

# **SCRIPT GRAMMAR FOR BANGLA 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
Node.....	17
Mātrā/Kār .....	17
Modifier .....	17
Mātrā/Kār+Modifier .....	17
6.3 SCRIPT GRAMMAR PROPER.....	18
6.3.1. The Character Set of Bangla. ....	18
6.3.2. Consonant Mātrā/Kār Combinations. ....	25
6.3.3. The Ligature Set of Bangla. ....	35
6.3.4 The Collation Order of Bangla. ....	57
7. REFERENCES .....	58
8. ANNEXURES.....	59
Annexure 1: Names of experts who have contributed to the script grammar.....	59
Annexure 2: Unicode Table Bengali.....	60

## 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 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 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 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. TERMINOLOGY<sup>1</sup>

**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 secondary<sup>2</sup>

**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

---


<sup>1</sup> 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.

<sup>2</sup> Wikipedia definition

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 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.<sup>3</sup>

**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**)

---

<sup>3</sup> 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 occur<sup>4</sup>
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.

---

<sup>4</sup> 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: BANGLA**  
**ISO Mnemonics: *ben***

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 Bangla is written using the Bengali 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 ঞ or alternatively form an entirely new shape  
ক্ + ষ = ক্ষ.

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 Bangla language

The experts from Paschimbanga Bangla Akademi have recommended the following change:

ঞ which is used in Assamese can also be accommodated in Bangla character set as a special character. However the said change is already incorporated in Unicode.

## 6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

**Bangla 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 Bangla.

### (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	শ	ষ
স	হ	ড়	ঢ়	য়
ৎ				

### (V) Vowels

অ	আ	ই	ঈ	উ	ঊ	ঋ	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.
◌: Avagraha	For extra length with long vowels as seen in the Sanskrit text transcribed in Bangla e.g. শ্রেয়োহনুপশ্যামি

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

(N)<sup>5</sup> Nukta ◌ - is used in Bangla (in ড়, ঢ় and ঞ)

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	<b>C</b>	N,M,D,H
C	<b>N</b>	C,M,D,H
-	<b>V<sup>6</sup></b>	D
C, N	<b>M</b>	D
C, N,V,M	<b>D</b>	-
C, N	<b>H</b>	C
	<b>Chandrabindu</b>	Anuswar/Visarga

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.

As a special case in Bangla alone, the Chandrabindu can be followed by an Anuswar or a Visarga in words like

হ্যাঁৎচা and হ্যাঁঃ respectively.

### 6.2.1.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:

<sup>5</sup> The nukta is a small dot placed under a character in Northern scripts to show that they are flapped or for deriving 5 other consonants in the Devanāgarī and Punjabi scripts, required for Urdu क़,ख़,ग़,ज़,फ़

<sup>6</sup> As a special case in Bangla two vowels অ and ঐ can be followed by a Hasanta to generate out অ্যা and ঐ্যা respectively representing the vowel /æ/. Both vowel combinations are in free variation i.e. can be used indifferently. However the Paschimbanga Bangla Akademi recommends the use of অ্যা

1. A vowel syllable (akshar): a full vowel.
2. A consonant syllable (akshar): a full consonant (having a weak vowel or a mātrā/kār)

The four other subsets viz. Mātrā/Kārs, Vowel Modifiers, Halanta/Hasanta 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 /i/, आः /āH/
- 1.3. As a special case in Bangla two vowels अ and ए can be followed by a Hasanta to generate out आ and ए (ja phalā+ā kār) respectively representing the vowel /æ/. Both vowel combinations are in free variation i.e. can be used indifferently. However the Bangla Akademi recommends the use of आ.

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 consonant<sup>7</sup> (with or without Nukta) followed by a mātrā/kār i.e. the inherent vowel being substituted by another vowel: कि /ki/, डि /ḍi/
- 2.3. A consonant (with or without Nukta) followed by a modifier: कँ /kẽ/, हः /haH<sup>8</sup>/
- 2.4. A consonant (with or without Nukta) followed by a mātrā/kār and a modifier: कू /kū/, दूः /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/kār, a modifier or a combination of both. These result in a ligature or what is often termed as *yuktakshara*<sup>9</sup>.

कु /kta/, क्त /ktā/, न्तः /ntaH/ क्तू /ktū/, न्दु /ndu/

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. However Bangla admits 4 Consonant clusters.<sup>10</sup>

This means that theoretically the following forms can be postulated:

1. Vowel Set: With the Vowel as the node.

V      VD

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

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

<sup>9</sup> The following theoretical consonant clusters are proposed

<sup>10</sup> Sanskrit admits a single case where five consonants can come together: कर्त्स्न्य /kārtsnya/ "wholeness", "entirety" (secondary derivative from the adjective कर्त्स्न /kārtsna/ meaning “whole, complete”).



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

Node	Mātrā/Kār	Modifier	Mātrā/Kār+Modifier
C <sup>11</sup>	CM	CD	CMD
CHC	CHCM	CHCD	CHCMD
CHCHC	CHCHCM	CHCHCD	CHCHCMD
CHCHCHC	CHCHCHCM	CHCHCHCD	CHCHCHCM

An exception in Bangla to the consonant set is the *khanda ta* which cannot be followed by a Mātrā/Kār, Hasanta or a Diacritic. Given this exception, a total number of 16 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.

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

---

<sup>11</sup> “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 Bangla. These are:

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

### 6.3.1. The Character Set of Bangla.

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 Bangla.

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 Bangla 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		

## Special consonant <sup>12</sup>

ৎ <sup>13</sup>
-----------------

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		

## Special consonant

<sup>12</sup> This categorization has been proposed by the experts of the Paschimbanga Bangla Akademi

<sup>13</sup> the *khanda ta* is a special consonant in Bangla since unlike other consonants it cannot be followed by a Mātrā/Kār, Hasanta or a Diacritic but it can form a typical conjunct like ত্ৰ (ত্ৰ + ত্ৰ)

### 6.3.1.2. The Vowel Set

The Vowel set of Bangla is as under:

অ	BENGALI LETTER A
আ	BENGALI LETTER AA
ই	BENGALI LETTER I
ঈ	BENGALI LETTER II
উ	BENGALI LETTER U
ঊ	BENGALI LETTER UU
ঋ	BENGALI LETTER VOCALIC R
এ	BENGALI LETTER E
ঐ	BENGALI LETTER AI
ও	BENGALI LETTER O
ঔ	BENGALI 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 Bangla is as under:

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

4. Bengali sign U	ু	উ	ক্ + উ = কু
5. Bengali sign UU	ূ	উ	ক্ + উ = কু
6. Bengali sign vocalic R	্	ঋ	ক্ + ঋ = কৃ
7. Bengali sign E <sup>14</sup>	ে/ ৈ	এ	ক্ + এ = কে/ কৈ
8. Bengali sign AI	ৈ/ ৈ	ঐ	ক্ + ঐ = কৈ/ কৈ
9..Bengali sign O	ো/ ৌ	ও	ক্ + ও = কো/ কৌ
10. Bengali sign AU	ৌ/ ৌ	ঔ	ক্ + ঔ = কৌ/ 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. Bangla 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

<sup>14</sup> In the case of E, AI, O, AU, variants exist which are shown above. Their distribution is as under. Without shirorekha (the top bar) vowel modifiers are used in initial position (at the beginning of the word) as in বেল, the vowel modifiers with shirorekha are used in Medial and Final positions as in অনেক and বলে. Similar is the case with বৈশাখ ~ মাইভৈং, কোল ~ সন্তোষ, মৌ ~ দদৌ

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

The र takes a variety of shapes known as rakar (ra phalā in Bangla) and rafar/repha (reph in Bangla) 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 ट, it is appended in the shape of a ^ to the bottom of the character (bottom rakar/ra phalā). Bangla has the following combinations of ra:

#### RAFAR/ REPH

ৰ্ক      ত

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

তৰ্ক, কৰ্তব্য, অৰ্জন

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): নৈখতি

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

#### RA PHALĀ (RAKARS)

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

বক্ৰ, রাষ্ট্ৰ, পুত্ৰ

### 6.3.1.6. Diacritics

These are as under in the case of Bangla:

ং - Anuswar অংশ

ঁ - Chandrabindu/Anunasika চাঁদ

্ - Halanta/Hasanta প্রাক্কথন

ঃ - Visarga দুঃখ

২ - Avagraha for extra length with long vowels e.g. / শ্ৰেয়োহনুপশ্যামি/

### 6.3.1.7. Numerals

Following are the numbers used in Bangla language.

There is no fixed policy for use of Numbers. Both Latino-Arabic set: (0,1,2,3,4,5,6,7,8,9) and Bangla numerals are used in official documents as well as in day to day use.

১ ২ ৩ ৪ ৫ ৬ ৭ ৮ ৯ ০

Numeral Shapes	Explanation
০	Bengali Digit Zero
১	Bengali Digit One
২	Bengali Digit Two
৩	Bengali Digit Three
৪	Bengali Digit Four
৫	Bengali Digit Five
৬	Bengali Digit Six
৭	Bengali Digit Seven
৮	Bengali Digit Eight
৯	Bengali Digit Nine

### 6.3.1.8. Punctuation Markers

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

Purna and Deerga Virama (full-stop/danda/danri) Devanagari code block: U+0964,

U+0965 ।,॥. । is used to mark the full stop and ॥ is used for writing poetry of middle Bangla.

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.	Bengali Danda/Danri	
20.	Bengali 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

ট: Bengali rupee/tākā sign

৳: Bengali rupee mark

৳: Bengali currency numerator one (not in current usage)

৳: Bengali currency numerator two (not in current usage)

৳: Bengali currency numerator three (not in current usage)

৳: Bengali currency numerator four

৳: Bengali currency numerator one less than the denominator

৳: Bengali currency denominator sixteen

৳: Bengali Ishwar



### 6.3.2. Consonant Mātrā/Kār Combinations.

These refer to the shapes generated when a Mātrā/Kār 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:	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	শ	ষ
	স	হ	ড়	ঢ়	য়					

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 + ্ is not acceptable in the language, to ensure that such a combination if entered by the user, should be displayed as: ঙ্

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 Paschimbanga Bangla Akademi have stated as under:

*“The traditional clusters will be given, and by their side, the new ‘transparent clusters’ suggested by the Akademi”*

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	ঞা
ি	কি	খি	গি	ঘি	ঙি	চি	ছি	জি	ঝি	ঞি
ী	কী	খী	গী	ঘী	ঙী	চী	ছী	জী	ঝী	ঞী
ু	কু	খু	গু <sup>15</sup>	ঘু	ঙু	চু	ছু	জু	ঝু	X
ূ	কূ	খূ	গূ	ঘূ	X	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	ঘৈ	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

Remark 1- ঙ and ঞ are not used as the first members of clusters

---

<sup>15</sup> Variant shape as per traditional orthography is 

## 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
া	ট্টা	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
ূ	X	X	X	X	ণূ	তূ	থূ	দূ	ধূ	নূ
্	ট্	ঠ্	ড্	ঢ্	ণ্	ত্	থ্	দ্	ধ্	ন্
ে	টে	ঠে	ডে	ঢে	ণে	তে	থে	দে	ধে	নে
ৈ	টৈ	ঠৈ	ডৈ	ঢৈ	ণৈ	তৈ	থৈ	দৈ	ধৈ	নৈ
ো	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	শ	ষ	স	হ	ড়	ঢ়	য়
া	পা	ফা	বা	ভা	মা	যা	রা	লা	শা	ষা	সা	হা	ড়া	ঢ়া	য়া
ি	পি	ফি	বি	ভি	মি	যি	রি	লি	শি	ষি	সি	হি	ড়ি	ঢ়ি	য়ি
ী	পী	ফী	বী	ভী	মী	যী	রী	লী	শী	ষী	সী	হী	ড়ী	ঢ়ী	য়ী
ু	পু	ফু	বু	ভু	মু	যু	রু <sup>16</sup>	লু	শু <sup>17</sup>	ষু	সু <sup>18</sup>	হু <sup>19</sup>	ড়ু	ঢ়ু	য়ু
ূ	পূ	ফূ	বূ	ভূ	মূ	যূ	রূ <sup>20</sup>	লূ	শূ	ষূ	সূ	হূ	X	X	য়ূ
্	প্	ফ্	ব্	ভ্	ম্	X	X	ল্	শ্	X	স্	হ্ <sup>21</sup>	X	X	X
ে	পে	ফে	বে	ভে	মে	যে	রে	লে	শে	ষে	সে	হে	ড়ে	ঢ়ে	য়ে
ৈ	পৈ	ফৈ	বৈ	ভৈ	মৈ	যৈ	রৈ	লৈ	শৈ	ষৈ	সৈ	হৈ	ড়ৈ	ঢ়ৈ	য়ৈ
ো	পো	ফো	বো	ভো	মো	যো	রো	লো	শো	ষো	সো	হো	ড়ো	ঢ়ো	য়ো
ৌ	পৌ	ফৌ	বৌ	ভৌ	মৌ	যৌ	রৌ	লৌ	শৌ	ষৌ	সৌ	হৌ	X	X	য়ৌ

<sup>16</sup> Variant shape as per traditional orthography is রু

<sup>17</sup> Variant shape as per traditional orthography is শু

<sup>18</sup> Variant shape as per traditional orthography is সু

<sup>19</sup> Variant shape as per traditional orthography is হু

<sup>20</sup> Variant shape as per traditional orthography is রূ

<sup>21</sup> Variant shape as per traditional orthography is হ্

### 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 <sup>22</sup>	ঘুং	X	চুং	ছুং	জুং	ঝুং	X
ূং	X	X	X	X	X	X	X	X	X	X
্ং	ক্ং	খ্ং	গ্ং	ঘ্ং	X	চ্ং	ছ্ং	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

<sup>22</sup> Variant shape as per traditional orthography is গুং

## 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
ীং	X	X	X	X	X	X	X	X	X	X
ুং	টুং	ঠুং	ডুং	ঢুং	ণুং	তুং	থুং	দুং	ধুং	নুং
ূং	X	X	X	X	X	X	X	X	X	X
্ং	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	শ	ষ	স	হ	ড়	ঢ	য়
ং	পং	ফং	বং	ভং	মং	যং	রং	লং	শং	ষং	সং	হং	ড়ং	ঢং	য়ং
াং	পাং	ফাং	বাং	ভাং	মাং	যাং	রাং	লাং	শাং	ষাং	সাং	হাং	ড়াং	ঢাং	য়াং
িং	পিং	ফিং	বিং	ভিং	মিং	যিং	রিং	লিং	শিং	ষিং	সিং	হিং	ড়িং	ঢিং	য়িং
ীং	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ুং	পুং	ফুং	বুং	ভুং	মুং	যুং	রুং <sup>23</sup>	লুং	শুং <sup>24</sup>	ষুং	সুং <sup>25</sup>	হুং <sup>26</sup>	ড়ুং	ঢুং	য়ুং
ূং	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
্ং	প্ং	ফ্ং	ব্ং	ভ্ং	ম্ং	X	X	ল্ং	শ্ং	X	স্ং	হ্ং <sup>27</sup>	X	X	X
েং	পেং	ফেং	বেং	ভেং	মেং	যেং	রেং	লেং	শেং	ষেং	সেং	হেং	ড়েং	ঢেং	য়েং
ৈং	পৈং	ফৈং	বৈং	ভৈং	মৈং	যৈং	রৈং	লৈং	শৈং	ষৈং	সৈং	হৈং	ডৈং	ঢৈং	য়ৈং
োং	পোং	ফোং	বোং	ভোং	মোং	যোং	রোং	লোং	শোং	ষোং	সোং	হোং	ডোং	ঢোং	য়োং
ৌং	পৌং	ফৌং	বৌং	ভৌং	মৌং	যৌং	রৌং	লৌং	শৌং	ষৌং	সৌং	হৌং	X	X	য়ৌং

<sup>23</sup> Variant shape as per traditional orthography is 

<sup>24</sup> Variant shape as per traditional orthography is 

<sup>25</sup> Variant shape as per traditional orthography is 

<sup>26</sup> Variant shape as per traditional orthography is 

<sup>27</sup> Variant shape as per traditional orthography is 

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

	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
ঁ	কঁ	খঁ	গঁ	ঘঁ	X	চঁ	ছঁ	জঁ	ঝঁ	X
াঁ	কাঁ	খাঁ	গাঁ	ঘাঁ	X	চাঁ	ছাঁ	জাঁ	ঝাঁ	X
িঁ	কিঁ	খিঁ	গিঁ	ঘিঁ	X	চিঁ	ছিঁ	জিঁ	ঝিঁ	X
ীঁ	কীঁ	খীঁ	গীঁ	ঘীঁ	X	চীঁ	ছীঁ	জীঁ	ঝীঁ	X
ুঁ	কুঁ	খুঁ	গুঁ <sup>28</sup>	ঘুঁ	X	চুঁ	ছুঁ	জুঁ	ঝুঁ	X
ূঁ	কুঁ	খুঁ	গুঁ	ঘুঁ	X	চুঁ	ছুঁ	জুঁ	ঝুঁ	X
্ঁ	ক্ঁ	খ্ঁ	গ্ঁ	ঘ্ঁ	X	চ্ঁ	ছ্ঁ	জ্ঁ	ঝ্ঁ	X
েঁ	কেঁ	খেঁ	গেঁ	ঘেঁ	X	চেঁ	ছেঁ	জেঁ	ঝেঁ	X
ৈঁ	কৈঁ	খৈঁ	গৈঁ	ঘৈঁ	X	চৈঁ	ছৈঁ	জৈঁ	ঝৈঁ	X
োঁ	কোঁ	খোঁ	গোঁ	ঘোঁ	X	চোঁ	ছোঁ	জোঁ	ঝোঁ	X
ৌঁ	কৌ	খৌ	গৌ	ঘৌ	X	চৌ	ছৌ	জৌ	ঝৌ	X

<sup>28</sup> Variant shape as per traditional orthography is 



## 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	X	তঁ	থঁ	দঁ	ধঁ	X
াঁ	টাঁ	ঠাঁ	ডাঁ	ঢাঁ	X	তাঁ	থাঁ	দাঁ	ধাঁ	X
িঁ	টিঁ	ঠিঁ	ডিঁ	ঢিঁ	X	তিঁ	থিঁ	দিঁ	ধিঁ	X
ীঁ	টীঁ	ঠীঁ	ডীঁ	ঢীঁ	X	তীঁ	থীঁ	দীঁ	ধীঁ	X
ুঁ	টুঁ	ঠুঁ	ডুঁ	ঢুঁ	X	তুঁ	থুঁ	দুঁ	ধুঁ	X
ূঁ	টুঁ	ঠুঁ	ডুঁ	ঢুঁ	X	তুঁ	থুঁ	দুঁ	ধুঁ	X
্ঁ	ট্ঁ	ঠ্ঁ	ড্ঁ	ঢ্ঁ	X	ত্ঁ	থ্ঁ	দ্ঁ	ধ্ঁ	X
েঁ	টেঁ	ঠেঁ	ডেঁ	ঢেঁ	X	তেঁ	থেঁ	দেঁ	ধেঁ	X
ৈঁ	টেঁ	ঠেঁ	ডেঁ	ঢেঁ	X	তেঁ	থেঁ	দেঁ	ধেঁ	X
োঁ	টোঁ	ঠোঁ	ডোঁ	ঢোঁ	X	তোঁ	থোঁ	দোঁ	ধোঁ	X
ৌঁ	টৌ	ঠৌ	ডৌ	ঢৌ	X	তৌ	থৌ	দৌ	ধৌ	X

### 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	শ	ষ	স	হ	ড়	ঢ	য়
ঁ	পঁ	ফঁ	বঁ	ভঁ	মঁ	যঁ	রঁ	লঁ	শঁ	ষঁ	সঁ	হঁ	ড়ঁ	ঢঁ	য়ঁ
াঁ	পাঁ	ফাঁ	বাঁ	ভাঁ	মাঁ	যাঁ	রাঁ	লাঁ	শাঁ	ষাঁ	সাঁ	হাঁ	ড়াঁ	ঢ়াঁ	য়াঁ
িঁ	পিঁ	ফিঁ	বিঁ	ভিঁ	মিঁ	যিঁ	রিঁ	লিঁ	শিঁ	ষিঁ	সিঁ	হিঁ	ড়িঁ	ঢ়িঁ	য়িঁ
ীঁ	পীঁ	ফীঁ	বীঁ	ভীঁ	মীঁ	যীঁ	রীঁ	লীঁ	শীঁ	ষীঁ	সীঁ	হীঁ	ড়ীঁ	ঢ়ীঁ	য়ীঁ
ুঁ	পুঁ	ফুঁ	বুঁ	ভুঁ	মুঁ	যুঁ	রুঁ <sup>29</sup>	লুঁ	শুঁ <sup>30</sup>	ষুঁ	সুঁ <sup>31</sup>	হুঁ <sup>32</sup>	ড়ুঁ	ঢ়ুঁ	য়ুঁ
ূঁ	পূঁ	ফূঁ	বূঁ	ভূঁ	মূঁ	যূঁ	রূঁ <sup>33</sup>	লূঁ	শূঁ	ষূঁ	সূঁ	হূঁ	ড়ূঁ	ঢ়ূঁ	য়ূঁ
্ঁ	প্ঁ	ফ্ঁ	ব্ঁ	ভ্ঁ	ম্ঁ	X	X	ল্ঁ	শ্ঁ	X	স্ঁ	হ্ঁ <sup>34</sup>	X	X	X
েঁ	পেঁ	ফেঁ	বেঁ	ভেঁ	মেঁ	যেঁ	রেঁ	লেঁ	শেঁ	ষেঁ	সেঁ	হেঁ	ড়েঁ	ঢ়েঁ	য়েঁ
ৈঁ	পৈঁ	ফৈঁ	বৈঁ	ভৈঁ	মৈঁ	যৈঁ	রৈঁ	লৈঁ	শৈঁ	ষৈঁ	সৈঁ	হৈঁ	ড়েঁ	ঢ়েঁ	য়েঁ
োঁ	পোঁ	ফোঁ	বোঁ	ভোঁ	মোঁ	যোঁ	রোঁ	লোঁ	শোঁ	ষোঁ	সোঁ	হোঁ	ড়োঁ	ঢ়োঁ	য়োঁ
ৌঁ	পৌ	ফৌ	বৌ	ভৌ	মৌ	যৌ	রৌ	লৌ	শৌ	ষৌ	সৌ	হৌ	X	X	য়ৌ

<sup>29</sup> Variant shape as per traditional orthography is

<sup>30</sup> Variant shape as per traditional orthography is

<sup>31</sup> Variant shape as per traditional orthography is

<sup>32</sup> Variant shape as per traditional orthography is

<sup>33</sup> Variant shape as per traditional orthography is

<sup>34</sup> Variant shape as per traditional orthography is

রুঁ  
শুঁ  
সুঁ  
হুঁ  
রুঁ  
শুঁ  
সুঁ  
হুঁ

### 6.3.3. The Ligature Set of Bangla.

Bangla 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

6.3.3.3. CHCHCHC

#### 6.3.3.1. CHC ( combination of two consonanats)

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 Bangla 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 “খ” is ligature “ক্খ”.

CHC( combination of two consonants) - Set 1										
	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
ক্	ক্	কখ <sup>35</sup>	X	X	X	X	X	X	X	X
খ্	X	X	X	X	X	X	X	X	X	X
গ্	X	X	গ্গ	X	X	X	X	X	X	X
ঘ্	X	X	X	ঘ্ঘ	X	X	X	X	X	X
ঙ্	ঙ্ক <sup>36</sup>	ঙ্খ	ঙ্গ <sup>37</sup>	ঙ্ঘ	ঙ্ঙ <sup>38</sup>	X	X	X	X	X
চ্	X	X	X	X	X	চ্চ	চ্ছ	X	X	চ্ঞ

<sup>35</sup> This conjunct might be needed for transcribing Pali and Prakrit text in Bangla

<sup>36</sup> Variant shape as per traditional orthography is ক্

<sup>37</sup> Variant shape as per traditional orthography is ঙ্

<sup>38</sup> This conjunct might be needed for transcribing Pali and Prakrit text in Bangla

ছ	X	X	X	X	X	X	X	X	X	X
জ্	X	X	X	X	X	X	X	জ্জ	জ্ঝ	জ্জ্জ
ঝ	X	X	X	X	X	X	X	X	X	X
ঞ	X	X	X	X	X	ঞ <sup>39</sup>	ঞ <sup>40</sup>	ঞ <sup>41</sup>	ঞ্	ঞ্জ <sup>42</sup>
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	ব্জ	X	X
ভ	X	X	X	X	X	X	X	X	X	X
ম	X	X	X	X	X	X	X	X	X	X

<sup>39</sup> Variant shape as per traditional orthography is ঞ

<sup>40</sup> Variant shape as per traditional orthography is ঞ্

<sup>41</sup> Variant shape as per traditional orthography is ঞ্জ

<sup>42</sup> This conjunct might be needed for transcribing Pali and Prakrit text in Bangla

য্	X	X	X	X	X	X	X	X	X	X
র্	ক্	খ্	গ্	ঘ্	X	চ্	ছ্	জ্	ঝ্	X
ল্	ক্	খ্	গ্	X	X	X	X	X	X	X
শ্	X	X	X	X	X	শ্চ	শ্ছ	X	X	X
ষ্	ক্	খ্	X	X	X	X	X	X	X	X
স্	ক্	খ্	X	X	X	X	X	X	X	X
হ্	X	X	X	X	X	X	X	X	X	X
ড্	X	X	ড্গ	X	X	X	X	X	X	X
ঢ্	X	X	X	X	X	X	X	X	X	X
য়্	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
ক্	ক্ট <sup>43</sup>	X	X	X	X	ক্ত <sup>44</sup>	X	X	X	ক্ন
খ্	X	X	X	X	X	X	X	X	X	X
গ্	X	X	X	X	গ্ণ	X	X	গ্দ	গ্ধ <sup>45</sup>	গ্ন
ঘ্	X	X	X	X	X	X	X	X	X	ঘ্ন
ঙ্	X	X	X	X	X	X	X	X	X	X
চ্	X	X	X	X	X	X	X	X	X	X
ছ্	X	X	X	X	X	X	X	X	X	X
জ্	X	X	X	X	X	X	X	X	X	X
ঝ্	X	X	X	X	X	X	X	X	X	X
ঞ্	X	X	X	X	X	X	X	X	X	X
ট্	ট্ট	ট্ট <sup>46</sup>	X	X	X	X	X	X	X	X
ঠ্	X	X	X	X	X	X	X	X	X	X
ড্	X	X	ড্‌	ডঢ <sup>47</sup>	X	X	X	X	X	X

<sup>43</sup> This conjunct is needed for writing loan words in Bangla

<sup>44</sup> Variant shape as per traditional orthography is ক্ত

<sup>45</sup> Variant shape as per traditional orthography is গ্ধ

<sup>46</sup> This conjunct might be needed for transcribing Pali and Prakrit text in Bangla

<sup>47</sup> This conjunct might be needed for transcribing Pali and Prakrit text in Bangla

ঢ়	X	X	X	ঢ়	X	X	X	X	X	X
ণ্	ণ্	ণ্	ণ্ <sup>48</sup>	ণ্	ণ্	X	X	X	X	X
ত্	X	X	X	X	X	ত্	ত্	X	X	ত্
থ্	X	X	X	X	X	X	X	X	X	X
দ্	X	X	X	X	X	X	X	দ্	দ্ <sup>49</sup>	X
ধ্	X	X	X	X	X	X	X	X	X	ধ্
ন্	ন্ <sup>50</sup>	ন্	ন্	ন্	X	ন্	ন্ <sup>51</sup>	ন্	ন্ <sup>52</sup>	ন্
প্	প্ <sup>53</sup>	X	X	X	X	প্	প্	X	X	প্
ফ্	X	X	X	X	X	X	X	X	X	X
ব্	X	X	X	X	X	X	X	ব্	ব্ <sup>54</sup>	X
ভ্	X	X	X	X	X	X	X	X	X	X
ম্	X	X	X	X	X	X	X	X	X	ম্
য্	X	X	X	X	X	X	X	X	X	X
র্	র্	র্	র্	র্	র্	র্	র্	র্	র্	র্
ল্	ল্	X	ল্ <sup>55</sup>	X	X	X	X	X	X	X

<sup>48</sup> Variant shape as per traditional orthography is ণ্

<sup>49</sup> Variant shape as per traditional orthography is দ্

<sup>50</sup> ন্ with retroflex letters are improvisations in the orthography of Bangla

<sup>51</sup> Variant shape as per traditional orthography is ঞ্

<sup>52</sup> Variant shape as per traditional orthography is ঞ্

<sup>53</sup> This conjunct is needed for writing loan words in Bangla

<sup>54</sup> Variant shape as per traditional orthography is ব্

<sup>55</sup> This conjunct is needed for writing loan words in Bangla

শ্	X	X	X	X	X	X	X	X	X	শ্
ষ্	ষ্ট	ষ্ঠ	X	X	ষ্ <sup>56</sup>	X	X	X	X	X
স্	স্ট <sup>57</sup>	X	X	X	X	স্ত	স্ <sup>58</sup>	X	X	স্
হ্	X	X	X	X	হ্	X	X	X	X	হ্
ড্	X	X	X	X	X	X	X	X	X	X
ঢ়্	X	X	X	X	X	X	X	X	X	X
ঝ্	X	X	X	X	X	X	X	X	X	X

---

<sup>56</sup> Variant shape as per traditional orthography is ষ্

<sup>57</sup> This conjunct is needed for writing loan words in Bangla

<sup>58</sup> Variant shape as per traditional orthography is ষ্



### 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	শ	ষ	স	হ	ড়	ঢ	য়
ক্	X	X	ক্	X	ক্ম	ক্য	ক্র <sup>59</sup>	ক্ল	X	ক্ষ	ক্স	X	X	X	X
খ্	X	X	খ্	X	X	খ্য	খ্র	X	X	X	X	X	X	X	X
গ্	X	X	গ্	X	গ্ম	গ্য	গ্র	গ্ল	X	X	X	X	X	X	X
ঘ্	X	X	X	X	X	ঘ্য	ঘ্র	ঘ্ল <sup>60</sup>	X	X	X	X	X	X	X
ঙ্	X	X	ঙ্	X	ঙ্ম	X	X	X	X	X	X	X	X	X	X
চ্	X	X	চ্	X	X	চ্য	X	X	X	X	X	X	X	X	X
ছ্	X	X	ছ্	X	X	ছ্য	ছ্র	X	X	X	X	X	X	X	X
জ্	X	X	জ্	X	X	জ্য	জ্র	X	X	X	X	X	X	X	X
ঝ্	X	X	X	X	X	ঝ্য	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
ট্	X	X	ট্	X	ট্ম	ট্য	ট্র	X	X	X	X	X	X	X	X
ঠ্	X	X	X	X	X	ঠ্য	X	X	X	X	X	X	X	X	X
ড্	X	X	ড্	X	ড্ম	ড্য	ড্র	X	X	X	X	X	X	X	X
ঢ্	X	X	X	X	X	ঢ্য	X	X	X	X	X	X	X	X	X
ণ্	X	X	ণ্	X	ণ্ম	ণ্য	X	X	X	X	X	X	X	X	X

<sup>59</sup> Variant shape as per traditional orthography is ক্র

<sup>60</sup> This conjunct might be needed for transcribing tantric texts in Bangla

ত্	X	X	ত্ব	X	ত্বা	ত্যা	ত্র	X	X	X	X	X	X	X	X
থ্	X	X	থ্ব	X	থ্বা	থ্যা	থ্র	X	X	X	X	X	X	X	X
দ্	X	X	দ্ব	ডু	দ্বা	দ্য <sup>61</sup>	দ্র	X	X	X	X	X	X	X	X
ধ্	X	X	ধ্ব	X	ধ্বা	ধ্য	ধ্র	X	X	X	X	X	X	X	X
ন্	X	X	ন্ব	X	ন্বা	ন্য <sup>62</sup>	X	X	X	X	ঞ	X	X	X	X
প্	ঞ	X	X	X	X	প্যা	প্র	প্প	X	X	ঞ	X	X	X	X
ফ্	X	X	X	X	X	ফ্যা	ফ্র	ফ্ফ <sup>63</sup>	X	X	X	X	X	X	X
ব্	X	X	ব্ব	X	X	ব্য	ব্র	ব্ব	X	X	X	X	X	X	X
ভ্	X	X	ভ্ব	X	X	ভ্যা	ভ্র	ভ্ভ <sup>64</sup>	X	X	X	X	X	X	X
ম্	ঞ	ম্ফ	ম্ব	ম্ভ	ম্ম	ম্যা	ম্র	ম্ম	X	X	X	X	X	X	X
য্	X	X	X	X	X	য্যা	X	X	X	X	X	X	X	X	X
র্	র্প	র্ফ	র্ব	র্ভ	র্ম	র্য <sup>65</sup>	র্	র্ল	র্শ	র্ষ	র্স	ই	X	X	X
ল্	ল্ল	ল্ফ	ল্ব	ল্ভ	ল্ম	ল্যা	X	ল্ল	X	X	X	X	X	X	X
শ্	X	X	শ্ব	X	শ্ম	শ্য <sup>66</sup>	শ্র	শ্ল	X	X	X	X	X	X	X

<sup>61</sup> Variant shape as per traditional orthography is ঙ্

<sup>62</sup> Variant shape as per traditional orthography is ঞ্

<sup>63</sup> This conjunct is needed for writing loan words in Bangla

<sup>64</sup> This conjunct is needed for transcribing Russian text in Bangla

<sup>65</sup> The combination of র্ + য় primarily forms the conjunct র্য but in rare cases it might form the conjunct

র্য (ra + ja phalā) in loan words like র্যাপার particularly to represent the sound /ræ/. This is handled by the use of the Zero Width Joiner U+200D

<sup>66</sup> Variant shape as per traditional orthography is শ্

ষ্	ঞ	ঞ্	ষ	X	ঞ	ষ্য <sup>67</sup>	X	X	X	X	X	X	X	X	X
স্	ঞ	ঞ্	স	X	স্ম	স্য <sup>68</sup>	স্র	স্র	X	X	স্স <sup>69</sup>	X	X	X	X
হ্	X	X	হু <sup>70</sup>	X	হ্ম <sup>71</sup>	হ্য <sup>72</sup>	হ্র	হ্র	X	X	X	X	X	X	X
ড্	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ঢ়্	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
য়্	X	X	X	X	X	য়্য	X	X	X	X	X	X	X	X	X

The CHC + Mātrā/Kār combinations are not listed separately since in most of the cases there is no change in the shape of the consonant cluster when the Mātrā/ Kār is adjoined to it. However, the Paschimbanga Bangla Akademi has suggested ‘transparent variant’ shapes for some of the traditional CHC + Mātrā/Kār combinations. Some of them are listed below.

Traditional Orthography	Transparent Orthography
গ্, গ	গু, গু
ক্, ক	কু, কু
খ্	খু
ঢ়, ঢ়	ঢ়ু, ঢ়ু
ভ্, ভ	ভু, ভু
প্	পু
ব্, ব	বু, বু
জ্, জ	জু, জু
ট্, ট	টু, টু
ড়, ড়	ডু, ডু

<sup>67</sup> Variant shape as per traditional orthography is ঞ

<sup>68</sup> Variant shape as per traditional orthography is স্ম

<sup>69</sup> This conjunct might be needed for transcribing Pali and Prakrit texts in Bangla

<sup>70</sup> Variant shape as per traditional orthography is হু

<sup>71</sup> Variant shape as per traditional orthography is হ্ম

<sup>72</sup> Variant shape as per traditional orthography is হ্য

### 6.3.3.2 CHCHC ( combination of three consonants)

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

The traditional variants are listed as footnotes.

ক + ট + র	ক্ট্র
ক + ত + ব	ক্ভ
ক + ত + র	ক্ভ
ক + ষ্ + ণ	ক্ণ
ক + ষ্ + ব	ক্ণ
ক + ষ্ + ম	ক্ণ
গ্ + ধ্ + র	গ্ধ্র <sup>73</sup>
ঙ + ক + র	ঙক্র <sup>74</sup>
ঙ + ক + ষ	ঙক্ণ
ঙ + গ্ + র	ঙগ্র <sup>75</sup>
ঙ + ঘ্ + র	ঙঘ্র
চ্ + ছ্ + ব	চ্ছ

---

<sup>73</sup> Variant shape as per traditional orthography is ক্ধ্র

<sup>74</sup> Variant shape as per traditional orthography is ক্ণ

<sup>75</sup> Variant shape as per traditional orthography is ক্ণ

চ্ + ছ্ + র	চ্ছ
জ্ + জ্ + ব	জ্জ্ব
জ্ + জ্ + র	জ্জ্র
ণ্ + ড্ + র	ণ্ড্র <sup>76</sup>
ত্ + ত্ + ব	ত্ত্ব
দ্ + দ্ + ব	দ্ব
দ্ + ধ্ + র	দধ্র <sup>77</sup>
দ্ + ভ্ + র	দ্ব
ন্ + ট্ + র	ন্ট্র
ন্ + ড্ + র	ন্ড্র
ন্ + ত্ + ব	ন্ত্ব
ন্ + ত্ + র	ন্ত্র
ন্ + থ্ + র	ন্থ্র <sup>78</sup>
ন্ + দ্ + ব	ন্দ্ব

<sup>76</sup> Variant shape as per traditional orthography is

<sup>77</sup> Variant shape as per traditional orthography is

<sup>78</sup> Variant shape as per traditional orthography is

ন্ + দ্ + র	দ্র
ন্ + ধ্ + র	দ্র <sup>79</sup>
ম্ + প্ + র	ম্র
ম্ + ব্ + র	ব্র
ম্ + ভ্ + র	ভ্র
ল্ + ট্ + র	ল্ট্র
ল্ + ড্ + র	ল্ড্র
ষ্ + ক্ + র	ক্ক <sup>80</sup>
ষ্ + ট্ + র	ট্ট্র
ষ্ + প্ + র	প্প
স্ + ক্ + র	ক্ক <sup>81</sup>
স্ + ট্ + র	ট্ট্র
স্ + ত্ + ব	স্ব
স্ + ত্ + র	স্র

<sup>79</sup> Variant shape as per traditional orthography is

দ্র

<sup>80</sup> Variant shape as per traditional orthography is

ক্ক

<sup>81</sup> Variant shape as per traditional orthography is

ক্ক

স্ + প্ + র	স্প্র
স্ + প্ + ল	স্প্ল
ক্ + ত্ + য	কৃত্য <sup>৪২</sup>
ক্ + র্ + য	ক্র্য <sup>৪৩</sup>
ক্ + ল্ + য	ক্ল্য
ক্ + য্ + য	ক্ষ্য
গ্ + ধ্ + য	গ্ধ্য <sup>৪৪</sup>
গ্ + ন্ + য	গ্ন্য
গ্ + র্ + য	গ্র্য
গ্ + ল্ + য	গ্ল্য
ঘ্ + র্ + য	ঘ্র্য
ঙ্ + ক্ + য	ঙ্ক্য <sup>৪৫</sup>
ঙ্ + খ্ + য	ঙ্খ্য

<sup>৪২</sup> Variant shape as per traditional orthography is কৃত্য

<sup>৪৩</sup> Variant shape as per traditional orthography is ক্র্য

<sup>৪৪</sup> Variant shape as per traditional orthography is গ্ধ্য

<sup>৪৫</sup> Variant shape as per traditional orthography is ঙ্ক্য

ঙ + গ্ + য	ঙ্য <sup>৪৬</sup>
ঙ + ঘ্ + য	ঙ্য
জ্ + জ্ + য	জ্য
ট্ + র্ + য	ট্র্য
ড্ + র্ + য	ড্র্য
ণ্ + ঠ্ + য	ণ্য
ণ্ + ড্ + য	ণ্ড্য <sup>৪৭</sup>
ত্ + ত্ + য	ত্ৰ্য
ত্ + ম্ + য	ত্ম্য
ত্ + র্ + য	ত্র্য
থ্ + র্ + য	থ্র্য
দ্ + ধ্ + য	দধ্য <sup>৪৮</sup>
দ্ + ব্ + য	দ্ব্য
দ্ + র্ + য	দ্র্য

<sup>৪৬</sup> Variant shape as per traditional orthography is



<sup>৪৭</sup> Variant shape as per traditional orthography is



<sup>৪৮</sup> Variant shape as per traditional orthography is





ন্ + ট্ + য	ন্ট্য
ন্ + ড্ + য	ন্ড্য
ন্ + ত্ + য	ন্ত্য
ন্ + থ্ + য	ন্থ্য <sup>89</sup>
ন্ + দ্ + য	ন্দ্য
ন্ + ধ্ + য	ন্ধ্য <sup>90</sup>
ন্ + ন্ + য	ন্ন্য
প্ + প্ + য	প্প্য
প্ + র্ + য	প্র্য
প্ + ল্ + য	প্ল্য
ফ্ + র্ + য	ফ্র্য
ফ্ + ল্ + য	ফ্ল্য
ব্ + দ্ + য	ব্দ্য
ব্ + র্ + য	ব্র্য

---

<sup>89</sup> Variant shape as per traditional orthography is

স্থ্য

<sup>90</sup> Variant shape as per traditional orthography is

ক্ক্য

ব্ + ল্ + য	ব্ল্য
ম্ + প্ + য	ম্প্য
ল্ + ক্ + য	ল্ক্য
ল্ + প্ + য	ল্ল্য
ল্ + ল্ + য	ল্ল্য
ষ্ + ট্ + য	ষ্ট্য
ষ্ + ঠ্ + য	ষ্ঠ্য
ষ্ + ণ্ + য	ষ্ণ্য <sup>৭১</sup>
স্ + ক্ + য	স্ক্য
স্ + ট্ + য	স্ট্য
স্ + ত্ + য	স্ত্য
স্ + থ্ + য	স্থ্য <sup>৭২</sup>
স্ + ন্ + য	স্ন্য
স্ + প্ + য	স্প্য

<sup>৭১</sup> Variant shape as per traditional orthography is ফ্য

<sup>৭২</sup> Variant shape as per traditional orthography is স্থ্য

স্ + ম্ + য	স্ম্য
স্ + র্ + য	স্র্য
স্ + ল্ + য	স্ল্য
র্ + ক্ + য	ক্য
র্ + ক্ + ষ	ক্ষ
র্ + ক্ + স	ক্স
র্ + থ্ + য	থ্য
র্ + গ্ + য	গ্য
র্ + গ্ + র	গ্র
র্ + ঘ্ + য	ঘ্য
র্ + ঙ্ + গ	ঙগ <sup>93</sup>
র্ + চ্ + চ	চ্চ
র্ + চ্ + য	চ্য
র্ + চ্ + ছ	চ্ছ

---

<sup>93</sup> Variant shape as per traditional orthography is ঙ

ৰ্ + ছ্ + য	ছ্য
ৰ্ + জ্ + জ	জ্জ
ৰ্ + জ্ + ঞ্	জ্জ্
ৰ্ + জ্ + য	জ্য
ৰ্ + ট্ + য	ট্য
ৰ্ + ড্ + য	ড্য
ৰ্ + ঢ্ + য	ঢ্য
ৰ্ + ণ্ + য	ণ্য
ৰ্ + ত্ + ত্	ত্ত
ৰ্ + ত্ + য	ত্য
ৰ্ + ত্ + ম	ত্ম
ৰ্ + ত্ + র	ত্ৰ
ৰ্ + থ্ + য	থ্য
ৰ্ + থ্ + র	ত্ৰ
ৰ্ + দ্ + দ	দ্দ

ৰ্ + দ্ + ধ	দধ্ <sup>94</sup>
ৰ্ + দ্ + ব	দ্ব
ৰ্ + দ্ + য	দ্য <sup>95</sup>
ৰ্ + দ্ + র	দ্র
ৰ্ + ধ্ + য	ধ্য
ৰ্ + ধ্ + ব	ধ্ব
ৰ্ + ন্ + য	ন্য <sup>96</sup>
ৰ্ + ন্ + ত	ন্ত
ৰ্ + প্ + য	প্য
ৰ্ + ফ্ + য	ফ্য
ৰ্ + ব্ + য	ব্য
ৰ্ + ভ্ + য	ভ্য
ৰ্ + ম্ + য	ম্য
ৰ্ + ম্ + য	ম্য

<sup>94</sup> Variant shape as per traditional orthography is

ঈ

<sup>95</sup> Variant shape as per traditional orthography is

ঈ

<sup>96</sup> Variant shape as per traditional orthography is

ঈ

ৰ্ + য্ + য	ৰ্য্য
ৰ্ + ল্ + ট	ল্ট
ৰ্ + ল্ + ড	ল্ড
ৰ্ + ল্ + য	ল্য
ৰ্ + শ্ + ব	শ্ব
ৰ্ + শ্ + য	শ্য
ৰ্ + য্ + ট	ট্ট
ৰ্ + য্ + ণ	ষ্ণ <sup>97</sup>
ৰ্ + য্ + য	ষ্য <sup>98</sup>
ৰ্ + স্ + ক	ক্ক
ৰ্ + স্ + ট	ক্ট
ৰ্ + স্ + থ	ক্খ <sup>99</sup>
ৰ্ + স্ + য	স্য
ৰ্ + হ্ + য	হ্য <sup>100</sup>

<sup>97</sup> Variant shape as per traditional orthography is ষ্ণ

<sup>98</sup> Variant shape as per traditional orthography is ষ্য

<sup>99</sup> Variant shape as per traditional orthography is ক্খ

### 6.3.3.3.CHCHCHC ( Combination of four Consonanats)

This is rare and is found only in the following:

ক্ + ষ্ + ম্ + য	ক্ষ্ম্য
ঙ্ + ক্ + ষ + য	ঙক্ষ্য
র্ + জ্ + ঞ্ + য	র্জ্ঞ্য
র্ + ত্ + ত্ + য	র্ত্ভ্য
র্ + ম্ + ম্ + য	র্ম্ম্য
র্ + ষ্ + ণ্ + য	র্ষ্ণ্য <sup>101</sup>
ন্ + ট্ + র্ + য	ন্ট্র্য
ন্ + ত্ + র্ + য	ন্ত্র্য
ন্ + দ্ + র্ + য	ন্দ্র্য
র্ + ব্ + ব্ + য	র্ব্ব্য
স্ + ক্ + র্ + য	স্ক্র্য <sup>102</sup>
স্ + ট্ + র্ + য	স্ট্র্য

---

<sup>100</sup> Variant shape as per traditional orthography is

হ

<sup>101</sup> Variant shape as per traditional orthography is

ক্ষ্য

<sup>102</sup> Variant shape as per traditional orthography is

ক্ষ্য

স্ + প্ + ল্ + য়	অপ্ল্য
-------------------	--------



### 6.3.4 The Collation Order of Bangla.

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.

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

অ    আ    ই    ঈ    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    শ    ষ    স    হ    ড়  
 ঢ়    া    ি    ী    ু    ূ    ্    ে    ৈ    ো    ৌ

In Tabular format:

অ	আ	ই	ঈ	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	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>**