

SCRIPT GRAMMAR FOR MANIPURI 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	133
Node.....	17
Mātrā/Kār	17
Modifier	17
Mātrā/Kār+Modifier	17
6.3 SCRIPT GRAMMAR PROPER.....	18
6.3.1. The Character Set of Manipuri.	18
6.3.2. Consonant Mātrā/Kār Combinations.	24
6.3.3. The Ligature Set of Manipuri.	35
6.3.4 The Collation Order of Manipuri.....	49
7. REFERENCES	51
8. ANNEXURES.....	52
Annexure 1: Names of experts who have contributed to the script grammar.....	52
Annexure 2: Unicode Table of Bangla.....	53

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 Nepali 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 Nepali /la/ is different from the Marathi /la/ in so far as the placement of the stem is concerned Nepali 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 Nepali /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 ISCII'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 catenators i.e. characters such as Mātrā/kārs 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 ण 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 ae 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 available in the current font. In the absence of a conjunct glyph, the one or more dead

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

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 /ɪ̄/ 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 **eyelash ra**)

eyelash ra: The eyelash ra is used in Konkani, Nepali and Marathi. It is treated as different from the ॠ (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: MANIPURI

ISO Mnemonics: *mni*

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 Manipuri is written using the Manipuri 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 ꯀ or alternatively form an entirely new shape ꯁ
+ $\text{ꯂ} = \text{ꯃ}$.

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 Manipuri language

6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

Manipuri language complies with the syllable (akshar) structure described above. It can admit up to 4 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 Manipuri.

(C) Consonants

ক	খ	গ	ঘ	ঙ
চ	ছ	জ	ঝ	ঞ
ট	ঠ	ড	ঢ	ণ
ত	থ	দ	ধ	ন
প	ফ	ব	ভ	ম
য	র	ল	ৱ	শ
ষ	স	হ	ক্ষ	ড়
ড়	য়	ং	ং	ং
ৱ				

(V) Vowels

অ	আ	ই	ঈ	উ	ঊ	ঋ	ৠ	এ	ঐ	ও	ঔ
---	---	---	---	---	---	---	---	---	---	---	---

(M) Mātrās/Kārs 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.
◌ँ: 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.

(H): Halanta/Hasanta ◌ँ – Halanta/Hasanta used in most writing systems to signify the lack of an inherent vowel.

(N)⁵ Nukta ◌ँ - is used in Manipuri in 𑜏, 𑜐 and in 𑜑 for Bangla

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	Subtype	Followed	
H,M	C	C,M,D	
–	V	C,D	
C	M	C,D	
V,C	D		
M	H	C	M

⁵ The nukta is a small dot placed under a character in certain scripts to show that they are flapped or for deriving 5 other consonants for Urdu loans क,ख,ग,ज,फ

Examples:

1. A vowel syllable is of the following types

1.1. A pure vowel all by itself: ই (i) , উ (u), এ (e), ও (o), ঐ (ei), ঔ (ou),

1.2. A vowel followed by a consonant: ইন (in) (v+c), উন (un) (v+c), উৎ (ut) (v +c)

1.3. A vowel followed by a modifier i.e. া : অং (ang) (V + D), উং (ung) (V+D)

2. A consonant syllable can be of the following types.

2.1. A full consonant without matra or modifier: ক (ka), ল (la)

2.2. A consonant preceded by a half consonant i.e. (consonant +halanta)

ক্র (kra) (H +C)

2.3. A consonant preceded by a consonant cluster (consonant + halanta) and followed by a consonant: ক্রক (krak) (H + C +C), ক্রল (trat) (H + C +C)

2.4. A consonant preceded by a consonant halanta and followed by a matra: কা (kwa) (H + C + M), শ্বা (shwa) (H + C + M)

2.5. A consonant preceded by a matra: পে /pe/(M + C), চে (che) (M + C)

2.6. A consonant preceded by a matra and followed by a consonant: চেক /chek) (M + C
হিক (hik) (M + C + C)

2.7. A consonant preceded and followed by matras: মো (mo) (M + C + M), লৌ(lou) (M + C + M)

2.8. A consonant followed by another consonant: হক (hak) (C + C),

থক (thak) (C + C)

2.9. A consonant followed by a matra: চা (chā) (C + M), লা(lā) (C + M), তা(tā) (C + M)

2.10. A consonant followed by a modifier i.e. কং (kang) (C + D), লং (lang) (C +D)

3. Matra

3.1. A matra preceded by a consonant: পী (pa) (C +M)

3.2. A matra preceded and followed by consonants: ঙাক (ngāk) (C + M + C), চাক(chāk)
(C + M + C)

3.3. A matra preceded by a consonant and followed by a modifier i.e. ফাঁ (pha) (C + M +
D), বাঁ (ba) (C + M + D)

4. Diacritics

4.1. A diacritic preceded by a vowel: অং (ang) (V + D)

4.2. A diacritic preceded by a consonant: কং (kang) (C + D), সং (sang) (C + D)

5. Halanta

5.1. A consonant cluster (consonant halanta + consonant) preceded by a matra and followed by a consonant: প্রে (pre) (M + H + C) , চ্রে (chre) (M + H + C)

5.2. A consonant cluster (consonant halanta) preceded by a matra and followed by a consonant and a matra and followed by a consonant and a matra again: প্রো (pro) (M + H + C + M) খ্রো (khro) (M + H + C + M) [It is a new addition in the template above, which has four boxes]

Node	Mātrā	Modifier	Mātrā/Kār+Modifier
C ⁶	CM, MC, MCM	CD	CMD
CHC	CHCM	CHCD	CHCMD, MCHCMD
CHCHC	CHCHCM	-	-

Addition

MC ঢে (ce), পে (pe)

MCM মো (mo), সো (so)

MCHCMD ক্রোং (krong)

8.1. BASIC SET OF CHARACTERS

The basic set of characters has been provided in this inventory.

These are arranged as per their class:

CONSONANT/VOWEL/MATRAS/DIACRITICS

The allographs are presented at the end.

⁶ “C” here will automatically be treated as being also consonant+nukta, C+N to simplify the explanation

6.3 SCRIPT GRAMMAR PROPER

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

- 6.3.1. The Character Set of Manipuri.
- 6.3.2. The Consonant mātrā/kār combinations of Manipuri.
- 6.3.3. The Ligature Set of Manipuri.
- 6.3.4. Collation Order of Manipuri

6.3.1. The Character Set of Manipuri.

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 Manipuri.

This comprises the following:

- 6.3.1.1. The Consonant Set
- 6.3.1.2. The Vowel Set
- 6.3.1.3. The Mātrā/Kār 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 Manipuri comprises the following characters:

A basic Consonant inventory arranged as per their Vargas.

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	ক	খ	গ	ঘ	ঙ
Palatal	চ	ছ	জ	ঝ	ঞ
Retroflex	ট	ঠ	ড	ঢ	ণ
Dental	ত	থ	দ	ধ	ন
B-labial	প	ফ	ব	ভ	ম

Flaps

ড়	ঢ়
----	----

Other consonants

য	র	ল	শ	ষ
স	হ	য়	ৱ	

Special consonant ⁷

৳ ⁸

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

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	ক	খ	গ	ঘ	ঙ
Palatal	চ	ছ	জ	ঝ	ঞ
Retroflex	ট	ঠ	ড	ঢ	ণ
Dental	ত	থ	দ	ধ	ন
B-labial	প	ফ	ব	ভ	ম

Flaps

ড়	ঢ়
----	----

Other consonants

য	র	ল	শ	ষ
স	হ	য়	ৱ	

Special consonant

৳

⁷ This categorization has been proposed by the experts of the Manipuri Sahitya Parishad

⁸ the *khanda ta* is a special consonant in Manipuri since unlike other consonants it cannot be followed by a Mātrā/Kār, Hasanta or a Diacritic but it can form a typical conjunct like ঳ (ৱ + ঳)

6.3.1.2. The Vowel Set

The Vowel set of Manipuri is as under:

অ	MANIPURI LETTER A
আ	MANIPURI LETTER AA
ই	MANIPURI LETTER I
ঈ	MANIPURI LETTER II
উ	MANIPURI LETTER U
ঊ	MANIPURI LETTER UU
ঋ	MANIPURI LETTER VOCALIC R
৺	MANIPURI LETTER L
এ	MANIPURI LETTER E
ঐ	MANIPURI LETTER AI
ও	MANIPURI LETTER O
ঔ	MANIPURI LETTER AU

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

অ আ ই ঈ উ ঊ ঋ ৺ এ ঐ ও ঔ

6.3.1.3. The Mātrā/Kār Set

The Mātrā/Kār (Vowel Modifier Set) of Manipuri is as under:

Mātrā/Kār Names	Mātrā/kārs Sign	Where is it used ?	Consonant Shapes formed
1. Manipuri sign AA	া	আ	ক্ + আ = কা
2. Manipuri sign I (stands to the left of the consonant)	ি	ই	ক্ + ই = কি

3. Manipuri sign II	ী	ঈ	ক্ + ঈ = কী
4. Manipuri sign U	ু	উ	ক্ + উ = কু
5. Manipuri sign UU	ূ	ঊ	ক্ + ঊ = কূ
6. Manipuri sign vocalic R	্	ঋ	ক্ + ঋ = ক্
7. Manipuri sign E	ে	এ	ক্ + এ = কে
8. Manipuri sign AI	ৈ	ঐ	ক্ + ঐ = কৈ
9..Manipuri sign O	ো	ও	ক্ + ও = কো
10. Manipuri sign AU	ৌ	ঔ	ক্ + ঔ = কৌ

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

া ি ী ু ূ ্ ে ৈ ো ৌ

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. Manipuri admits the following displaced catenators.

CATENATOR	POSITION	EXAMPLE
ি	To left of Consonant	কি
ে	To left of Consonant	কে
ৈ	To left of Consonant	কৈ

TWO PART DEPENDENT VOWEL SIGNS

ো	To right and left of the consonant	কো
ৌ	To right and left of the consonant	কৌ

6.3.1.5. Shape of the combination of ra (rakar/ra phalā, rafar/repha/reph)

The ৱ takes a variety of shapes known as **rakar (ra phalā in Manipuri) and rafar/repha (reph in Manipuri)** depending on its position. When conjoined before a consonant by means of the halanta/hasanta, it changes shape and is placed on top of the consonant or consonant clusters to which it relates. This is called a repha/reph or rafar. When it is

conjoined after a consonant with the help of a halanta/hasanta, it appends to the consonant in the shape of a slanting stroke attached to the stem (side rakar) or in the case of consonants which have no stem such as ɽ, it is appended in the shape of a ^ to the bottom of the character (bottom rakar/ra phalā). Manipuri has the following combinations of ra:

RAFAR/ REPH

ৰ্ক ৰ্ম ৰ্ন

for eg. reph will be formed in case of following words.

টুকীশিং, মুক্তি

In addition to the reph being adjoined to the Consonant, Bangla like Sanskrit admits a special case of the reph being adjoined to the Vocalic RA as shown below. This is only in the case of a tatsama word (from Sanskrit): লৈৰ্ৰত

The reph can also be adjoined as mentioned above to the *khanda ta* as in ভঁসৰ্ৰা

RA PHALĀ (RAKARS)

There exists only ONE ra phalā (rakar) in Manipuri which is appended to the bottom of the consonant. Examples of words usin ra phalā in Manipuri language are given below:

দ্ৰ, প্র, ব্ৰ

Other Allographs

ক্ জ্ ব্ খ্ ত্ প্ দ্ ক্ ম্ ফ্ স্ শ্

Example: ক্ৰ্ (kri-kri), ব্ৰ্ (bri-bri), জ্ৰ্ (jri-jri)

6.3.1.6. Diacritics

These are as under in the case of Manipuri:

◌ং - Anuswar মঙংগি

◌ঁ - Chandrabindu/Anunasika

◌্ - Halanta/Hasanta

◌ঃ - Visarga

6.3.1.7. Numerals

Following are the numbers used in Manipuri language (၀ ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ ၀).

Both Latino-Arabic set: (0,1,2,3,4,5,6,7,8,9) and Manipuri numerals are used in official documents as well as in day to day use.

Numeral Shapes	Explanation
၀	Manipuri Digit Zero
၁	Manipuri Digit One
၂	Manipuri Digit Two
၃	Manipuri Digit Three
၄	Manipuri Digit Four
၅	Manipuri Digit Five
၆	Manipuri Digit Six
၇	Manipuri Digit Seven
၈	Manipuri Digit Eight
၉	Manipuri Digit Nine

6.3.1.8. Punctuation Markers

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

???

A list of punctuations is provided below:

Sr. No.	Name of the marker	Marker Shape
1.	Question Mark	?
2.	Exclamation Mark	!
3.	Comma	,
4.	Apostrophe	'
5.	Semi Colon	;
6.	Colon	:
7.	Hyphen	-

8.	Dash	--
9.	Ellipsis mark	...
10.	Oblique	/
11.	Double quotation mark	" "
12.	Single quotation mark	' '
13.	Cross	XXX
14.	As Above	-- " --
15.	Round Brackets	()
16.	Square Brackets	[]
17.	Curly Brackets	{ }
18.	Abbreviation Sign	(◌)/(◌)/(◌)
19.	Manipuri Danda/Danri	
20.	Manipuri Double Danda/Double Danri	

6.3.1.9 Other Symbols

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

₹: Rupee Sign as mandated by Government of India

6.3.2. Consonant Mātrā/Kār Combinations.

This set is divided into three parts:

CM: The combination of Consonant and Matra

CMD (Anunasika) i.e. Consonant + Matra + Anunasika

CMD (Chandrabindu) i.e. Consonant + Matra + Chandrabindu.

In case you do not see any issues just tick the VALID box. In case you see issues tick invalid and provide the necessary correction for the combination in question.

- Please do not forget that some combinations are dead clusters but are still needed by the font designer to generate out the grammar.
- In case you feel a particular Consonant + Matra combination has been left out, please specify the same.
- In case a particular character combination is not used in your script, please cross it out.

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

Table 1:	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
Table 2:	ট	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন
Table 3:	প	ফ	ব	ভ	ম	য	র	ল	ৱ	শ
	স	হ	ড়	ঢ়	য়					ষ

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 ঙ + Mātrā/Kār is used only in few cases, it needs to be handled at the font level in the anticipation that a user could type this combination.

Although normally the combination of ঙ + ঞ is not acceptable in the language, to ensure that such a combination if entered by the user, should be displayed as: ঙ্ণ

The classes are as under:

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

6.3.2.2. refers to a concatenation of Consonant and Mātrā/Kār + 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.

6.3.2.1 Consonant and Mātrā/Kār combinations.

This set refers to a simple concatenation of Consonant and Mātrā/Kār. The experts from Manipuri Sahitya Parisad have stated as under:

In what follows these traditional clusters are provided as footnotes.

Consonant and Mātrā/Kār combinations Set 1

	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
া	কা	খা	গা	ঘা	ঙা	চা	ছা	জা	ঝা	X
ি	কি	খি	গি	ঘি	ঙি	চি	ছি	জি	ঝি	X
ী	কী	খী	গী	ঘী	ঙী	চী	ছী	জী	ঝী	X
ু	কু	খু	গু	ঘু	ঙু	চু	ছু	জু	ঝু	X
ূ	কূ	খূ	গূ	ঘূ	ঙূ	চূ	ছু	জূ	ঝূ	X
্	ক্	খ্	গ্	ঘ্	X	চ্	ছ্	জ্	ঝ্	X
ে	কে	খে	গে	ঘে	ঙে	চে	ছে	জে	ঝে	X
ৈ	কৈ	খৈ	গৈ	X	ঙৈ	চৈ	ছৈ	জৈ	ঝৈ	X
ো	কো	খো	গো	ঘো	ঙো	চো	ছো	জো	ঝো	X
ৌ	কৌ	খৌ	গৌ	ঘৌ	ঙৌ	চৌ	ছৌ	জৌ	ঝৌ	X

Consonant and Mātrā/Kār combinations Set 2

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

	ট	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন
া	টা	ঠা	ডা	ঢা	ণা	তা	থা	দা	ধা	না
ি	টি	ঠি	ডি	ঢি	ণি	তি	থি	দি	ধি	নি
ী	টী	ঠী	ডী	ঢী	ণী	তী	থী	দী	ধী	নী
ু	টু	ঠু	ডু	ঢু	ণু	তু	থু	দু	ধু	নু
ূ	টু	ঠু	ডু	ঢু	ণু	তু	থু	দু	ধু	নু
্	ট্	ঠ্	ড্	ঢ্	ণ্	ত্	থ্	দ্	ধ্	ন্
ে	টে	ঠে	ডে	ঢে	ণে	তে	থে	দে	ধে	নে
ৈ	টে	ঠে	ডে	ঢে	ণে	তে	থে	দে	ধে	নে
ো	টো	ঠো	ডো	ঢো	ণো	তো	থো	দো	ধো	নো
ৌ	টৌ	ঠৌ	ডৌ	ঢৌ	ণৌ	তৌ	থৌ	দৌ	ধৌ	নৌ

Consonant and Mātrā/Kār combinations Set 3

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

	প	ফ	ব	ভ	ম	য	র	ল	ৱ	শ	ষ	স	হ	ক্ষ	ড়	ঢ	য
া	পা	ফা	বা	ভা	মা	যা	রা	লা	ৱা	শা	ষা	সা	হা	ক্ষা			য়া
ি	পি	ফি	বি	ভি	মি	যি	রি	লি	ৱি	শি	ষি	সি	হি	ক্ষি			য়ি
ী	পী	ফী	বী	ভী	মী	যী	রী	লী	ৱী	শী	ষী	সী	হী	ক্ষী			য়ী
ু	পু	ফু	বু	ভু	মু	যু	রু	লু	ৱু	শু	ষু	সু	হু	ক্ষু			য়ু
ূ	পূ	ফূ	বূ	ভূ	মূ	যূ	রূ	লূ	ৱূ	শূ	ষূ	সূ	হূ	ক্ষূ			য়ূ
্	প্	ফ্	ব্	ভ্	ম্	য্				শ্	ষ্	স্	হ্	ক্ষ্			
ে	পে	ফে	বে	ভে	মে	যে	রে	লে	ৱে	শে	ষে	সে	হে	ক্ষে			য়ে
ৈ	পৈ	ফৈ	বৈ	ভৈ	মৈ	যৈ	রৈ	লৈ	ৱৈ	শৈ	ষৈ	সৈ	হৈ	ক্ষে			য়ে
ো	পো	ফো	বো	ভো	মো	যো	রো	লো	ৱো	শো	ষো	সো	হো	ক্ষো			য়ো
ৌ	পৌ	ফৌ	বৌ	ভৌ	মৌ	যৌ	রৌ	লৌ	ৱৌ	শৌ	ষৌ	সৌ	হৌ	ক্ষৌ			য়ৌ

6.3.2.2 Consonant and Mātrā/Kār +Nasal combinations.

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

Consonant and Mātrā/Kār + Nasal combinations: With Anuswar - Set 1

	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
ং	কং	খং	গং	ঘং	ঙং	চং	ছং	জং	ঝং	X
াং	কাং	খাং	গাং	ঘাং	X	চাং	ছাং	জাং	ঝাং	X
িং	কিং	খিং	গিং	ঘিং	X	চিং	ছিং	জিং	ঝিং	X
ীং	X	X	X	X	X	X	X	X	X	X
ুং	কুং	খুং	গুং	ঘুং	X	চুং	ছুং	জুং	ঝুং	X
ূং	X	X	X	X	X	X	X	X	X	X
্ং	ক্ং	খ্ং	গ্ং	ঘ্ং	X	চ্ং	চ্ং	জ্ং	ঝ্ং	X
েং	কেং	খেং	গেং	ঘেং	X	চেং	ছেং	জেং	ঝেং	X
ৈং	কৈং	খৈং	গৈং	ঘৈং	X	চৈং	ছৈং	জৈং	ঝৈং	X
োং	কোং	খোং	গোং	ঘোং	X	চোং	ছোং	জোং	ঝোং	X
ৌং	কৌং	খৌং	গৌং	ঘৌং	X	চৌং	ছৌং	জৌং	ঝৌং	X

Consonant and Mātrā/Kār + Nasal combinations: With Anuswar - Set 2

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

	ট	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন
ং	টং	ঠং	ডং	ঢং	ণং	তং	থং	দং	ধং	নং
াং	টাং	ঠাং	ডাং	ঢাং	ণাং	তাং	থাং	দাং	ধাং	নাং
িং	টিং	ঠিং	ডিং	ঢিং	ণিং	তিং	থিং	দিং	ধিং	নিং
ীং	টীং	ঠীং	ডীং	ঢীং	ণীং	তীং	থীং	দীং	ধীং	নীং
ুং	টুং	ঠুং	ডুং	ঢুং	ণুং	তুং	থুং	দুং	ধুং	নুং
ূং	টুং	ঠুং	ডুং	ঢুং	ণুং	তুং	থুং	দুং	ধুং	নুং
্ং	ট্ং	ঠ্ং	ড্ং	ঢ্ং	ণ্ং	ত্ং	থ্ং	দ্ং	ধ্ং	ন্ং
েং	টেং	ঠেং	ডেং	ঢেং	ণেং	তেং	থেং	দেং	ধেং	নেং
ৈং	টৈং	ঠৈং	ডৈং	ঢৈং	ণৈং	তৈং	থৈং	দৈং	ধৈং	নৈং
োং	টোং	ঠোং	ডোং	ঢোং	ণোং	তোং	থোং	দোং	ধোং	নোং
ৌং	টৌং	ঠৌং	ডৌং	ঢৌং	ণৌং	তৌং	থৌং	দৌং	ধৌং	নৌং

Consonant and Mātrā/Kār + Nasal combinations: With Anuswar - Set 3

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

	প	ফ	ব	ভ	ম	য	র	ল	ৱ	শ	ষ	স	হ	ড়	ঢ়	য়
ং	পং	ফং	বং	ভং	মং	যং	রং	লং	ৱং	শং	ষং	সং	হং	ড়ং	ঢ়ং	য়ং
া ং	পাং	ফাং	বাং	ভাং	মাং	যাং	রাং	লাং	ৱাং	শাং	ষাং	সাং	হাং	ড়াং	ঢ়াং	য়াং
ি ং	পিং	ফিং	বিং	ভিং	মিং	যিং	রিং	লিং	ৱিং	শিং	ষিং	সিং	হিং	ড়িং	ঢ়িং	য়িং
ী ং	পীং	ফীং	বীং	ভীং	মীং	যীং	রীং	লীং	ৱীং	শীং	ষীং	সীং	হীং	ড়ীং	ঢ়ীং	য়ীং
ুং	পুং	ফুং	বুং	ভুং	মুং	যুং	রুং	লুং	ৱুং	শুং	ষুং	সুং	হুং	ড়ুং	ঢ়ুং	য়ুং
ূং	পূং	ফূং	বূং	ভূং	মূং	যুং	রুং	লুং	ৱুং	শুং	ষুং	সুং	হুং	ড়ুং	ঢ়ুং	য়ুং
্ং	প্ং	ফ্ং	ব্ং	ভ্ং	ম্ং	X	X	ল্ং	ৱ্ং	শ্ং	ষ্ং	স্ং	হ্ং	X	X	X
ে ং	পেং	ফেং	বেং	ভেং	মেং	যেং	রেং	লেং	ৱেং	শেং	ষেং	সেং	হেং	ড়েং	ঢ়েং	য়েং
ৈ ং	পৈং	ফৈং	বৈং	ভৈং	মৈং	যৈং	রৈং	লৈং	ৱৈং	শৈং	ষৈং	সৈং	হৈং	ড়েং	ঢ়েং	য়েং
ো ং	পোং	ফোং	বোং	ভোং	মোং	যোং	রোং	লোং	ৱোং	শোং	ষোং	সোং	হোং	ড়েং	ঢ়োং	য়োং
ৌ ং	পৌং	ফৌং	বৌং	ভৌং	মৌং	যৌং	রৌং	লৌং	ৱৌং	শৌং	ষৌং	সৌং	হৌং	ড়েং	ঢ়ৌং	য়ৌং

Consonant and Mātrā/Kār + Nasal combinations: With Chandrabindu - Set 1

	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
ঁ	কঁ	খঁ	গঁ	ঘঁ	ঙঁ	চঁ	ছঁ	জঁ	ঝঁ	X
াঁ	কাঁ	খাঁ	গাঁ	ঘাঁ	ঙাঁ	চাঁ	ছাঁ	জাঁ	ঝাঁ	X
িঁ	কিঁ	খিঁ	গিঁ	ঘিঁ	ঙিঁ	চিঁ	ছিঁ	জিঁ	ঝিঁ	X
ীঁ	কীঁ	খীঁ	গীঁ	ঘীঁ	ঙীঁ	চীঁ	ছীঁ	জীঁ	ঝীঁ	X
ুঁ	কুঁ	খুঁ	গুঁ	ঘুঁ	ঙুঁ	চুঁ	ছুঁ	জুঁ	ঝুঁ	X
ূঁ	কূঁ	খূঁ	গূঁ	ঘূঁ	ঙূঁ	চূঁ	ছুঁ	জূঁ	ঝূঁ	X
্ৰঁ	ক্ৰঁ	খ্ৰঁ	গ্ৰঁ	ঘ্ৰঁ	ঙ্ৰঁ	চ্ৰঁ	ছ্ৰঁ	জ্ৰঁ	ঝ্ৰঁ	X
েঁ	কেঁ	খেঁ	গেঁ	ঘেঁ	ঙেঁ	চেঁ	ছেঁ	জেঁ	ঝেঁ	X
ৈঁ	কৈঁ	খৈঁ	গৈঁ	ঘৈঁ	ঙৈঁ	চৈঁ	ছৈঁ	জৈঁ	ঝৈঁ	X
োঁ	কোঁ	খোঁ	গোঁ	ঘোঁ	ঙোঁ	চোঁ	ছোঁ	জোঁ	ঝোঁ	X
ৌঁ	কৌঁ	খৌঁ	গৌঁ	ঘৌঁ	ঙৌঁ	চৌঁ	ছৌঁ	জৌঁ	ঝৌঁ	X

Consonant and Mātrā/Kār +Nasal combinations: With Chandrabindu - Set 2


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

	ট	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন
ঁ	টঁ	ঠঁ	ডঁ	ঢঁ	ণঁ	তঁ	থঁ	দঁ	ধঁ	নঁ
াঁ	টাঁ	ঠাঁ	ডাঁ	ঢাঁ	ণাঁ	তাঁ	থাঁ	দাঁ	ধাঁ	নাঁ
িঁ	টিঁ	ঠিঁ	ডিঁ	ঢিঁ	ণিঁ	তিঁ	থিঁ	দিঁ	ধিঁ	নিঁ
ীঁ	টীঁ	ঠীঁ	ডীঁ	ঢীঁ	ণীঁ	তীঁ	থীঁ	দীঁ	ধীঁ	নীঁ
ুঁ	টুঁ	ঠুঁ	ডুঁ	ঢুঁ	ণুঁ	তুঁ	থুঁ	দুঁ	ধুঁ	নুঁ
ূঁ	টূঁ	ঠূঁ	ডূঁ	ঢূঁ	ণূঁ	তূঁ	থূঁ	দূঁ	ধূঁ	নূঁ
্ঁ	ট্ঁ	ঠ্ঁ	ড্ঁ	ঢ্ঁ	ণ্ঁ	ত্ঁ	থ্ঁ	দ্ঁ	ধ্ঁ	ন্ঁ
েঁ	টেঁ	ঠেঁ	ডেঁ	ঢেঁ	ণেঁ	তেঁ	থেঁ	দেঁ	ধেঁ	নেঁ
ৈঁ	টেঁ	ঠেঁ	ডেঁ	ঢেঁ	ণেঁ	তেঁ	থেঁ	দেঁ	ধেঁ	নেঁ
োঁ	টোঁ	ঠোঁ	ডোঁ	ঢোঁ	ণোঁ	তোঁ	থোঁ	দোঁ	ধোঁ	নোঁ
ৌঁ	টৌঁ	ঠৌঁ	ডৌঁ	ঢৌঁ	ণৌঁ	তৌঁ	থৌঁ	দৌঁ	ধৌঁ	নৌঁ

Consonant and Mātrā/Kār +Nasal combinations: With Chandrabindu - Set 3

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

	প	ফ	ব	ভ	ম	য	র	ল	ৱ	শ	ষ	স	হ	ড়	ঢ	য়
ঁ	পঁ	ফঁ	বঁ	ভঁ	মঁ	যঁ	রঁ	লঁ	ৱঁ	শঁ	ষঁ	সঁ	হঁ	ড়ঁ	ঢঁ	য়ঁ
াঁ	পাঁ	ফাঁ	বাঁ	ভাঁ	মাঁ	যাঁ	রাঁ	লাঁ	ৱাঁ	শাঁ	ষাঁ	সাঁ	হাঁ	ড়াঁ	ঢাঁ	য়াঁ
িঁ	পিঁ	ফিঁ	বিঁ	ভিঁ	মিঁ	যিঁ	রিঁ	লিঁ	ৱিঁ	শিঁ	ষিঁ	সিঁ	হিঁ	ড়িঁ	ঢিঁ	য়িঁ
ীঁ	পীঁ	ফীঁ	বীঁ	ভীঁ	মীঁ	যীঁ	রীঁ	লীঁ	ৱীঁ	শীঁ	ষীঁ	সীঁ	হীঁ	ড়ীঁ	ঢীঁ	য়ীঁ
ুঁ	পুঁ	ফুঁ	বুঁ	ভুঁ	মুঁ	যুঁ	রুঁ	লুঁ	ৱুঁ	শুঁ	ষুঁ	সুঁ	হুঁ	ড়ুঁ	ঢুঁ	য়ুঁ
ূঁ	পূঁ	ফূঁ	বূঁ	ভূঁ	মূঁ	যূঁ	রূঁ	লূঁ	ৱূঁ	শূঁ	ষূঁ	সূঁ	হূঁ	ড়ূঁ	ঢূঁ	য়ূঁ
্ঁ	প্ঁ	ফ্ঁ	ব্ঁ	ভ্ঁ	ম্ঁ	য্ঁ	X	ল্ঁ	ৱ্ঁ	শ্ঁ	ষ্ঁ	স্ঁ	হ্ঁ	X	X	X
েঁ	পেঁ	ফেঁ	বেঁ	ভেঁ	মেঁ	যেঁ	রেঁ	লেঁ	ৱেঁ	শেঁ	ষেঁ	সেঁ	হেঁ	ড়েঁ	ঢেঁ	য়েঁ
ৈঁ	পৈঁ	ফৈঁ	বৈঁ	ভৈঁ	মৈঁ	যৈঁ	রৈঁ	লৈঁ	ৱৈঁ	শৈঁ	ষৈঁ	সৈঁ	হৈঁ	ড়ৈঁ	ঢৈঁ	য়ৈঁ
োঁ	পোঁ	ফোঁ	বোঁ	ভোঁ	মোঁ	যোঁ	রোঁ	লোঁ	ৱোঁ	শোঁ	ষোঁ	সোঁ	হোঁ	ড়োঁ	ঢোঁ	য়োঁ
ৌঁ	পৌঁ	ফৌঁ	বৌঁ	ভৌঁ	মৌঁ	যৌঁ	রৌঁ	লৌঁ	ৱৌঁ	শৌঁ	ষৌঁ	সৌঁ	হৌঁ	ড়ৌঁ	ঢৌঁ	য়ৌঁ

⁹ Variant shape as per traditional orthography is 

6.3.3. The Ligature Set of Manipuri.

Manipuri has a large set of ligatural forms. These are combinations of Consonant+Halanta+Consonant (CHC) or CHCHC. 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 consonants)

These ligatures are presented as in the earlier case of Consonant+Mātrā/Kār 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 nature.

As in the case of Consonant+ Mātrā/Kār combinations, the Paschimbanga Manipuri Akademi experts have provided for traditional variants which are listed as footnotes.

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 “ক” and “ক” is ligature “ক্”.

CHC(combination of two consonants) - Set 1										
	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
ক	ক্		X	X	X	X	X	X	X	X
খ	X	X	X	X	X	X	X	X	X	X
গ	X	X		X	X	X	X	X	X	X
ঘ	X	X	X		X	X	X	X	X	X
ঙ	ক্	খ্	গ্	ঘ্	ঙ্	X	X	X	X	X
চ	X	X	X	X	X	চ্	চ্	X	X	চ্
ছ	X	X	X	X	X	X	X	X	X	X
জ	X	X	X	X	X	X	X	জ্	X	জ্
ঝ	X	X	X	X	X	X	X	X	X	X
ঞ	X	X	X	X	X	ঞ্	ঞ্	ঞ্	ঞ্	
ট	X	X	X	X	X	X	X	X	X	X

ঠ	X	X	X	X	X	X	X	X	X	X
ড	X	X	X	X	X	X	X	X	X	X
ঢ	X	X	X	X	X	X	X	X	X	X
ণ	X	X	X	X	X	X	X	X	X	X
ত	X	X	X	X	X	X	X	X	X	X
থ	X	X	X	X	X	X	X	X	X	X
দ	X	X	দগ	X	X	X	X	X	X	X
ধ	X	X	X	X	X	X	X	X	X	X
ন	X	X	X	X	X	X	X	X	X	X
প	X	X	X	X	X	X	X	X	X	X
ফ	X	X	X	X	X	X	X	X	X	X
ব	X	X	X	X	X	X	X	ভ	X	X
ভ	X	X	X	X	X	X	X	X	X	X
ম	X	X	X	X	X	X	X	X	X	X
য	X	X	X	X	X	X	X	X	X	X
র	ক	খ	গ	ঘ	X	চ	X	জ	ঝ	X
ল	ক্ষ	X	X	X	X	X	X	X	X	X
শ	X	X	X	X	X	শ্চ	শ্ছ	X	X	X
স	ক্ষ	ষ্ঠ	X	X	X	X	X	X	X	X
স্	ক্ষ	ষ্	X	X	X	X	X	X	X	X
হ	X	X	X	X	X	X	X	X	X	X
ড়	X	X	X	X	X	X	X	X	X	X
ঢ়	X	X	X	X	X	X	X	X	X	X
য্	X	X	X	X	X	X	X	X	X	X

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 “ক” and “ট” is ligature “ক্ট”.

CHC(combination of two consonants) - Set 2										
	ট	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন
ক	ক্ট	X	X	X	X	ক্‌ত	X	X	X	ক্‌ন
খ	X	X	X	X	X	X	X	X	X	X
গ	X	X	X	X	X	X	X	X	X	গ্‌ন
ঘ	X	X	X	X	X	X	X	X	X	X
ঙ	X	X	X	X	X	X	X	X	X	X
চ	X	X	X	X	X	X	X	X	X	X
ছ	X	X	X	X	X	X	X	X	X	X
জ	X	X	X	X	X	X	X	X	X	X
ঝ	X	X	X	X	X	X	X	X	X	X
ঞ	X	X	X	X	X	X	X	X	X	X
ট	ট্ট	X	X	X	X	X	X	X	X	X
ঠ	X	X	X	X	X	X	X	X	X	X
ড	X	X	X	X	X	X	X	X	X	X
ঢ	X	X	X	X	X	X	X	X	X	X
ণ	ণ্ট	ণ্ঠ	ণ্ড	X	X	X	X	X	X	X
ত	X	X	X	X	X	ত্‌ত	ত্‌থ	X	X	ত্‌ন
থ	X	X	X	X	X	X	X	X	X	X
দ	X	X	X	X	X	X	X	দ্‌দ	দ্‌ধ	X
ধ	X	X	X	X	X	X	X	X	X	X
ন	ন্ট	ন্ঠ	ন্ড	X	X	ন্‌ত	ন্‌থ	ন্‌দ	ন্‌ধ	ন্‌ন
প	প্ট	X	X	X	X	প্‌ত	X	X	X	প্‌ন

ফ	X	X	X	X	X	X	X	X	X	X
ব	X	X	X	X	X	X	X	ব্	ব্ধ	X
ভ	X	X	X	X	X	X	X	X	X	X
ম	X	X	X	X	X	মত	মথ	X	X	ম্ন
য	X	X	X	X	X	X	X	X	X	X
র	ট		ড		ণ	ত	থ	দ	ধ	ন
ল	ল্ট	X	ল্ড	X	X	ল্ত	X	ল্দ	X	X
শ	X	X	X	X	X	X	X	X	X	শ্ন
ষ	ষ্ট	ষ্ঠ	X	X	ফ	X	X	X	X	ষ্ন
স	স্ট	X	X	X	X	স্ত	স্থ	X	X	স্ন
হ	X	X	X	X	হ্	X	X	X	X	হ্ন
ড়	X	X	X	X	X	X	X	X	X	X
ঢ়	X	X	X	X	X	X	X	X	X	X
য়	X	X	X	X	X	X	X	X	X	X

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 “ক” and “ম” is the ligature “ক্ম”.

CHC(combination of two consonants) - Set 3																
	প	ফ	ব	ভ	ম	য	র	ল	ব	শ	ষ	স	হ	ড়	ঢ়	য়
ক	X	X	ক্	X	ক্ম	ক্য	ক্র	ক্ল		X	ক্ষ	ক্স	X	X	X	X
খ	X	X	খ্	X	X	খ্য	খ্র	X		X	X	X	X	X	X	X
গ	X	X	গ্	X	গ্ম	গ্য	গ্র	গ্ল		X	X	X	X	X	X	X

য	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
ঙ	X	X	X	X	X	ঙ	ঙ	X		X	X	X	X	X	X	X
চ	X	X	X	X	X	চ	চ	X		X	X	X	X	X	X	X
ছ	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
জ	X	X	জ	X	X	জ্য	জ	X		X	X	X	X	X	X	X
ঝ	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
ঞ	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
ট	X	X	X	X	X	X	ট	X		X	X	X	X	X	X	X
ঠ	X	X	X	X	X	ঠ	X	X		X	X	X	X	X	X	X
ড	X	X	X	X	X	X	ড	X		X	X	X	X	X	X	X
ঢ	X	X	X	X	X	ঢ	X	X		X	X	X	X	X	X	X
ণ	X	X	X	X	X	ণ	X	X		X	X	X	X	X	X	X
ত	X	X	ত	X	ত	ত	ত	ত		X	X	X	X	X	X	X
থ	X	X	থ	X	X	X	X	X		X	X	X	X	X	X	X
দ	X	X	দ	X	দ	দ	দ	X		X	X	X	X	X	X	X
ধ	X	X	X	X	X	ধ	ধ	X		X	X	X	X	X	X	X
ন	X	X	ন	X	ন	ন	X	X		X	X	ন	X	X	X	X
প	প	X	X	X	X	প	প	প		X	X	প	X	X	X	X
ফ	X	X	X	X	X	ফ	ফ	ফ		X	X	X	X	X	X	X
ব	X	X	X	X	X	ব	ব	ব		X	X	X	X	X	X	X
ভ	X	X	X	X	X	ভ	ভ	X		X	X	X	X	X	X	X
ম	ম	ম	ম	ম	ম	ম	ম	ম		X	X	X	X	X	X	X
য	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
র	প	ফ	ব	ভ	ম	য	র	ল		শ	ষ	স	X	X	X	X

ল্	ল্ল	লফ	ল্ব	X	ল্ম	ল্য	ল্র	ল্ল		X	X	X	ল্হ	X	X	X
শ্	X	X	শ্ব	X	শ্ম	শ্য	শ্র	শ্ল		X	X	X	X	X	X	X
ষ্	X	X	X	X	ষ্ম	ষ্য	X	X		X	X	X	X	X	X	X
স্	স্প	X	স্ব	X	স্ম	স্য	স্র	স্ল		X	X	X	X	X	X	X
ত্	X	X	ত্ব	X	ত্ম	X	X	ত্ল		X	X	X	X	X	X	X
ক্ষ্	X	X	X	X	ক্ষ্ম	ক্ষ্য	X	X	X	X	X	X	X	X	X	X
ড্	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
ঢ্	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
য্	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X

Deviant list

Single glyphs:

Composing consonants	Resultant combination (s)
ক ্ট	ক্ট
ক ্ত	ক্ত
ক ্র	ক্ৰ
ক ্ষ	ক্ক্ষ
ক ্স	ক্ক্স

খ ্য	খ্য
গ ্য	গ্য
ঙ ্ক	ঙ্ক
ঙ ্গ	ঙ্গ
চ ্য	চ্য
জ ্ঞ	জ্ঞ
জ ্য	জ্য
ঞ ্চ	ঞ্চ
ঞ ্ছ	ঞ্ছ
ঞ ্জ	ঞ্জ
ন ্ড	ন্ড
ত ্ত	ত্ন্ত
ত ্থ	ত্থ
ত ্ম	ত্ম
ত ্য	ত্যা
ত ্র	ত্র

দ্‌	দ
দ্‌ধ	ধ
দ্‌য	দ্য
ধ্‌য	ধ্য
ন্‌থ	ন্থ
ন্‌ধ	নধ
ন্‌ম	নম
ন্‌ষ	ন্য /  ¹⁰
ন্‌স	নস
প্‌ট	প্ত
প্‌য	প্য
ব্‌দ	বদ
ব্‌য	ব্য
ভ্‌য	ভ্য
ভ্‌র	ভ্র /  ¹¹

¹⁰ Variant shape proposed by the Manipuri Expert

¹¹ Variant shape proposed by the Manipuri Expert

ম্ য	ম্য
শ্ ম	শ্ম
শ্ য	শ্য /  ¹²
য্ ঙ	য্ ঙ
য্ য	য্য
য্ ট	য্ ট /  ¹³
য্ ঠ	য্ ঠ
য্ ম	য্ম
য্ য	য্য
স্ ক	স্ক
স্ ট	স্ট
স্ ত	স্ত
স্ থ	স্ থ
স্ ঙ	স্ ঙ /  ¹⁴

¹² Variant shape proposed by the Manipuri Expert

¹³ Variant shape proposed by the Manipuri Expert

¹⁴ Variant shape proposed by the Manipuri Expert

স ্প	BR / স্প / স্প ¹⁵
স ্ব	স্ব / স্ব ¹⁶
স ্ম	স্ম / স্ম ¹⁷
স ্য	স্য
স ্র	স্র / স্র ¹⁸
হ ্ন	হ্ন
হ ্ম	হ্ম
হ ্ল	হ্ল
ক্ষ ্ম	ক্ষ্ম
ক্ষ ্য	ক্ষ্য
ফ ্য	ফ্য
ট ্ট	ট্ট
ল ্ট	ল্ট
ল ্ফ	ল্ফ

¹⁵ Variant shapes proposed by the Manipuri Expert

¹⁶ Variant shape proposed by the Manipuri Expert

¹⁷ Variant shape proposed by the Manipuri Expert

¹⁸ Variant shape proposed by the Manipuri Expert

ল ্ম	ল্ম
ল ্য	ল্য
ল ্ড	ল্ড
হ ্ল	হ্ল

Stacked Consonants:

Composing consonants Resultant combination

ক ্ক	ক্ক
ক ্ম	ক্ক্ম
ক ্ব	ক্ক্ব
ক ্ল	ক্ক্ল
খ ্ব	ক্ক্ব
খ ্র	ক্ক্র
গ ্ন	ক্ক্ন
গ ্ব	ক্ক্ব

গ ্র	গ্র
গ ল	গ্ৰ
ঘ ন	ঘ্ন
ঙ খ	ঙ্খ
ঙ য	ঙ্য
ঙ ঙ	ঙ্ঙ
ঙ র	ঙ্র
চ চ	চ্চ
চ ছ	চ্ছ
চ র	চ্ৰ
জ জ	জ্জ
জ ব	জ্ৰ
জ র	জ্র
ট র	ট্র
ড র	ড্র
ত ন	ত্ন

ত্ৰ	ত্ৰ
থ্ৰ	থ্ৰ
থ্ৰ	থ্ৰ
দ্ৰ	দ্ৰ
দ্ৰ	দ্ৰ
ধ্ৰ	ধ্ৰ
ধ্ৰ	ধ্ৰ / ধ্ব ¹⁹
ন্ত	ন্ত
ন্দ	ন্দ
ন্ন	ন্ন
ন্ৰ	ন্ৰ
প্ত	প্ত
প্ন	প্ন
প্প	প্প / প্প ²⁰
প্ৰ	প্ৰ

¹⁹ Variant shape proposed by the Manipuri Expert

²⁰ Variant shape proposed by the Manipuri Expert

প ্ল	প্ল
ফ ্র	ফ্র
ফ ্ল	ফ্ল
ফ ্ট	ফ্ট
ব ্দ	ব্দ
ব ্ল	ব্ল
ম ্ন	ম্ন
ম ্প	ম্প
ম ্ফ	ম্ফ
ম ্‌ব	ম্‌ব
ম ্‌ভ	ম্‌ভ
ম ্ম	ম্ম
ম ্র	ম্র
ম ্ল	ম্ল
ল ্ক	ল্ক
ল ্ত	ল্ত

ল ্ব	ল্ব
ল ্দ	ল্দ
ল ্প	ল্প / ল্প ²¹
শ ্ন	শ্ন
শ ্ব	শ্ব
শ ্র	শ্র
শ ্ৰ	শ্ৰ
হ ্ণ	হ্ণ
হ ্ব	হ্ব

6.3.4 The Collation Order of Manipuri.

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 Devanagari 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. In Nepali, the two conjunct glyphs are given at the end of the sort order. 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.

²¹ Variant shape proposed by the Manipuri Expert

In the case of Manipuri the following is the traditional sort order as determined by the experts.

ঁ ং ঃ অ আ ই ঙ্গ উ উ ঋ ঌ এ ঐ ও ঔ ক
 খ গ ঘ ঙ চ ছ জ ঝ ঞ ট ঠ ড ঢ ণ ত থ দ
 ধ ন প ফ ব ভ ম য র ল ঝ শ ষ স হ ঙ্গ ড় ঢ়
 য ং

In Tabular format:

ঁ	ং	ঃ	অ	আ	ই	ঙ্গ	উ	ঊ	ঋ
এ	ঐ	ও	ঔ	ক	খ	গ	ঘ	ঙ	চ
ছ	জ	ঝ	ঞ	ট	ঠ	ড	ঢ	ণ	ত
থ	দ	ধ	ন	প	ফ	ব	ভ	ম	য
র	ল	ঝ	র	শ	ষ	স	হ	ড়	ঢ়
া	ি	ী	ু	ূ	্	ে	ৈ	ো	ৌ
ং									

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 Bengali

Link: <http://unicode.org/charts/PDF/U0980.pdf>