTER AU
ँ ¹¹	DEVANAGARI LETTER CHANDRA E
ॉ ⁹	DEVANAGARI LETTER CHANDRA O

As per expert recommendations the character set should be written as under:

अ आ इ ई उ ऊ
ऋ ए ऐ ओ औ ँ ऑ

6.3.1.3. The Mātrā Set

The Mātrā (Vowel Modifier Set) of Maithili is as under:

Mātrā Names	Mātrās Sign	Where is it used ?	Consonant Shapes formed
1. Devanāgarī sign AA	ा	आ	क् + आ = का
2. Devanāgarī sign I (stands to the left of the consonant)	ि	इ	क् + इ = कि
3. Devanāgarī sign II	ी	ई	क् + ई = की
4. Devanāgarī sign U	ु	उ	क् + उ = कु

¹¹ The vowels ँ, ऑ and their corresponding vowel signs ँ, ॉ for foreign loan words which were earlier not used are becoming popular in usage and are hence included. Examples: पॉप, कैरियर, फॉर ऐण्ड etc.

5. Devanāgarī sign UU	ू	ऊ	क् + ऊ = कू
6. Devanāgarī sign vocalic R	ृ	ऋ	क् + ऋ = कृ
7. Devanāgarī sign E	े	ए (There can not be any matra on this)	क् + ए = के
8. Devanāgarī sign AI	ै	ऐ (There can not be any matra on this)	क् + ऐ = कै
9. Devanāgarī sign O	ो	ओ	क् + ओ = को
10. Devanāgarī sign AU	ौ	औ	क् + औ = कौ
11 Devanāgarī sign Candra E	ँ	ऐँ	क् + ँ = कैँ
12 Devanāgarī sign Candra O	ॉ	औँ	क् + ऌ = कौँ
13 Modifier Letter Apostrophe		Used as a syncopation mark ¹²	क + ' = क

As per expert recommendations the character set should be written as under:

ा ि िी ु ू ृ े ै ो ौ ँ ऌ

6.3.1.4. Displaced Catenators

Under normal circumstances Vowel Modifiers also known as catenators (since they concatenate to the preceding consonant) in Brahmi based scripts are written from left to right in linear order (with the exception of Consonant stacks). However certain modifiers are displaced and are placed to the left of the consonant to which they concatenate. As a general rule in all Devanāgarī script driven languages there is only one displaced

catenator: ि

CATENATOR	POSITION	EXAMPLE
ि	To left of character	कि, रि, पि

6.3.1.5. Shape of the combination of ra (rakar, repha)

The र takes a variety of shapes known as rakar and repha (rafar) depending on its position. When conjoined before a consonant by means of the halanta, it changes shape

¹² In Maithili, U+02BC “ ’ ” modifier letter apostrophe is used to indicate the prolongation of a short a and to indicate the truncation of words. This sign is called bikari kaamaa (Unicode 6.0 ch. 9 p.282)

and is placed on top of the consonant or consonant clusters to which it relates. This is called a repha or rafar. In the case of consonants which have no stem such as ट, it is appended in the shape of a ^ to the bottom of the character (bottom rakar). Maithili has the following combinations of ra:

RAFARS

Top rafar: कॅ तॅ for ex top rafars will be formed in case of following words.

दूरदर्शन, सर्वोच्च

RAKARS

1. Bottom rakar ट्र ड्र
2. Side rakar ज्र ब्र

Examples of words where Rakars are used in Maithili language are given below:

Bottom rakar	राष्ट्रपति
Side rakar	प्रश्न, ध्रुपद

6.3.1.6. Diacritics

These are as under in the case of Maithili:

ं	Anunasika	संपादकीय
ँ	Chandrabindu	सभकेँ
ः	Visarga	जन्मतः
्	Halanta	साक्षात्कार
ऌ	Avagraha	बालोऌहं, लऌगमें
’	bikari kaamaa	कत’ पड़ा’ गेल’?

6.3.1.7. Numerals

The experts have deemed that the International number set (Latino-Arabic set: 0,1,2,3,4,5,6,7,8,9) is used in official documents in Maithili.

Following are the numbers used in Maithili language written in Devanāgarī :

० १ २ ३ ४ ५ ६ ७ ८ ९

Numeral Shapes	Explanation
०	Devanāgarī Digit Zero

१	Devanāgarī Digit One
२	Devanāgarī Digit Two
३	Devanāgarī Digit Three
४	Devanāgarī Digit Four
५	Devanāgarī Digit Five
६	Devanāgarī Digit Six
७	Devanāgarī Digit Seven
८	Devanāgarī Digit Eight
९	Devanāgarī Digit Nine

6.3.1.8. Punctuation Markers

Maithili uses punctuation markers from the Latin set, such as . , ; : “ ‘ () [] etc.

Purna and *Deergha Virama* (full-stop/*danda*) Devanāgarī code block: U+0964, U+0965

।,॥ । is used to mark the full stop, ॥ is rarely used except in poetry and in religious texts.

बालोऽहं जगदानन्द ! न मे बाला सरस्वती । अपूर्णे पञ्चमे वर्षे वर्णयामि जगत्त्रयम् ॥

A list of punctuations is provided below:

Sr. No.	Name of the marker	Marker Shape
1.	Full Stop or Period Used in Maithili after numbers eg. 1.	.
2.	Question Mark	?
3.	Exclamation Mark	!
4.	Comma	,
5.	Apostrophe	'
6.	Semi Colon	;
7.	Colon	:
8.	Hyphen	-
9.	Dash	--
10.	Ellipsis mark	...
11.	Oblique	/
12.	Double quotation mark	" "
13.	Single quotation mark	' '
14.	Cross	XXX
15.	As Above	- - " - -
16.	Round Brackets	()
17.	Square Brackets	[]
18.	Curly Brackets	{ }
19.	Devanāgarī Danda	
20.	Devanāgarī Double Danda	॥

6.3.1.9 Other Symbols

These are religious, currency markers etc. included in Unicode:

ॐ : OM

₹: Rupee Sign as mandated by Government of India.¹³

6.3.2. Consonant Mātrā Combinations.

These refer to the shapes generated when a Mātrā is adjoined to the Consonant. The layout of these is in the shape of a matrix where the first horizontal row refers to the active consonant and the first vertical column refers to the vowel-modifier.

Due to constraints of space and also for reasons of clarity, for each class a series of 3 tables are provided.

Table 1: क ख ग घ ङ च छ ज झ ञ

Table 2: ट ठ ड ढ ण त थ द ध न

Table 3: प फ ब भ म य र ल व श

ष स ह

Table 4: ङ ङ

Wherever there is an X it implies that the combination does not exist. For the font developer this is an indication that for this particular combination which is not possible in the language but needs to be accommodated in the font table, a simple linear combination be provided.

e.g. Although the combination of ङ +Mātrā is theoretically not possible it needs to be handled at the font level in the anticipation that a user could type this combination. The font would show the following: ङ

The classes are as under:

6.3.2.1. refers to a simple concatenation of Consonant and Mātrā combinations.

6.3.2.2. refers to a concatenation of Consonant and Mātrā + Nasal marker combinations.

Other diacritics such as avagraha and visarga have been avoided, since these are linear in nature, are adjoined to the combination and do not in any way modify the structure of the shapes.

¹³ The earlier symbol for rupee in code block U+20A ₹ and ₹ U+0930 U+0941U+0970 is no longer in use.

6.3.2.1 Consonant and Mātrā combinations.

This set refers to a simple concatenation of Consonant and Mātrā¹⁴.

Consonant and Mātrā combinations Set 1

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ा	का	खा	गा	घा	ङा	चा	छा	जा	झा	जा
ि	कि	खि	गि	घि	ङि	चि	छि	जि	झि	जि
ी	की	खी	गी	घी	ङी	ची	छी	जी	झी	जी
ु	कु	खु	गु	घु	ङु	चु	छु	जु	झु	जु
ू	कू	खू	गू	घू	ङू	चू	छू	जू	झू	जू
ृ	कृ	खृ	गृ	घृ	ङृ	चृ	छृ	जृ	झृ	जृ
े	के	खे	गे	घे	ङे	चे	छे	जे	झे	जे
ै	कै	खै	गै	घै	ङै	चै	छै	जै	झै	जै
ो	को	खो	गो	घो	ङो	चो	छो	जो	झो	जो
ौ	कौ	खौ	गौ	घौ	ङौ	चौ	छौ	जौ	झौ	जौ
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	जँ
ॉ	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ

¹⁴ In all the combinations the Apostrophe Modifier Letter U+02BC is not listed, because it is linear in nature and is adjoined to the Character to which it relates and in no way modifies the shape of the character.

Consonant and Mātrā combinations Set 2

This set is in continuation of set 1 which shows consonant and Mātrā combinations.

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ा	टा	ठा	डा	ढा	णा	ता	था	दा	धा	ना
ि	टि	ठि	डि	ढि	णि	ति	थि	दि	धि	नि
ी	टी	ठी	डी	ढी	णी	ती	थी	दी	धी	नी
ु	टु	ठु	डु	ढु	णु	तु	थु	दु	धु	नु
ू	टू	ठू	डू	ढू	णू	तू	थू	दू	धू	नू
ृ	टृ	ठृ	डृ	ढृ	णृ	तृ	थृ	दृ	धृ	नृ
े	टे	ठे	डे	ढे	णे	ते	थे	दे	धे	ने
ै	टै	ठै	डै	ढै	णै	तै	थै	दै	धै	नै
ो	टो	ठो	डो	ढो	णो	तो	थो	दो	धो	नो
ौ	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ	धौ	नौ
ँ	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ॉ	टॉ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ

Consonant and Mātrā combinations Set 3

This set is in continuation of set 2 which shows consonant and Mātrā combinations.

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ा	पा	फा	बा	भा	मा	या	रा	ला	वा	शा	षा	सा	हा
ि	पि	फि	बि	भि	मि	यि	रि	लि	वि	शि	षि	सि	हि
ी	पी	फी	बी	भी	मी	यी	री	ली	वी	शी	षी	सी	ही
ु	पु	फु	बु	भु	मु	यु	रु	लु	वु	शु	षु	सु	हु
ू	पू	फू	बू	भू	मू	यू	रू	लू	वू	शू	षू	सू	हू
ृ	पृ	फृ	बृ	भृ	मृ	यृ	रृ	लृ	वृ	शृ	षृ	सृ	हृ
े	पे	फे	बे	भे	मे	ये	रे	ले	वे	शे	षे	से	हे
ै	पै	फै	बै	भै	मै	यै	रै	लै	वै	शै	षै	सै	है
ो	पो	फो	बो	भो	मो	यो	रो	लो	वो	शो	षो	सो	हो
ौ	पौ	फौ	बौ	भौ	मौ	यौ	रौ	लौ	वौ	शौ	षौ	सौ	हौ
ँ	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ॉ	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह

Consonant and Mātrā combinations Set 4

This set is in continuation of set 3 which shows consonant and Mātrā combinations for Nukta Consonants.

	ड़	ढ़	ज़
ा	ड़ा	ढ़ा	ज़ा
ि	ड़ि	ढ़ि	ज़ि
ी	ड़ी	ढ़ी	ज़ी
ु	ड़ु	ढ़ु	ज़ु
ू	ड़ू	ढ़ू	ज़ू
ृ	ड़ृ	ढ़ृ	ज़ू
े	ड़े	ढ़े	ज़े
ै	ड़े	ढ़े	ज़े
ो	ड़ो	ढ़ो	ज़ो
ौ	ड़ौ	ढ़ौ	ज़ौ
ँ	ड़ँ	ढ़ँ	ज़ँ
ॉ	ड़ाँ	ढ़ाँ	ज़ाँ

6.3.2.2 Consonant and Mātrā +Nasal combinations.

This set refers to a Consonant and Mātrā + Nasal marker combinations.

Consonant and Mātrā + Nasal combinations - Set 1

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ं	कं	खं	गं	घं	ङं	चं	छं	जं	झं	ञं
ां	कां	खां	गां	घां	ङां	चां	छां	जां	झां	जां
िं	किं	खिं	गिं	घिं	ङिं	चिं	छिं	जिं	झिं	ञिं
ीं	कीं	खीं	गीं	घीं	ङीं	चीं	छीं	जीं	झीं	जीं
ुं	कुं	खुं	गुं	घुं	ङुं	चुं	छुं	जुं	झुं	जुं
ूं	कूं	खूं	गूं	घूं	ङूं	चूं	छूं	जूं	झूं	जूं
ृं	कृं	खृं	गृं	घृं	ङृं	चृं	छृं	जृं	झृं	जृं
ें	कें	खें	गें	घें	ङें	चें	छें	जें	झें	जें
ैं	कैं	खैं	गैं	घैं	ङैं	चैं	छैं	जैं	झैं	जैं
ों	कों	खों	गों	घों	ङों	चों	छों	जों	झों	जों
ौं	कौं	खौं	गौं	घौं	ङौं	चौं	छौं	जौं	झौं	जौं
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	जँ
ॉं	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ

Consonant and Mātrā +Nasal combinations - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Nasal marker

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ं	टं	ठं	डं	ढं	णं	तं	थं	दं	धं	नं
ां	टां	ठां	डां	ढां	णां	तां	थां	दां	धां	नां
िं	टिं	ठिं	डिं	ढिं	णिं	तिं	थिं	दिं	धिं	निं
ीं	टीं	ठीं	डीं	ढीं	णीं	तीं	थीं	दीं	धीं	नीं
ुं	टुं	ठुं	डुं	ढुं	णुं	तुं	थुं	दुं	धुं	नुं
ूं	टूं	ठूं	डूं	ढूं	णूं	तूं	थूं	दूं	धूं	नूं
ृं	ट्रं	ठ्रं	ड्रं	ढ्रं	ण्रं	त्रं	थ्रं	द्रं	ध्रं	न्रं
ें	टे	ठे	डे	ढे	णे	ते	थे	दे	धे	ने
ैं	टै	ठै	डै	ढै	णै	तै	थै	दै	धै	नै
ों	टों	ठों	डों	ढों	णों	तों	थों	दों	धों	नों
ौं	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ	धौ	नौ
ँ	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ॉं	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ

Consonant and Mātrā +Nasal combinations - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Nasal marker

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ं	पं	फं	बं	भं	मं	यं	रं	लं	वं	शं	षं	सं	हं
ां	पां	फां	बां	भां	मां	यां	रां	लां	वां	शां	षां	सां	हां
िं	पिं	फिं	बिं	भिं	मिं	यिं	रिं	लिं	विं	शिं	षिं	सिं	हिं
ीं	पीं	फीं	बीं	भीं	मीं	यीं	रीं	लीं	वीं	शीं	षीं	सीं	हीं
ुं	पुं	फुं	बुं	भुं	मुं	युं	रुं	लुं	वुं	शुं	षुं	सुं	हुं
ूं	पूं	फूं	बूं	भूं	मूं	यूं	रूं	लूं	वूं	शूं	षूं	सूं	हूं
ृं	पृं	फृं	बृं	भृं	मृं	यृं	रृं	लृं	वृं	शृं	षृं	सृं	हृं
ें	पें	फें	बें	भें	में	यें	रें	लें	वें	शें	षें	सें	हें
ैं	पैं	फैं	बैं	भैं	मैं	यैं	रैं	लैं	वैं	शैं	षैं	सैं	हैं
ों	पों	फों	बों	भों	मों	यों	रों	लों	वों	शों	षों	सों	हों
ौं	पौं	फौं	बौं	भौं	मौं	यौं	रौं	लौं	वौं	शौं	षौं	सौं	हौं
ँ ¹⁵	पँ	फँ	बँ	भँ	मँ	यँ	रँ	लँ	वँ	शँ	षँ	सँ	हँ
ॉं	पाँ	फाँ	बाँ	भाँ	माँ	याँ	राँ	लाँ	वाँ	शाँ	षाँ	साँ	हाँ

¹⁵ Since ँ ,ँॉ are used only for the purpose of writing loan words mainly from English and since English does not have nasals, the nasal shapes of ँँ ,ँँॉ are rarely encountered in Maithili

Consonant and Mātrā +Nasal combinations - Set 4

This set is in continuation of set 3 above which shows combinations of Nukta Consonant and Mātrā + Nasal marker

	ड़	ढ़	ज़
ं	ड़ं	ढ़ं	ज़ं
ां	ड़ां	ढ़ां	ज़ां
िं	ड़िं	ढ़िं	ज़िं
ीं	ड़ीं	ढ़ीं	ज़ीं
ुं	ड़ुं	ढ़ुं	ज़ुं
ूं	ड़ूं	ढ़ूं	ज़ूं
ृं	ड़ृं	ढ़ृं	ज़ृं
ें	ड़ें	ढ़ें	ज़ें
ैं	ड़ैं	ढ़ैं	ज़ैं
ों	ड़ों	ढ़ों	ज़ों
ौं	ड़ौं	ढ़ौं	ज़ौं
ँ ¹⁶	ड़ँ	ढ़ँ	ज़ँ
ॉं	ड़ाँ	ढ़ाँ	ज़ाँ

¹⁶ Since ँ ,ँ are used only for the purpose of writing loan words mainly from English and since English does not have nasals, the nasal shapes of ँँ ,ँँ are rarely encountered in Maithili

Consonant and Mātrā + Nasal combinations: With Chandrabindu - Set 1¹⁷

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	ञँ
ाँ	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ
िँ	किँ	खिँ	गिँ	घिँ	ङिँ	चिँ	छिँ	जिँ	झिँ	ञिँ
ीँ	कीँ	खीँ	गीँ	घीँ	ङीँ	चीँ	छीँ	जीँ	झीँ	जीँ
ुँ	कुँ	खुँ	गुँ	घुँ	ङुँ	चुँ	छुँ	जुँ	झुँ	जुँ
ूँ	कूँ	खूँ	गूँ	घूँ	ङूँ	चूँ	छूँ	जूँ	झूँ	जूँ
ृँ	कृँ	खृँ	गृँ	घृँ	ङृँ	चृँ	छृँ	जृँ	झृँ	जृँ
ेँ	केँ	खेँ	गेँ	घेँ	ङेँ	चेँ	छेँ	जेँ	झेँ	जेँ
ैँ	कैँ	खैँ	गैँ	घैँ	ङैँ	चैँ	छैँ	जैँ	झैँ	जैँ
ोँ	कोँ	खोँ	गोँ	घोँ	ङोँ	चोँ	छोँ	जोँ	झोँ	जोँ
ौँ	कौँ	खौँ	गौँ	घौँ	ङौँ	चौँ	छौँ	जौँ	झौँ	जौँ

¹⁷ Since ँ, ौ are used only for the purpose of writing loan words mainly from English and since English does not have nasals, the nasal shapes of ँ, ौ are rarely encountered in Maithili and are therefore not included.

Consonant and Mātrā +Nasal combinations With Chandrabindu - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Chandrabindu

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ँ	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ा	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ
ि	टिँ	ठिँ	डिँ	ढिँ	णिँ	तिँ	थिँ	दिँ	धिँ	निँ
ी	टीँ	ठीँ	डीँ	ढीँ	णीँ	तीँ	थीँ	दीँ	धीँ	नीँ
ु	टुँ	ठुँ	डुँ	ढुँ	णुँ	तुँ	थुँ	दुँ	धुँ	नुँ
ू	टूँ	ठूँ	डूँ	ढूँ	णूँ	तूँ	थूँ	दूँ	धूँ	नूँ
े	टेँ	ठेँ	डेँ	ढेँ	णेँ	तेँ	थेँ	देँ	धेँ	नेँ
ै	टैँ	ठैँ	डैँ	ढैँ	णैँ	तैँ	थैँ	दैँ	धैँ	नैँ
ो	टौँ	ठौँ	डौँ	ढौँ	णौँ	तौँ	थौँ	दौँ	धौँ	नौँ
ौ	टौँ	ठौँ	डौँ	ढौँ	णौँ	तौँ	थौँ	दौँ	धौँ	नौँ

Consonant and Mātrā +Nasal combinations With Chandrabindu - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Chandrabindu

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ँ	पँ	फँ	बँ	भँ	मँ	यँ	रँ	लँ	वँ	शँ	षँ	सँ	हँ
ा	पाँ	फाँ	बाँ	भाँ	माँ	याँ	राँ	लाँ	वाँ	शाँ	षाँ	साँ	हाँ
ि	पिँ	फिँ	बिँ	भिँ	मिँ	यिँ	रिँ	लिँ	विँ	शिँ	षिँ	सिँ	हिँ
ी	पीँ	फीँ	बीँ	भीँ	मीँ	यीँ	रीँ	लीँ	वीँ	शीँ	षीँ	सीँ	हीँ
ु	पुँ	फुँ	बुँ	भुँ	मुँ	युँ	रुँ	लुँ	वुँ	शुँ	षुँ	सुँ	हुँ
ू	पूँ	फूँ	बूँ	भूँ	मूँ	यूँ	रूँ	लूँ	वूँ	शूँ	षूँ	सूँ	हूँ
ृ	पृँ	फृँ	बृँ	भृँ	मृँ	यृँ	रृँ	लृँ	वृँ	शृँ	षृँ	सृँ	हृँ
े	पेँ	फेँ	बेँ	भेँ	मेँ	येँ	रेँ	लेँ	वेँ	शेँ	षेँ	सेँ	हेँ
ै	पैँ	फैँ	बैँ	भैँ	मैँ	यैँ	रैँ	लैँ	वैँ	शैँ	षैँ	सैँ	हैँ
ो	पोँ	फोँ	बोँ	भोँ	मोँ	योँ	रोँ	लोँ	वोँ	शोँ	षोँ	सोँ	होँ
ौ	पौँ	फौँ	बौँ	भौँ	मौँ	यौँ	रौँ	लौँ	वौँ	शौँ	षौँ	सौँ	हौँ

Consonant and Mātrā +Nasal combinations With Chandrabindu - Set 4

This set is in continuation of set 3 above which shows combinations of Nukta Consonant and Mātrā + Chandrabindu

	ॐ	ॐ	ॐ
ँ	ॐ	ॐ	ॐ
ाँ	ॐ	ॐ	ॐ
िँ	ॐ	ॐ	ॐ
ीँ	ॐ	ॐ	ॐ
ुँ	ॐ	ॐ	ॐ
ूँ	ॐ	ॐ	ॐ
ृँ	ॐ	ॐ	ॐ
ेँ	ॐ	ॐ	ॐ
ैँ	ॐ	ॐ	ॐ
ोँ	ॐ	ॐ	ॐ
ौँ	ॐ	ॐ	ॐ
ँ	ॐ	ॐ	ॐ
ॉ	ॐ	ॐ	ॐ

6.3.3. The Ligature Set of Maithili.

Maithili has a large set of ligatural forms. These are combinations of Consonant+Halanta+Consonant (CHC) or CHCHC or even rarer CHCHCHC. The CHC combinations which are the most frequent are arranged in the shape of a matrix: the abscissa or horizontal axis refers to the Consonant which constitutes the ligature and the ordinate or vertical axis shows the consonant which forms the ligature and which is followed by a halanta.

As in 6.3.2. the ligature sets are divided into the following

6.3.3.1 CHC (in a matrix)

6.3.3.2 CHCHC

6.3.3.1. CHC (combination of two consonanats)

These ligatures are presented as in the earlier case of Consonant+Mātrā combinations in three sets. A lot of “slots” have an X marked, showing that the ligature is not possible in the language but is theoretically possible. In these cases, the font developer is to assume that the ligature is linear in nature¹⁸.

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “क” and “ख” is ligature “क्ख”.

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
क्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ख्	ख्क	ख्ख	ख्ग	ख्घ	ख्ङ	ख्च	ख्छ	ख्ज	ख्झ	ख्ञ
ग्	ग्क	ग्ख	ग्ग	ग्घ	ग्ङ	ग्च	ग्छ	ग्ज	ग्झ	ग्ञ
घ्	घ्क	घ्ख	घ्ग	घ्घ	घ्ङ	घ्च	घ्छ	घ्ज	घ्झ	घ्ञ
ङ्	ङ्क	ङ्ख	ङ्ग	ङ्घ	ङ्ङ	ङ्च	ङ्छ	ङ्ज	ङ्झ	ङ्ञ
च्	च्क	च्ख	च्ग	च्घ	च्ङ	च्च	च्छ	च्च	च्च	च्च
छ्	छ्क	छ्ख	छ्ग	छ्घ	छ्ङ	छ्च	छ्छ	छ्ज	छ्झ	छ्ञ
ज्	ज्क	ज्ख	ज्ग	ज्घ	ज्ङ	ज्च	ज्छ	ज्ज	ज्झ	ज्ञ
झ्	झ्क	झ्ख	झ्ग	झ्घ	झ्ङ	झ्च	झ्छ	झ्ज	झ्झ	झ्ञ
ञ्	ञ्क	ञ्ख	ञ्ग	ञ्घ	ञ्ङ	ञ्च	ञ्छ	ञ्ज	ञ्झ	ञ्ञ
ट्	ट्क	ट्ख	ट्ग	ट्घ	ट्ङ	ट्च	ट्छ	ट्ज	ट्झ	ट्ञ
ट्	ट्क	ट्ख	ट्ग	ट्घ	ट्ङ	ट्च	ट्छ	ट्ज	ट्झ	ट्ञ
ड्	ड्क	ड्ख	ड्ग	ड्घ	ड्ङ	ड्च	ड्छ	ड्ज	ड्झ	ड्ञ
ढ्	ढ्क	ढ्ख	ढ्ग	ढ्घ	ढ्ङ	ढ्च	ढ्छ	ढ्ज	ढ्झ	ढ्ञ
ण्	ण्क	ण्ख	ण्ग	ण्घ	ण्ङ	ण्च	ण्छ	ण्ज	ण्झ	ण्ञ

¹⁸ Alternate forms as proposed by the Experts are shown as preceded by a /

त्	त्क	त्ख	त्ग	त्घ	त्ङ	त्च	त्छ	त्ज	त्झ	त्ञ
थ्	थ्क	थ्ख	था	थ्घ	थ्ङ	थ्च	थ्छ	थ्ज	थ्झ	थ्ञ
द्	द्क	द्ख	द्ग	द्घ	द्ङ	द्च	द्छ	द्ज	द्झ	द्ञ
ध्	ध्क	ध्ख	धा	ध्घ	ध्ङ	ध्च	ध्छ	ध्ज	ध्झ	ध्ञ
न्	न्क	न्ख	ना	न्घ	न्ङ	न्च	न्छ	न्ज	न्झ	न्ञ
प्	प्क	प्ख	पा	प्घ	प्ङ	प्च	प्छ	प्ज	प्झ	प्ञ
फ्	फ्क	फ्ख	फा	फ्घ	फ्ङ	फ्च	फ्छ	फ्ज	फ्झ	फ्ञ
ब्	ब्क	ब्ख	बा	ब्घ	ब्ङ	ब्च	ब्छ	ब्ज	ब्झ	ब्ञ
भ्	भ्क	भ्ख	भा	भ्घ	भ्ङ	भ्च	भ्छ	भ्ज	भ्झ	भ्ञ
म्	म्क	म्ख	मा	म्घ	म्ङ	म्च	म्छ	म्ज	म्झ	म्ञ
य्	य्क	य्ख	या	य्घ	य्ङ	य्च	य्छ	य्ज	य्झ	य्ञ
र्	र्क	र्ख	र्ग	र्घ	र्ङ	र्च	र्छ	र्ज	र्झ	र्ञ
ल्	ल्क	ल्ख	ला	ल्घ	ल्ङ	ल्च	ल्छ	ल्ज	ल्झ	ल्ञ
व्	व्क	व्ख	वा	व्घ	व्ङ	व्च	व्छ	व्ज	व्झ	व्ञ
श्	श्क	श्ख	शा	श्घ	श्ङ	श्च	श्छ	श्ज	श्झ	श्ञ
ष्	ष्क	ष्ख	षा	ष्घ	ष्ङ	ष्च	ष्छ	ष्ज	ष्झ	ष्ञ
स्	स्क	स्ख	सा	स्घ	स्ङ	स्च	स्छ	स्ज	स्झ	स्ञ
ह्	ह्क	ह्ख	हा	ह्घ	ह्ङ	ह्च	ह्छ	ह्ज	ह्झ	ह्ञ
ज्ञ्	ज्ञ्क	ज्ञ्ख	ज्ञा	ज्ञ्घ	ज्ञ्ङ	ज्ञ्च	ज्ञ्छ	ज्ञ्ज	ज्ञ्झ	ज्ञ्ञ
फ़्	फ़्क	फ़्ख	फ़ा	फ़्घ	फ़्ङ	फ़्च	फ़्छ	फ़्ज	फ़्झ	फ़्ञ
ड्	ड्क	ड्ख	डा	ड्घ	ड्ङ	ड्च	ड्छ	ड्ज	ड्झ	ड्ञ
ढ्	ढ्क	ढ्ख	ढा	ढ्घ	ढ्ङ	ढ्च	ढ्छ	ढ्ज	ढ्झ	ढ्ञ

CHC Set 2:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “क” and “ट” is ligature “कट”.

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
क्	कट	कठ	कड	कढ	कण	क्त	कथ	कद	कध	क्र
ख्	खट	खठ	खड	खढ	खण	ख्त	खथ	खद	खध	ख्न
ग्	गट	गठ	गड	गढ	गण	गत	गथ	गद	गध	ग्न
घ्	घट	घठ	घड	घढ	घण	घ्त	घथ	घद	घध	घ्न
ङ्	ङट	ङठ	ङड	ङढ	ङण	ङ्त	ङथ	ङद	ङध	ङ्न
च्	चट	चठ	चड	चढ	चण	च्त	चथ	चद	चध	च्न
छ्	छट	छठ	छड	छढ	छण	छ्त	छथ	छद	छध	छ्न
ज्	जट	जठ	जड	जढ	जण	ज्त	जथ	जद	जध	ज्न
झ्	झट	झठ	झड	झढ	झण	झ्त	झथ	झद	झध	झ्न
ञ्	ञट	ञठ	ञड	ञढ	ञण	ञ्त	ञथ	ञद	ञध	ञ्न
ट्	टट	टठ	टड	टढ	टण	ट्त	टथ	टद	टध	ट्न
ठ्	ठट	ठठ	ठड	ठढ	ठण	ठ्त	ठथ	ठद	ठध	ठ्न
ड्	डट	डठ	डड	डढ	डण	ड्त	डथ	डद	डध	ड्न
ढ्	ढट	ढठ	ढड	ढढ	ढण	ढ्त	ढथ	ढद	ढध	ढ्न
ण्	णट	णठ	णड	णढ	णण	ण्त	णथ	णद	णध	ण्न
त्	त्ट	त्ठ	त्ड	त्ढ	त्ण	त्त	त्थ	त्द	त्ध	त्न
थ्	थट	थठ	थड	थढ	थण	थ्त	थथ	थद	थध	थ्न
द्	दट	दठ	दड	दढ	दण	द्त	दथ	दद	दध	द्न
ध्	धट	धठ	धड	धढ	धण	ध्त	धथ	धद	धध	ध्न
न्	नट	नठ	नड	नढ	नण	न्त	न्थ	न्द	न्ध	न्न
प्	पट	पठ	पड	पढ	पण	प्त	पथ	पद	पध	प्न
फ्	फट	फठ	फड	फढ	फण	फ्त	फथ	फद	फध	फ्न
ब्	बट	बठ	बड	बढ	बण	ब्त	बथ	बद	बध	ब्न
भ्	भट	भठ	भड	भढ	भण	भ्त	भथ	भद	भध	भ्न
म्	मट	मठ	मड	मढ	मण	म्त	मथ	मद	मध	म्न
य्	यट	यठ	यड	यढ	यण	य्त	यथ	यद	यध	य्न
र्	रट	रठ	रड	रढ	रण	र्त	रथ	रद	रध	र्र्न

लृ	लट	लठ	लड	लढ	लण	लत	लथ	लद	लध	लन
वृ	वट	वठ	वड	वढ	वण	वत	वथ	वद	वध	वन
शृ	शट	शठ	शड	शढ	शण	शत	शथ	शद	शध	शन
षृ	षट	षठ	षड	षढ	षण	षत	षथ	षद	षध	षन
सृ	सट	सठ	सड	सढ	सण	सत	सथ	सद	सध	सन
हृ	हट	हठ	हड	हढ	हण	हत	हथ	हद	हध	हन
जृ	जट	जठ	जड	जढ	जण	जत	जथ	जद	जध	जन
झृ	झट	झठ	झड	झढ	झण	झत	झथ	झद	झध	झन
ढृ	ढट	ढठ	ढड	ढढ	ढण	ढत	ढथ	ढद	ढध	ढन

CHC SET 3:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “क” and “प” is the ligature “क्प”.

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
क्	क्प	क्फ	क्ब	क्भ	क्म	क्य	क्र	क्ल	क्व	क्श	क्ष	क्स	क्ह
ख्	खप	खफ	खब	खभ	खम	ख्य	ख्र	खल	खव	खश	खष	खस	खह
ग्	गप	गफ	गब	गभ	गम	ग्य	ग्र	गल	गव	गश	गष	गस	गह
घ्	घप	घफ	घब	घभ	घम	घ्य	घ्र	घल	घव	घश	घष	घस	घह
ङ्	ङप	ङफ	ङब	ङभ	ङम	ङ्य	ङ्र	ङल	ङव	ङश	ङष	ङस	ङह
च्	चप	चफ	चब	चभ	चम	च्य	च्र	चल	चव	चश	चष	चस	चह
छ्	छप	छफ	छब	छभ	छम	छ्य	छ्र	छल	छव	छश	छष	छस	छह
ज्	जप	जफ	जब	जभ	जम	ज्य	ज्र	जल	जव	जश	जष	जस	जह
झ्	झप	झफ	झब	झभ	झम	झ्य	झ्र	झल	झव	झश	झष	झस	झह
ञ्	ञप	ञफ	ञब	ञभ	ञम	ञ्य	ञ्र	ञल	ञव	ञश	ञष	ञस	ञह
ट्	टप	टफ	टब	टभ	टम	ट्य	ट्र	टल	टव	टश	टष	टस	टह
ठ्	ठप	ठफ	ठब	ठभ	ठम	ठ्य	ठ्र	ठल	ठव	ठश	ठष	ठस	ठह
ड्	डप	डफ	डब	डभ	डम	ड्य	ड्र	डल	डव	डश	डष	डस	डह
ढ्	ढप	ढफ	ढब	ढभ	ढम	ढ्य	ढ्र	ढल	ढव	ढश	ढष	ढस	ढह
ण्	णप	णफ	णब	णभ	णम	ण्य	ण्र	णल	णव	णश	णष	णस	णह
त्	तप	तफ	तब	तभ	तम	त्य	त्र	तल	तव	तश	तष	तस	तह
थ्	थप	थफ	थब	थभ	थम	थ्य	थ्र	थल	थव	थश	थष	थस	थह
द्	दप	दफ	दब	दभ	दम	द्य	द्र	दल	दव	दश	दष	दस	दह
ध्	धप	धफ	धब	धभ	धम	ध्य	ध्र	धल	धव	धश	धष	धस	धह
न्	नप	नफ	नब	नभ	नम	न्य	न्र	नल	नव	नश	नष	नस	नह
प्	पप	पफ	पब	पभ	पम	प्य	प्र	पल	पव	पश	पष	पस	पह
फ्	फप	फफ	फब	फभ	फम	फ्य	फ्र	फल	फव	फश	फष	फस	फह
ब्	बप	बफ	बब	बभ	बम	ब्य	ब्र	बल	बव	बश	बष	बस	बह
भ्	भप	भफ	भब	भभ	भम	भ्य	भ्र	भल	भव	भश	भष	भस	भह
म्	मप	मफ	मब	मभ	मम	म्य	म्र	मल	मव	मश	मष	मस	मह
य्	यप	यफ	यब	यभ	यम	य्य	य्र	यल	यव	यश	यष	यस	यह

र्	र्ष	र्फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ल्	ल्प	ल्फ	ल्ब	ल्भ	ल्म	ल्य	ल	ल्ल	ल्व	ल्श	ल्ष	ल्स	ल्ह
व्	व्य	व्फ	व्व	व्वभ	व्वम	व्य	व्र	व्ल	व्व	व्वश	व्वष	व्वस	व्वह
श्	श्प	श्फ	श्ब	श्भ	श्म	श्य	श्र	शल	श्व	शश	शष	शस	शह
ष्	ष्प	ष्फ	ष्व	ष्वभ	ष्वम	ष्य	ष्	ष्ल	ष्व	ष्वश	ष्वष	ष्वस	ष्वह
स्	स्प	स्फ	स्ब	स्भ	स्म	स्य	स्र	सल	स्व	स्श	स्ष	स्स	सह
ह्	ह्य	ह्यफ	ह्व	ह्वभ	ह्वम	ह्य	ह्र	ह्ल	ह्व	ह्वश	ह्वष	ह्वस	ह्वह
ज़्	ज़्प	ज़्फ	ज़्ब	ज़्भ	ज़्म	ज़्य	ज़्र	ज़्ल	ज़्व	ज़्श	ज़्ष	ज़्स	ज़्ह
ड्	ड्प	ड्फ	ड्ब	ड्भ	ड्म	ड्य	ड्र	ड्ल	ड्व	ड्श	ड्ष	ड्स	ड्ह

CHC SET 4:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “क” and “ड़” is ligature “कड़”.

	ड़	ढ़	ज़
क्	कड़	कढ़	कज़
ख्	खड़	खढ़	खज़
ग्	गड़	गढ़	गज़
घ्	घड़	घढ़	घज़
ङ्	ङड़	ङढ़	ङज़
च्	चड़	चढ़	चज़
छ्	छड़	छढ़	छज़
ज्	जड़	जढ़	जज़
झ्	झड़	झढ़	झज़
ञ्	ञड़	ञढ़	ञज़
ट्	टड़	टढ़	टज़
ठ्	ठड़	ठढ़	ठज़
ड्	डड़	डढ़	डज़
ढ्	ढड़	ढढ़	ढज़
ण्	णड़	णढ़	णज़
त्	तड़	तढ़	तज़
थ्	थड़	थढ़	थज़
द्	दड़	दढ़	दज़
ध्	धड़	धढ़	धज़
न्	नड़	नढ़	नज़
प्	पड़	पढ़	पज़
फ्	फड़	फढ़	फज़
ब्	बड़	बढ़	बज़

भ्	भृ	भू	भृज्
म्	मृ	मू	मृज्
य्	यृ	यू	यृज्
र्	रृ	रू	रृज्
ल्	लृ	लू	लृज्
व्	वृ	वू	वृज्
श्	शृ	शू	शृज्
ष्	षृ	षू	षृज्
स्	सृ	सू	सृज्
ह्	हृ	हू	हृज्
ड्	डृ	डू	डृज्
ढ्	ढृ	ढू	ढृज्
ज्ञ्	ज्ञृ	ज्ञू	ज्ञृज्

6.3.3.2 CHCHC (combination of three consonants)

These are not as frequent as the CHC combinations. Only the major are listed below.

With a few exceptions these are mainly linear in nature. Maithili admits the following clusters:

हृत्य=हृत्या

स्ट=वस्ट

स्थ=स्वास्थ्य

स्र=शास्त्रीय

स्ट्र=डिस्ट्रीब्यूशन

ष्ट्र=राष्ट्रीय

ष्क्र=निष्क्रिय

श्रव=पाश्च

ल्ट=वल्ड

तृत्य=अमृत्य

र्त=संग्रहकर्ताक

षृत्य=इष्या

म्पु=काम्प्लेक्स

म्यु=कम्प्यूटर

न्फ्र=कान्फ्रेंस

न्द्र=उपेन्द्र

त्स्य=मत्स्य

टर्स=स्पोर्ट्समैन

ज्ज्व=उज्ज्वल

क्ष्य=लक्ष्य

क्ष्म=लक्ष्मीश्वर

क्त्र=इलेक्ट्रिक

6.3.4 The Collation Order of Maithili.

Collation is one of the most important features of a script grammar. It determines the order in which a given culture indexes its characters. This is best seen in a dictionary sort where for easy search words are sorted and arranged in a specific order. Within a given script, each allo-script may have a different sort-order. Thus in Devanāgarī the conjunct glyph क्ष is sorted along with क, since the first letter of that conjunct is क and on a similar principle ज्ञ is sorted along with ज. Different scripts admit different sort orders and for all high-end NLP applications, sort is a crucial feature to ensure that the applications index data as per the cultural perception of that community. In quite a few States, sort order is clearly defined by the statutory bodies of that state and hence it is crucial that such sort order be ascertained and introduced in the script grammar.

In the case of Maithili the following is the traditional sort order as determined by the experts. The order as given below is pertinent to sorting by a computer program and is compliant with CLDR as laid down by Unicode and W3C.

ँ ँ ॊ अ आ इ ई उ ऊ ऋ ए ऐ
 ओ ओ औ क ख ग घ ङ च छ ज झ
 ञ ट ठ ड ढ ढ ढ ण त थ द ध
 न प फ ब भ म य र ल व श ष
 स ह क्ष ज्ञ ा ि ी ु ू ृ े ै
 ो ौ ¹⁹

In Tabular format:

ँ	ं	ॊ	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ
ओ	ओ	औ	क	ख	ग	घ	ङ	च	छ	ज	झ
ञ	ट	ठ	ड	ड़	ढ	ढ़	ण	त	थ	द	ध
न	प	फ	ब	भ	म	य	र	ल	व	श	ष
स	ह	क्ष	ज्ञ	ा	ि	ी	ु	ू	ृ	े	ै
ो	ौ										

¹⁹ The tonal marker 02BC is placed at the end of the collation order since it is a tone marker and is appended to a given character.

7. REFERENCES

1. <http://www.unicode.org>
2. ISCII'91

8. ANNEXURES

Annexure 1: Names of experts who have contributed to the script grammar

Annexure 2: Unicode Table of Devanāgarī ²⁰

Link: <http://unicode.org/charts/PDF/U0900.pdf>

²⁰ The Unicode chart provided is for version 5.1 since the Script Grammar was prepared at that time. No considerable change in the script grammar can be seen in the updated versions of Unicode, with the possible addition of the Rupee Sign U+02B9

