

# Proposal for a Bangla (or Bengali) Script Root Zone Label Generation Ruleset (LGR)

---

*LGR Version:* 4.0

*Current Date:* 2020-05-20

*Document version:* 5

*Authors:* Neo-Brahmi Generation Panel [NBGP]

## 1. General Information

This document lays down the Label Generation Rule Set (LGR) for the Bangla (or ‘Bengali’)<sup>1</sup> script under the general rubric of the Neo-Brāhmī Writing System. Three main components of the Bangla Script LGR i.e. (i) Code point repertoire, (ii) Variants and (iii) Whole Label Evaluation Rules which have been described in detail here, having given a brief historical background of the Script under Section 3.

All these components will be incorporated in a machine-readable format in an XML file named "proposal-bengali-lgr-20mar20-en.xml". Labels for testing can be found in the accompanying text document "bangla-test-labels-20mar20-en.txt".

## 2. Script for Which the LGR Is Proposed

ISO 15924 Code: Beng

ISO 15924 Key N°: 325

ISO 15924 English Name: Bengali (Bangla)

Latin transliteration of native script names [in IPA]: ba:ŋla:, ôxômiya

Native names of the script: বাংলা, অসমীয়া

Maximal Starting Repertoire (MSR) version : MSR-4

## 3. Background on Script & Principal Languages Using It

### 3.0. Introduction

‘Bangla’ (or Bengali) is historically and genealogically regarded as an eastern Indo-Aryan language with around 178.2 million speakers in Bangladesh (98% speakers), and 83.4 million speakers in the Indian states of West Bengal (68.37 million), Tripura (2.15 million), South Assam (7.3 million), Odisha (0.49 million) and Delhi (0.21 million) as

---

<sup>1</sup> The term ‘Bangla’ is used in the descriptive text and the term ‘Bengali’ is used in the normative part of this proposal.

well as in the Andaman and Nicobar Islands (close to a hundred thousand) - accounting for 8.3% of India. It is a major language in Jharkhand (2.6 million), too and a language with a sizable population in Bihar (0.44 million). Apart from these, there are a huge number of Bangla-speaking diasporas spread all over the world. It is the seventh largest spoken and written language in the world. Bangla is the national and official language of Bangladesh, and one of the 22 Official languages in India (listed in the 8<sup>th</sup> Schedule of the Indian Constitution). It is also one of the official languages of Sierra Leone. The script is also called Bangla [102], which is an eastern variety of the 'Brāhmī' Writing System, written from left to right. Historically it derives from the Brāhmī alphabet as used in the Ashokan inscriptions (269-232 BC).

Bangla and its cognate languages, as mentioned above, together form a linguistic group known as the Eastern New Indo-Aryan (NIA). There is a gross inadequacy of the inscriptions and manuscripts in the Eastern Apabhraṃśa or 'Avahaṭṭha' except for small inscriptions and the manuscripts of the Tantric Buddhist text titled '*Caryyācaryyaviniścaya*' or the *Caryā-Pada* [114] dating back to the 9th-11th century. As a result, there is not much epigraphic evidence for the development of its writing system. However, what evidence is available of the genesis of Bangla writing system is discussed in the section 3.1 [109].

Historically, the Bangla language is divided into three periods as evident from various sources:

- (i) Firstly, Old Bangla Period (roughly 950/1000 to A.D.1200/1350) of which three specimens are found: (a) 47 Caryā songs, the *Dohākōṣa* of Saraha and the *Dohākōṣa* of Kānha (mostly in Apabhraṃśa), and the *Ḍākārṇava* (in a variety of Prākṛt), (b) Old Bangla specimens of over 300 words in a commentary [141].
- (ii) Then there is Middle Bangla Period - 1200-1800 AD, again divided into three stages: (a) Transitional Middle Bangla (1200-1300 A.D, for which no genuine specimens are found) [147], (b) Early Middle Bangla (1300-1500 A.D), and (c) Late Middle Bangla (1500-1800 A.D).
- (iii) Finally, after 1800 AD, we find the Modern or New Bangla, marked by the introduction of written prose [109] in the books of Fort William College (established in 1800) The colloquial variety of Bangla based on the speech variety of Calcutta (called 'Kolkata' now) made its first appearance through the *Hutōm Pēcāra Nakśā* (1862) by Peari Chand Mitra. The influence of English in the vocabulary, idioms, and expressions as well as in the writing styles of Bangla is significant by this time. The fonts and types for Bangla developed during this time also spread to all parts of Bangla speech community [101, 120]. The same fonts with some extensions were also used for the neighbouring languages deploying this writing system.

Bangla prose had developed two literary styles during the 19th-20th Century: The *Sādhubhāṣā* (সাধুভাষা - "Elegant Language or Style") and the *Calitabhāṣā* (চলিতভাষা "Current Language, or Modern Style"). It is the latter style that is prevalent today in written prose.

The Language Movement in Bangladesh (the then East Pakistan) began in 1948, as civil society dissented to the elimination of the Bangla script from currency and stamps, which were in use since the British Raj. The movement reached its pinnacle in 1952, when on 21 February the police fired on demonstrating students and civilians, triggering numerous injuries and deaths<sup>2</sup>. Later, following the Language movement, on 27 April 1952, the All Party National Language Committee decided to demand establishment of an organization for the promotion of Bengali language. Bangla Academy, Dhaka right from its inception in 1955 has been engaged in promoting and fostering Bangla as the *lingua franca* of the country before and after independence from Pakistan in 1971. Through the various commissions and committees constituted by the Government of Bangladesh (Bāṅlādeśa Jātīya Śikṣā Kāmīśana in 1972, Jātīya Śikṣā Upadeṣṭā Pariṣad in 1979, Bāṅlā Bhāṣā Bāstabāyana Sela in 1982, Bāṅlā Bhāṣā Kāmīṭi in 1983, etc.<sup>3</sup>) after independence in 1971 Bangla was made the primary medium of instruction/communication in all Governmental and educational activities. Through a great struggle and bloodshed, the Bengalis established Bangla as an official language of the state.<sup>4</sup>

### 3.1. Written Bangla

The 'Bangla alphabet' (বাংলা লিপি - Bāṅlā lipi, ISO 15924) is derived from the Brāhmī writing system, which is related to the Nāgarī (*also known as Devanāgarī*<sup>5</sup>) script [108] as well as to Tirhutā writing system [106]. Considered to be fifth most widely used writing system in the world, this combined Bangla-Asamiyā-Manipuri Script (showing some variations for Asamiyā and Meitei or Biṣṇupriyā Manipuri) (130), was used in the eastern Indian Sanskrit manuscripts too. For Chākmā in India and Bangladesh and for Kokborok in Tripurā, it was and still is one of the scripts used. A close variant, called *Tirhutā* (123; now available also in UNICODE 10.0 as 11480 114DF; See 110) or

---

<sup>2</sup> The UN declared Ekuśe February (21st February) as the International Mother Language Day at the UNESCO General Conference in Paris on 17 November 1999 "in recognition of the sanctity and preservation of all vernacular languages in the world."<sup>22</sup>

<sup>3</sup> Bāṅlā Bhāṣā Kāmīṭi. 1983. *Bāṅlā Bhāṣā Kāmīṭi Riport* (Report of the Bangla Bhasha Committee). Dhaka: Śikṣā, Dharma, Krīrā O Saṅskṛti Mantraṇālaya, Peoples Republic of Bangladesh.

<sup>4</sup> Chakraborty, Rajib. 2018. *The Fishermen's Community: A Language-Culture Interplay (A Study of Post-1971 Select Bangla Novels)*. Unpublished Ph.D. Dissertation, Visva-Bharati.

<sup>5</sup> William Dwight Whitney in his *Sanskrit Grammar* unequivocally said, "This name (Devanāgarī) is of doubtful origin and value" (Whitney, William Dwight. 1994 reprint. *Sanskrit Grammar*. New Delhi: Motilal Banarasiidass Publishers, p. 1)

*Mithilākṣara* was used for Maithili from the 14th Century until the early-20th century [106]. In this context, one finds a mention of ‘Sylheti Nāgarī lipi’ or ‘Siloṭi’ (added to the Unicode Standard in March 2005 with the release of version 4.1) the details of which could be of interest only to historians and historical linguists (See 137 and 144). But Sylheti Bangla is generally written by many in the modern-day Bangla script now for all practical purposes. Originally, during the reign of the Pāla dynasty (750-1154 AD) in the eastern India, and even earlier, perhaps during the Malla period (694 AD onwards), the present-day Bangla writing system got a shape comparable to the modern-day ones [111, 119]. A pictorial description of Brāhmī to Modern Bangla Script could be presented here in a tabular form:

300 BCE	†	ε	୪	।	୮	୩
200 CE	†	Ε	୪	।	୮	୩
400 CE	†	Ε	୮	।	୮	୩
600 CE	†	Ε	୮	।	୮	୩
800 CE	†	Ε	୮	।	୮	୩
900 CE	†	Ε	୮	।	୮	୩
1100 CE	†	Ε	୮	।	୮	୩
1300 CE	†	Ε	୮	।	୮	୩
Modern	କ	ଜ	ମ	ର	ସ	ଅ
	k	j	m	r	s	a

Table 1: Pictorial depiction of Evolution of Brāhmī to Bangla

The inscripational evidence in Brāhmī is found in the Archaic Brāhmī from the 3rd century B.C. to the 1st century B.C, and in Middle Brāhmī – soon after (1<sup>st</sup>-3<sup>rd</sup> Century A.D.) and then on in the Late Brāhmī (4<sup>th</sup>-6<sup>th</sup> Century A.D.). This evidence could be seen in both Bangladesh and West Bengal [108] by 1) The Mahāsthānagaṛa (Bogra district, Bangladesh — the ancient name being Puṇḍranagara or Pauṇḍravardhanapura) inscriptions, 2) Brāhmī (and Kharoṣṭhī) inscriptions from the lower ‘Gangetic Bengal’ and (3) Copper plate inscriptions of the Imperial Guptas from Northern part of West Bengal and North-West Bangladesh — in the areas under Dharmāditya, Gopachandra and Samācāradeva (about whom one only knows from five Copper-plates found in Kotālipārā in the Faridpur district in Bangladesh, one in Mallasārul in the Burdwan district (West Bengal), and one in Jayrāmapura (Balleśvara district, now in Odisha).

These epigraphs from the eastern part of Undivided India (dating back to the 4th-6th Centuries A.D.) showed some characteristic features of letters (especially in म 'ma', ल 'la', स 'sa', श 'sa' and ह 'ha'), which led to the development of eastern variety of Gupta script. Epigraphic records from Bangladesh demonstrate remarkable developments in Eastern Brāhmī. In this context, the Tippera copper plate inscription of the 'Samatāṭa' rulers (139, pp 265) such as Lokanātha (dated 7<sup>th</sup> Century A.D., during the latter half), the Kailan inscription of Śrīdharaṇa Rāta as well as the Astafpur copper plates. The letters seem to hang down from wedge shaped solid triangles with right hand verticals bending down at the bottom, because of which it was described by Prinsep and Fleet as *Kuṭīla-lipi* (literally, 'Cursive writing style'), whereas the term *Siddhamātrikā* (as a mātrā or bar is placed over each of the letters) was used by Al Biruni (973-1048) to designate the script of Northern India. The next stage of development is illustrated by the 9th Century copper plate inscriptions from Khalimpur of the reign of Dharmapāla, from Monghyr and Nālandā of the time of Devapāla in Bihar, and from Jagjīvanpura (Malda) of the reign of Mahendrapāla. The Siddhamātrikā (mentioned as 'Siddham' in Chinese sources) is said to have been prevalent also in this region up to the end of the tenth century. Also called the Gauri (i.e. Gandhi) in Pūrvadeśā or the Eastern country, it was regarded as the same script to which is given the appellative Proto-Bangla characteristics in rudimentary forms, in the period between A.D. 875 and A.D. 1025.

In some epigraphs it is considered as belonging to the second quarter of the eleventh century A.D. Flattening of head-marks becomes prominent in comparison to the wedge-shaped serifs. An important landmark in the development of the Bangla script is the Rāmagañja copper plate inscription of Mahāmāṇḍalika in the last quarter of the eleventh century A.D. It is the earliest document from this entire region which bears the letter m, with a tick rising upwards. The full vowel i develops a tick at the right end of the upper horizontal bar above and a curved hook below. Initial e approaches the modern Bangla character. A mature form of Proto-Bangla, the immediate precursor of Bangla script, is illustrated in the inscriptions of the Varmaṇa, Sena and Deva rulers of the twelfth and thirteenth centuries [104].

The evolution of the Bangla script (Cf. 136) is aligned with the story of advancement of printing technology. The first "Movable type" scripts technically created and used while printing Nathaniel Brassey Halhed's (1751-1830) 1778-book titled, '*A Grammar of the Bengal Language*'. In 1785, Governor-General Warren Hastings (1732-1818) requested another civilian, Charles Wilkins (1749-1836) to cut punches for Bangla printing characters. The current printed form of Bangla script appeared soon after. It is generally agreed that Wilkins developed Bangla print script [111]. He passed on this knowledge to Pañcānana Karmakāra (?-1804), a renowned artist in Bengal. Later it was Karmakar and his family that became famous in Bangla printing technology. Shepherd was another assistant of Wilkins in this designing of script, which became more angular with sharper turns and edges [133]. A few archaic letters were modernized during the 19th

century. It was standardized by Pandit Ishwar Chandra Vidyasagar when the Bangla type fonts were to be used to publish on a large scale under the Calcutta School Book Society [116 for several references].

Much later, in 1935, the Linotype technique, invented by Ottmar Mergenthaler (1854-1899) in 1886, was introduced into Bangla printing in 1935, by the efforts of Suresh Chandra Majumdar (1888-1954), Rajsekhar Basu (1880-1960), Jatindra Kumar Sen (1882-1966) and his disciple, Sushil Kumar Bhattacharya and had begun being used by the Ānandabāzara Patrikā group, later followed by others. Within a few years the more advanced monotype technology came to be used in Bangla printing. However, in Bangla printing culture, monotype has a very limited acceptance and linotype held stage till, eventually, the digital technology came in to replace all earlier techniques.

All these could be presented in a table:

PERIOD	DESCRIPTION	NAMES
3 <sup>rd</sup> Century B.C.	Use of Brāhmī and Kharoṣṭhī scripts begin in the subcontinent. Brāhmī was widely used during the Mauryan King, Aśoka. In one theory, Brāhmī is based on North Semitic alphabet but suitably modified to fit the need of local languages. It is currently believed to have been an independent development.	Brāhmī
1 <sup>st</sup> -3 <sup>rd</sup> Century AD	The Kuṣāṇa script, named after the Kuṣāṇa royal dynasty.	Kuṣāṇa script
4 <sup>th</sup> -5 <sup>th</sup> Century AD	The next stage of its evolution was into the Gupta script, named after the Gupta royal dynasty.	Gupta script
7 <sup>th</sup> Century AD	Epigraphic records from Bangladesh demonstrate remarkable developments in Eastern Brāhmī, giving rise to the <i>Kuṭila-lipi</i>	Kuṭila-lipi
8 <sup>th</sup> Century AD	Some copper plate inscriptions are found in the Khalimpur, Bangladesh during the reign of Dharmapāla, from Monghyr and Nālandā in Bihar, of the time of Devapāla, and from Jagjīvanapura in West Bengal of the reign of Mahendrapāla.	<i>Siddhamātikā</i>

PERIOD	DESCRIPTION	NAMES
9th Century AD until 1025 AD	Proto-Bangla characteristics in rudimentary forms develop. An important landmark in the development of the Bangla script is the Rāmagañja copper plate inscription of Mahāmāṇḍalika found in the last quarter of the eleventh century A.D.	Proto-Bangla Script & Language
12 <sup>th</sup> -13 <sup>th</sup> Century AD	A mature form of Proto-Bangla, the immediate precursor of Bangla script, is found in the inscriptions of the Varmaṇa, Sena and Deva rulers of the twelfth and thirteenth centuries.	Matured Proto-Bangla
14 <sup>th</sup> -15 <sup>th</sup> Century AD	The characteristics of typical Bangla script began to develop, as could be seen in the copper plate inscription of Vijayamāṇikya-I of Tripura dated 1478 AD - also illustrates forms of Bangla letters in the fifteenth century A.D.	Modern Bangla Script era begins (See Ross 1999)
16 <sup>th</sup> -17 <sup>th</sup> Century AD	The chart of the Bangla alphabet, appended to the China Monuments, published from Amsterdam in 1667 and <i>The code of Gentoo law</i> , published from London in 1776, both show a chart of the Bangla alphabet. They show 16 Vowel letters, including the Long ‘ঔ’ ‘ঐ’, Anusvāra and <i>Visarga</i> , and 34 Consonants.	Printed Charts of Bangla
18 <sup>th</sup> -19 <sup>th</sup> Century AD	Charles Wilkins develops printing in Bangla in 1778 and Vidyasagar reforms it.	Bangla Type Fonts

Table 2: Development of the Bangla Writing System

The overall development of Bangla Script from the Kuṭila-lipi period to Modern Bangla could be seen here in Table 3 ([102 and 146] and also see the web-page in 147).



EGIDS Scale 1	EGIDS Scale 2	EGIDS Scale 3	EGIDS Scale 4	EGIDS Scale 5	EGIDS 6
Bangla (Bengali)				Santali, Bodo, Riang, Khumi, Mru(ng), Asho	Lepcha Pnar, Koda/ Kora, Chak
	Asamiyā (Assamese)			Koch or Rājabaṅśī	Mālto or Mālpāhāriyā
	Maṇipuri or Meitei		Biṣṇupriyā Maṇipuri, Kok-Borok (Tripura & Bangladesh)	Chākmā, Hājong, Muṇḍāri & Kurux (of Bangladesh)	Toto, Rohingyā, Tippera, Megam, Tanchangya
			Usoi	Limbu, Sadri or Oraon	Bhumij or Muṇḍāri, Bawm, Chin

Table 4: Main languages in India and Bangladesh  
that use Bangla Script on the EGIDS Scale

### 3.3. Notable Features of Bangla Script [150]

Bangla Writing System has certain features that show how it has to be written in or how type-setting in Bangla could be done. This section is followed by a section that explains the Code-points (and fixed Code-point sequences) which show certain distinctive characteristics of Bangla and which make the Repertoire. The next sections will also cover the ‘akshar’-formation rules (ABNF) showing character class, Word Level Evaluation (WLE) and Context Rules as well as In-Script and Cross-Script Variants. Here, we present some basic features of the Script and Pronunciation:

- The Bangla script is an alpha-syllabic writing system in which writing of all consonants are assumed to contain an accompanying ‘inherent’ vowel (theoretically before or after each consonant). It varies between /ɔ/ and /o/ depending on the position of the consonant in the word. At times, these ‘assumed’ or ‘inherent’ vowels are not pronounced at all [142].

- Vowels can be written as independent letters, or by using a variety of diacritical marks which are written above, below, before, after or both of the last two positions the consonant they follow in pronunciation [105].
- All Bangla consonants when pronounced in isolation are uttered with an inherent vowel - / ɔ/; hence ক 'k', খ 'kh' or গ 'g' are usually pronounced as [kɔ], [khɔ], or [gɔ], etc. Phonologically, Bangla vowel - / ɔ/ corresponds to the Hindi schwa /ə/
- When consonants occur together in clusters, special conjunct letters are formed. In printed Bangla, many of these consonantal clusters or conjoined consonants are in use. The letters for the consonants other than the final one in the group are generally reduced. But there are a few special conjunct characters which are compounds of the consonant characters, e.g. ক্(k)+ষ(ṣ)=ক্ষ(kṣ)/, ঞ্(ñ)+জ(j)=ঞ্জ(ñj), জ্(j)+ঞ(ṇ)=জ্ঞ(jṇ), হ্(h)+ম(m)=হ্ম(hm). There are other issues also—র as the second member of a cluster is reduced to a secondary symbol, e.g. প্(p)+র(r)=প্র(pr), ষ্(ṣ)+ট্(t)+র(r)=ষ্ট্র(ṣṭr) (as in উষ্ট্র uṣṭra “camel”); য (y), when used as a primary symbol, represents /jɔ/ in Bangla. But its secondary symbol (allograph) য়-phalā has two phonetic values. When added to the initial consonant in a word, it is a vowel /æ/ (as in শ্যামল (śyāmala) “green”, র্যাপার (ryāpāra) “wrapper”, etc.). But after a non-initial consonant, it just doubles it in pronunciation (as in কার্য, ধার্য, etc.). The র্(r)+য(y) combination has two renderings—র্য(ry) and র্য(ry). In case of দ্(d)+ধ(dh), গ্(g)+ধ(dh), ন্(n)+ধ(dh) the shape of the second member is changed—e.g. দ্ধ(ddh), গ্ধ(gdh), and ন্ধ(ndh) respectively. The solitary example of র্(r)+ঋ(ṛ)=ঋ (as in নৈঋত nairṛt “Southwest”) – used mostly in cases of Classical borrowings, shows the use of secondary symbol of a consonant followed by the primary symbol of a vowel. The inherent vowel only applies to the final consonant of the cluster.
- In consonant clusters, many consonants took a completely different form. Some typical examples are ক্ (kt), ক্র (kr), ক্‌ (kṣ), গ্‌ (gdh), জ্‌ (jñ), ঞ্‌ (ñc), ঞ্‌ (ñj), ট্‌ (tṭ), ত্‌ (nt), ন্‌ (ndh), ব্‌ (bdh), ভ্‌ (bhr), ষ্‌ (mb), স্‌ (st) etc. র has two allographs, apart from this full shape : one is ‘repha’, as found in র্‌ (rk), র্‌ (rp); and another is ra- phalā, as in প্র (pr), ক্র (kr). ঞ্‌ (ṣ+ṇ) is another one, where the cerebral nasal consonant sign takes a queer shape. [151]
- The Bangla script has at least fifty-two primary symbols and quite a few allographs (positional variants of them), corresponding to forty-four (7 oral and 7 nasal vowels and 30 consonants) phonemes (150) or functional speech sounds, with some obvious redundancies, although in one of the first phonemic analysis, the number was thought to be thirty-five phonemes [140].

- As mentioned above, in Bangla, several graphemic symbols have secondary shapes, technically called ‘allographs’ with a complementary distribution in each case. These graphs or markings are generally added to the following positions of the primary symbol [113] in the following manner:
  - 1) Below (e.g. কু (ku), ন্ত (nta), কূ (kū), হ্র (hra), etc.)
  - 2) Above (e.g. চ্চ (cā), র্ক (rka), etc.)
  - 3) Right side (e.g. কা (kā), কং (kaṅ), etc.)
  - 4) Left side (e.g. কে (ke))
  - 5) Left Side and above simultaneously (e.g. কৈ (kai), কি (ki) etc.)
  - 6) Right side and above simultaneously (e.g. কৌ (kī))
  - 7) Right side and left side simultaneously (e.g. কো (ko))
  - 8) Right side, left side and above simultaneously (e.g. কৌ (kau)).
- As for complementary distribution of vowel letters (word- or syllable-initial) and Vowel Mātrās, which are relevant for ABNF, let us consider the following. Besides some simple Vowel Modifiers called ‘Kārs’ in Bangla (also referred to as Mātrā in the other LGR documents of Neo-Brāhmī) there are some combinatory modifiers of Bangla Vowels with certain consonants. For example, whereas

আ U+0986 BENGALI LETTER AA is substituted by

া U+09BE BENGALI VOWEL SIGN AA,

ই U+0987 BENGALI LETTER I is substituted by

pre-posed ি U+09BF BENGALI VOWEL SIGN I,

ঐ U+0988 BENGALI LETTER II is substituted by

ী U+09C0 BENGALI VOWEL SIGN II or

উ U+0989 BENGALI LETTER U is substituted by

ূ U+09C1 BENGALI VOWEL SIGN U by marking below the primary grapheme, there are some special vowel modifiers of উ as in the following combined letters:

গু gu, rather than writing as গ (g) + ূ (u)

রু ru, rather than writing as র (r) + ূ (u)

সু su, rather than writing as শ (ś) + ূ (u)

হু hu, rather than writing as হ (h) + ূ (u)

ন্তু/তু ntu, rather than writing as ন্ (n) + ত (t) + ূ (u)

Similarly, there could be vowel modifiers of উ or ‘(Long) ū’ as well; e.g.

ভ্ (bh) + র (r) (ভ্র bhrū “eyebrow”), শ্ (ś) + র (r) (শ্র śrū), ঋ (ṛ) after হ (h) (হ্র hr), etc.

- There have been many notable contributions in simplifying and modifying Bangla spellings and combinatory techniques, especially by scholars such as Pabitra Sarkar (1992) [134]. In this there has been an attempt to reduce the number of allographs of both vowels and consonants in clusters, and it has been widely accepted in the printing of school texts in both Bangladesh and West Bengal [151, 152]. As of now, two systems, the old (traditional), and the new, go on side by side, operative in different domains.

However, in preparation of this LGR document, the aim has been to consider the widely used and usable sequences and combinations and their variations across the sister scripts belonging to the basket of Brāhmī writing systems.

Bangla Academy, Dhaka published Standard Bangla Spelling Rules in 1992 following the recommendations of a committee constituted through a workshop jointly organized by the Jātīya Śikṣākrama and Pāṭhyapustaka Board in 1988. A thoroughly revised edition of the Rules was published in September 2012.<sup>6</sup>

After the establishment of Bāñlā Ākāḍemi of West Bengal in 1986, its first President, Annadasankar Ray (1904-2002), in his inaugural address, gave a direction for standardization of Bangla alphabet, script, the spelling system and clearly argued that they would not blindly follow the Sanskritic model of conventional grammar. A broad list of proposals was sent to experts on Bangla, and a broad agreement was reached for ‘homogenization of Bangla spelling’ by 1988. Based on opinions received from different quarters, a unanimous list of ‘rules’ was agreed upon. This was published by a ‘Spelling Dictionary’ titled, *Ākāḍemi Bāñāna Abhidhāna* (1997), which was obviously more comprehensive than ‘The University of Calcutta proposals’, made in 1936. Along with the ‘rationalization’ of spellings, another step was taken to make the writing system easier to read, by making the symbols used, both single and combined ones, more ‘transparent’. These reforms were originally suggested by Sarkar (1987, first published in 1978) [134] [153] where he used the terms *Swaccha* (‘Transparent’) and *Aswaccha* (‘Opaque’ or non-transparent), even adding *Ardha Swaccha* (‘half transparent’) in between the two. Some sample examples are:

**Transparent:** ণ (nn), প্ত (pt), স্ত (st), where both member of the cluster can be recognized.

---

<sup>6</sup> Bangla Academy. 2012. *Bāñlā Ekaḍemī Pramita Bāñlā Bāñānera Niyama* (Bangla Academy Standard Bangla Spelling Rules). Dhaka: Bangla Academy.

**Opaque:** where neither of the two could be (easily) recognized—ক্ষ ks (ক্ k + ষ ষ), জ্ঞ jñ (জ্ j + ঞ্ ñ), ঙ্ ng (ঙ্ ñ + গ্ g), হ্ম hm (হ্ h + ম্ m).

**Semi-transparent:** প্র (pr), র্প (rp) where one symbol is recognizable and the other is not. In case of three-term clusters, at least one symbol will not be transparent, e.g. স্ট্র str (স্ s+ত্ t+র r), স্ট্র (স্ s+ট্ ট্+র r), etc.

There were, in fact, two types of proposals. One concerned the shape of the letters, those of consonant + vowel (CV) combinations and conjuncts, which is consonant + consonant combinations. There were further complex shapes, i.e. those of consonant + consonant+ (consonant+) vowel (CC(CV) signs, as in প্রু (pru), or স্ক্রু (skru). Some decisions in this area were necessary because a few of the CC(C) symbols represented complexities that made learning them difficult for the children. The other dealt with the spellings of words only, without any reference to the shapes of letters in which they were written. The basic objective here was 'one word, one spelling', to the greatest extent that was possible. [151]

Below we place a statement of the most salient changes that affect the consonant + vowel combinations. [153]

- The variants of the short u (হ্রস্ব উ-কার hrasva u-kāra) vowel sign have been brought down to one, i.e., ঁ. So গু (gu) is now গু. Similarly রু (ru) > রু, শু (śu) > শু, হু (hu) > হু and therefore, cluster + short u sign : ন্তু (ntu) > ত্তু (ন+ত্+উ), স্তু (stu) > স্তু (স+ত্+উ)
- The variants of long u (দীর্ঘ উ-কার dirgha u-kāra) have also been reduced. রূ (rū) > রু; ব্রূ (bhrū) > ব্রু (ভ bh+র r+উ ū); দ্রূ (drū) > দ্রু (দ d+র r+উ ū); শ্রূ (śrū) > শ্রু (শ ś+র r+উ ū)
- The variants of ঝ-কার (ṛ-kāra "secondary symbol of ṛ") have been brought down to one: হ্র (hr) > হ্

Regarding consonant + consonant + (consonant)...+ (vowel) clusters Paschimbanga Bangla Akademi proposed transparent or semi-transparent shapes for clusters to the extent admissible in Bangla writing system. Some examples will clarify the proposal (A slash will mean that the traditional cluster-shape precedes it, while the Bangla Akademi innovation follows.) [153]

ব্ধ bdh (ব্ b+ধ dh), দ্ধ ddh (দ্ d+ধ dh), ত্ব/ত্থ, ত্থ nth (ন্ n+ত্থ th), ধ্/ধ্ধ, ধ্ধ ñc (ঞ ñ+চ c), ঙ্/ঞ্, ঙ্ ঙ্চ (ঞ্+ছ), ঙ্/ঞ্, ঙ্ ñj (ঞ্ ñ+জ j), ক্ত/ক্, ক্ত kt (ক্ k+ত t), ক্র/ক্, ক্র kr (ক্ k+র r), গ্ধ/গ্, গ্ধ gdh (গ্ g+ধ dh), ক্/ক্, ক্ ñk (ঙ্ ñ+ক k), ঙ্/গ্, ঙ্ ñg (ঙ্ ñ+গ g), ষ্/স্, ষ্ ষ্ণ (ষ্ ষ্+ণ ñ), ক্/ন্থ, ক্ ndhr (ন্ n+ধ্ dh+র r), ঙ্/ঙ্, ঙ্ ñdr (ঙ্ ñ+ড্ d+র r), ক্/ক্, ক্ ktr (ক্ k+ত্ t+র r)

### 3.3.1 The Consonants

As per traditional classification Bangla Consonants are categorized according to their phonetic properties, especially in terms of place and manner of articulation [107]. There are Five ‘Varga’ (pronounced as ‘Barga’ in Bangla) or Groups (sets or classes) distinguished by Place of Articulation, and one Non-‘varga’ group [105]. Each Varga, which corresponds to Stops at a certain place of articulation, contains a series of five consonants classified as per their phonetic qualities (i.e. manner of articulation), beginning from Unvoiced and Unaspirated to Voiced and Aspirated (in the fourth column), finally ending with a Homorganic or Corresponding nasal [107]. Consider the following table:

‘Varga’ or Sets	Unvoiced		Voiced		Nasal
	-Asp	+Asp	-Asp	+Asp	
<b>Velar</b>	ক ‘K’ U+0995	খ ‘KH’ U+0996	গ ‘G’ U+0997	ঘ ‘GH’ U+0998	ঙ ‘NG’ U+0999
<b>Palatal</b>	চ ‘C’ U+099A	ছ ‘CH’ U+099B	জ ‘J’ U+099C	ঝ ‘JH’ U+099D	ঞ ‘NY’ U+099E
<b>Retroflex</b>	ট ‘TT’ U+099F	ঠ ‘TTH’ U+09A0	ড ‘DD’ U+09A1	ঢ ‘DDH’ U+09A2	ণ ‘NN’ U+09A3
<b>Dental</b>	ত ‘T’ U+09A4	থ ‘TH’ U+09A5	দ ‘D’ U+09A6	ধ ‘DH’ U+09A7	ন ‘N’ U+09A8
<b>Bilabial</b>	প ‘P’ U+09AA	ফ ‘PH’ U+09AB	ব ‘B’ U+09AC	ভ ‘BH’ U+09AD	ম ‘M’ U+09AE

Table 5: Varga classification of Bangla consonants  
(Falling into a Pattern of Five Sets of Unvoiced Unaspirated, Unvoiced Aspirated, Voiced Unaspirated, Voiced Aspirated and Nasals, called five ‘Varga’)

<b>Non-Varga</b>	য ‘Y’ U+09AF	য় ‘YY’ U+09DF	র ‘R’ U+09B0	ড় ‘RR’ U+09DC	ঢ় ‘RH’ U+09DD
	ল ‘L’ U+09B2	শ ‘SH’ U+09B6	ষ ‘SS’ U+09B7	স ‘S’ U+09B8	হ ‘H’ U+0939

Table 6: Non-Varga consonants (Not falling into any of the five categories)

### 3.3.2 The Implicit Vowel Killer: Hasanta (called 'Halant' or 'Halanta' in other Brāhmī-based scripts)

As stated earlier, all consonants are pronounced in isolation with an implicit vowel (central back /-ɔ/ in Bangla as the neutral vowel) assumed to be associated with them [121]. The 'Hasanta' (= 'Halant' or 'Halanta' in other Brāhmī-based scripts) or the term 'Virāma'<sup>7</sup> (= 'Dāri' in Bangla) as preferred in UNICODE (cf. Unicode 3.0 and above) have been used in this report as terms that have been used to denote the character that mark the absence of this inherent vowel. It may be noted that the term *virāma* has been adopted in UNICODE in a sense that is different from the traditional definition of grammar, and hence it requires some explanation here. Considering the importance of the document this note should be a part of this LGR document, so that anybody referring to it should be able to know the proper grammatical explanation of the term. Because a special sign is needed whenever this implicit vowel is stripped off, the symbol is known as the *Hasanta* (= *Halant*) "◌̣" (U+09CD). By placing the *Hasanta* under the first consonant of a combination or cluster, one could – in common parlance, "kill" its vowel, and create conjuncts. In this manner, conjunct characters can be generally written by joining two to four consonant combinations. In rare cases, this process can join up to five consonants. However, the notion of a maximum number of consonants joining to form one *akṣara*<sup>8</sup> is to be bounded empirically. This is an observation based on the CIIL-Emille Corpora of Bangla words [132 & 133] as seen in print these days. Given the mixture of scripts and languages happening on the web, the possibility that one may want a generic Top Level Domain [gTLD] which may have more than the observed maximum cannot be ruled out. This can be the case when a foreign language word, which admits a large number of consonants, is transliterated into Bangla. Hence, in the Bangla LGR work, this limit will not be enforced.

### 3.3.3 Vowels

Separate symbols exist for all 'Swara' or Vowels in Bangla, which are pronounced independently either at the beginning of the word or after another vowel or consonant sound. To indicate a Vowel sound other than the implicit one, a Vowel sign, called 'kāṛ' in Bangla or *Mātrā* in Nāgarī<sup>9</sup> is attached to the consonant. Since the consonant has this built in neutral vowel at the end, there are equivalent *kāras* (*Mātrās*) for all vowels except the अ (pronounced /-ɔ/). The correlation is shown as follows:

---

<sup>7</sup> *Virāma*, as used here, is also a misnomer according to the Indian grammatical traditions. No where mere absence of a vowel is marked as *virāma*. *Hasanta* just marks the absence of a vowel, nothing else. (Abhyankar, Kashinath Vasudev & J. M. Shukla. 1961. *A Dictionary of Sanskrit Grammar*. Baroda: Oriental Institute.)

<sup>8</sup> This term needs to be disambiguated. *Akṣara* also means 'syllable' in Indian grammatical traditions

<sup>9</sup> Although the term 'Mātrā' in Bangla stands for an altogether different concept, viz. the top bar placed over a letter – typically available in Hindi and Bangla but missing in Gujarati.

Vowel	Corresponding vowel sign (kāras (Mātrās))
অ 'A' U+0985	
আ 'AA' U+0986	া U+09BE
ই 'I' U+0987	ি U+09BF
ঈ 'II' U+0988	ী U+09C0
উ 'U' U+0989	ূ U+09C1
ঊ 'UU' U+098A	্ৰ U+09C2
ঋ Vocalic 'R' U+098B	্র U+09C3
ঌ Vocalic 'RR' U+09E0	্র U+09C4
঍ Vocalic 'L' U+098C	্র U+09E2
ঞ Vocalic 'LL' U+09E1	্র U+09E3
এ 'E' U+098F	ে U+09C7
ঐ 'AI' U+0990	ৈ U+09C8
ও 'O' U+0993	ৌ U+09CB
ঔ 'AU' U+0994	ৌ U+09CC
-	ী U+09D7
Could appear on top of অ 'A' U+0985 or	ঁ U+0981 Candrabindu
Could appear after অ 'A' U+0985 or	ং U+0982 Anusvāra
Could appear after অ 'A' U+0985	ঃ U+0983 Visarga
After any consonant	্ U+09CD (Hasanta )
-	৲ U+09BD Avagraha

Table 7: Bangla Vowels with corresponding **kārs**

### 3.3.4 The Anusvāra /onuffār/ (◌ং - U+0982)

The Anusvāra or /onuffār/ in Bangla at times represents a homorganic nasal but not always. It replaces a conjunct group of a ‘Nasal Consonant+Hasanta +Consonant’ where the second consonant belongs to the Velar *varga* or set as in লংকা. But it often appears also for such combinations involving non-velars appearing as the last member of the combination as in ল্যাংটা “naked”, or ল্যাংচা “a kind of sweet/to limp”. Before a non-*varga* consonant, the Anusvāra represents a nasal sound that may have an alternative conjoined writing symbol representing the corresponding nasal consonant of the particular set. Although Modern Hindi, Marathi and Konkani prefer the anusvāra to the corresponding Half-nasal, in Bangla it is clearly demarcated as to where one must use the Anusvāra and where it has to be a conjunct cluster with a nasal as the first or the second component.

### 3.3.5 Nasalization: Candrabindu (◌ঁ - U+0981)

Candrabindu denotes nasalization of the preceding vowel as in চাঁদ /cāṁd/ ‘moon’ (U+099A U+09BE U+0981 U+09A6). This sign with a dot inside the half-moon mark is used as nasalization marker in many Brāhmī-based scripts. [143]

### 3.3.6 Nukta (◌্ - U+09BC)

The nukta sign does not exist in Bangla orthography. It is predominantly used in many Brāhmī derived scripts, such as Devanāgarī (for Hindi, Bodo, Maithili, Santali, Kashmiri and Sindhi. The term and the concept of *nukta* are borrowed in Bangla.

The IDNA Protocol (RFC 5891) states that IDNs must be in Unicode Normalization Form C (NFC). RFC 7940 applies this requirement to LGRs. The definition of NFC in the Unicode Standard contains a number of composition exclusions. As a result, the Bangla letters য় YYA, ঙ্গ RRA and ঙ্গ RRHA have to be represented in the this LGR by using the sequences (YA +Nukta: U+9AF + U+09BC), (DDA + Nukta: U+9A1 + U+09BC), and (DDHA + Nukta: U+9A2 + U+09BC) instead of the single code points YYA (U+9DF), RRA (U+09DC), and RRHA (U+09DD), although the use of ‘Nukta’ is otherwise completely unnatural in Bangla.

It is noted that in the current Unicode Standard chart, these characters are listed as additional consonants. As per the LGR Procedure, however, these decisions depend on the IDNA Protocol through a set of procedures developed by the IETF. Even though the Unicode Standard also prescribes methods to produce these three characters both as atomic characters (for example, 09DC for ঙ্গ [r], 09DD for ঙ্গ [rh], and 09DF as য় [y] as single key stroke), the IDNA protocol requires that we treat them as conjunct characters and then allocate codes for these in the Unicode Bengali Block.

It may be noted that there could be sporadic attempts or cases of writing Muslim names, Urdu poetic words and Perso-Arabic loan words with *nukta* under ক (k), খ (kh), গ (g), জ (j) and ফ (ph) only for the sake of correct pronunciation and for maintaining the sanctity of the loan word. These were also like using Bangla writing system to work like the IPA script. It is, however, not in use in Bangla writing in printing.

### 3.3.7 Visarga /bifɔrgo/ (◌ঃ - U+0983) and Avagraha (◌্ - U+09BD)

The Visarga /bifɔrgo/ U+0983 is frequently used in Bangla loanwords borrowed from Sanskrit and represents a sound very close to /h/. One could quote, as an example: দুঃখ /duhkho/ “sorrow”, “unhappiness” (U+0926 U+0941 U+0983 U+0916).

The Avagraha "◌্" (U+09BD) is mainly used in Sanskrit, Pāli, Prākṛt or Maithili texts written in Bangla. It is gradually being replaced by an upper comma (e.g. বরোঃপরগি re-written as বরোপরগি). It is rarely used now even in other languages using Bangla script. In case of LGR, the Avagraha is not part of the repertoire. It has been decided, therefore, not to retain Avagraha (◌্) (U+09BD) because it is blocked in TLDs as per the Maximal Starting Repertoire (MSR).

Please see Appendix II in section 11 for a complete list of Bangla consonants and their allographs.

### 3.3.8 Zero Width Non-joiner (U+200C) and Zero Width Joiner (U+200D)

This note is pertinent to the use of Zero Width Joiner (ZWJ) and Zero Width Non Joiner (ZWNJ) as used in Bangla. It needs to be noted that Nepali, Konkani and Hindi use these two signs in a different manner.

ZWJ (U+0200D) and ZWNJ (U+0200C) are code points that have been provided by the Unicode standard to instruct the rendering of a string where the script has the option between joining and non-joining characters. Without the use of these control codes, the string may be rendered in an alternate form from what is intended.

#### Use of ZWJ

- Insofar as Bangla is concerned ZWJ is used for the proper rendering of characters such as *khaṇḍa-ta* /ৎ/ as in সত্যজিৎ (satyajit) “Satyajit” and সৎ (sat) “honest”. This is typed as follows:  
ta + Hasanta + ZWJ (U+0200D)

- However, ZWJ is more important where same combination of consonantal characters is represented differently depending upon the contexts. E.g. র+্+য have two representations in Bangla—as র্ and as র. To get the form র্ one has to type in the following manner—র+্+য, but for র the sequence would be র+ZWJ+্+য. [154]. In other words, ZWJ is used in the rendering of words demanding *ya-phalā* after *ra* which is otherwise not possible to type (render) due to the same order of *ra+hasanta+antastha ja* in the medial and/or final position. Interestingly, *ra+hasanta+antastha ja* is used to type *repha* on the consonant - *antastha ja* as in কার্ (kaarjo). In order to get a *ya-phalā* after the consonant -*ra* it is therefore obligatory to use ZWJ after -*ra* as in রাপার (wrapper), রাশ (rash), রালি (rally) etc. The typing sequence is given below:

*ra* (র) + ZWJ + *hasanta* (্) + *antastha ja* (য) = র্

### Use of ZWNJ

- The use of ZWNJ in Bangla is used to represent the *explicit Hasanta* or *Halant*. In order to avoid conjunct formation in cases where there is an explicit hasanta before the succeeding consonant the ZWNJ is used.

Consonant + hasanta + ZWNJ + consonant = explicit hasanta

Example: প্রাক্কথন (prākkathana /prakkɔtʰon)

The use of ZWJ/ZWNJ have been ruled out from the root zone by the [Procedure]. Used in Bangla, to create alternate renderings, the insertion of these two signs can affect searching as well as NLP.

The Zero Width Non-joiner (ZWNJ) is an invisible character used in certain cases (after Hasanta) where default conjunct formation is to be explicitly restricted and the Hasanta joining the two consonants participating in the conjunct formation needs to be explicitly shown.

### 3.3.9 Use of Ya-phalaa

Ya-Phalaa sequences are two instances in Bangla where Hasanta is preceded by a full vowel (U+0985 অ - BENGALI LETTER A and U+098F ঐ - BENGALI LETTER E).

- অ্যা 0985 09CD 09AF 09BE  
BENGALI LETTER A + BENGALI SIGN VIRAMA +  
BENGALI LETTER YA + BENGALI VOWEL SIGN AA
- ঐ্যা 098F 09CD 09AF 09BE

#### BENGALI LETTER E+ BENGALI SIGN VIRAMA + BENGALI LETTER YA + BENGALI VOWEL SIGN AA

For rendering *Ya-phalā* followed by অ and ঞ, it is necessary to type U+09CD Hasanta plus U+09AF *ya* preceded by the said vowels. This is a purely ligatural entity and the addition of *Ya-phalā* and ākāra is used to elicit the /æ/ sound as in English 'acid' অ্যাসিড, 'association' অ্যাসোসিয়েশন, 'bat' ব্যাট, 'fat' ফ্যাট, 'mat' ম্যাট, 'cap' ক্যাপ etc.

The Brāhmī script, by nature does not have Hasanta after a vowel. Hasanta is generally described as 'vowel killer', although it actually indicates absence of a vowel after the marked consonant. Only the consonants can have the Hasanta marked. But as we see here, Bangla ends up with a deviant feature in the orthography here in which Hasanta comes immediately after a vowel in ligatures অ্যা and ঞ্যা (Cf Unicode 10.0 p. 473 [100]).

#### 3.3.10 Formation of Ra-phalaa and Ref Sequences

This case refers to the formation of *repha* and *ra-phalā* as follows:

Ra-Hasanta = (C2 H)  
where C2 is either  
09B0 (৳ - BENGALI LETTER RA) or  
09F0 (ৰ - ASSAMESE LETTER RA/ Unicode name:  
BENGALI LETTER RA WITH MIDDLE DIAGONAL)  
H is 09CD (্ - BENGALI SIGN VIRAMA)

Owing to co-occurrence with HASANTA, RA either loses its own implicit vowel (REPHA), or suppresses the implicit vowel of the preceding consonant (RA-PHALĀ). For instance, *repha* = *ra* + Hasanta + C (e.g. ৰ্ক i.e. *ra* + Hasanta + *ka*, as in অৰ্ক *arka* "the sun"); *ra-phalā* = C + Hasanta + *ra* (e.g. ৰ্চ i.e. *ka* + Hasanta + *ra*, as in চৰ্চ *chakra* "cycle"). The point is in both the cases the slot for *ra* could be Bangla *ra* ৳ (U+09B0) or the Assamese *ra* ৰ (U+09F0), followed/ preceded by the common Hasanta (U+09CD), whereas the shapes of *repha* and *ra-phalā* in both the cases remain the same. The LGR makes a note of this point of concern with respect to the two RAs in disguise as it would be completely impossible to distinguish between them with naked eyes in a label so generated which may consequently lead to concerns related to spoofing and other kind of cyber irregularities. The motive to class these two CPs as (blocking) variants is because fully rendered labels may mask the distinction between Bangla *ra* ৳ (U+09B0) or the Assamese *ra* ৰ (U+09F0). That provides the justification for Variant Set 4, though only in the context of following Hasant. The difference between the RAs is only distinguishable if one looks into their Unicode values. Therefore, labels such as অৰ্ক *arka*, শীৰ্ষ *śīrṣa* 'top/apex', অভ্র *abhra* 'cloud/the sky', শ্রম *śrama* 'physical labour' could be extremely dangerous as the web-user may never verify the digital content (the labels) with its unicode value/code points. This point is made explicitly, with reference to Table 9 (of sequences, p. 36) and Table 16 (of WLE Symbols, p. 47) that are to follow. Moreover, it

is noteworthy that the REPHA can also occur with KHANDA TA. The conditions in this context of KHANDA TA are liable to be such that the C should be either RA U+09B0 (ꠕ) (used in Bangla) or RA U+09F0 (ꠕ) (used in Assamese).

## 4. Overall Development Process and Methodology

The Neo-Brāhmī Generation Panel (NBGP) has been formed by members having experience in Linguistics (especially in NLP / Computational linguistics), Literature, Language History and Epigraphy. Under the Neo-Brāhmī Generation Panel, Bangla and eight other scripts belonging to separate Unicode blocks are being taken up to assign a separate LGR for each. However, an attempt is made to ensure that the fundamental philosophy behind building those LGRs consistent with all other Brāhmī-derived scripts. The present LGR will cater to multiple languages belonging to EGIDS scale 1 to 4 (see Table 4) that use Bangla script.

The following guiding principles are used in making decisions about Bangla LGR Code-points:

### 4.1 Guiding Principles

The NBGP adopts following broad principles for selection of code-points in the code-point repertoire across the board for all the Neo-Brāhmī scripts within its ambit.

#### 4.1.1 Inclusion Principles

##### 4.1.1.1 Modern Usage

Every character proposed should be in the everyday usage of a particular linguistic community. The characters, which have been encoded in the Unicode for transcription purposes only or for archival purposes, will not be considered for inclusion in the code-point repertoire.

##### 4.1.1.2 Unambiguous Use

Every character proposed should have unambiguous understanding among linguists about its usage in the language.

### 4.2 Exclusion Principles

The main exclusion principle is that of External Limits on Scope. These consist of protocols or standards, which are prerequisites to the Label Generation Rule-sets. All further principles are in fact subsumed under these limitations but have been spelt out separately for the sake of clarity.

#### 4.2.1 External Limits of Scope

The code point repertoire for root zone being a very special case, at the top of protocol hierarchies, the canvas of available characters for selection as a part of the Root Zone code point repertoire is already constrained by various protocol layers beneath it. The following three main protocols/standards act as successive filters:

##### i. *The Unicode Chart*

Out of all the characters that are needed by the script in question, if a particular character is not encoded in Unicode, it cannot be incorporated in the code point repertoire. Such cases are quite rare, and especially so in Bangla-Asamiyā-Manipuri Writing System, given the elaborate and exhaustive character inclusion efforts made by the Unicode consortium.

##### ii. *IDNA Protocol*

Unicode being the character-encoding standard for providing the maximum possible representation of a given script/language, it has encoded as far as possible all the possible characters needed by the script. However, the Domain name being a specialized case, it is governed by an additional protocol known as IDNA (Internationalized Domain Names in Applications). The IDNA protocol excludes some characters out of Unicode repertoire from being part of the domain names.

##### iii. *Maximal Starting Repertoire (MSR)*

The Root-zone LGR being the repertoire of characters which are going to be used for creation of the Root-zone TLDs, which in turn constitute an even more specialized case of domain names, the ROOT LGR procedure introduces additional exclusions on the IDNA's allowed set of characters.

Example: Bangla Sign Avagraha "ঐ" (U+093D) even if allowed by IDNA protocol, is not permitted in the Root Zone Repertoire as per the MSR.

To sum up, the restrictions start off with admitting only such characters as are part of the code-block of the given script/language. The IDNA Protocol further narrows this down and finally an additional filter in the form of Maximal Starting Repertoire restricts the character set associated with the given language even more.

#### 4.2.1.1 No Punctuation Marks

The TLDs being identifiers, punctuation markers present in BraHami-based scripts will not be included.

#### 4.2.1.2 No Symbols and Abbreviations

Abbreviations, weights and measures and other such iconic characters like BANGLA ISSHAR "৳" (U+09FA), BANGLA CURRENCY DENOMINATOR SIXTEEN "৳" (U+09F9) etc. will also not be included.

#### 4.2.1.3 No Rare and Obsolete Characters

There are characters which have been added to Unicode to accommodate rare forms such as Sanskritic VOCALIC RR "ঐ" (U+09E0) and VOCALIC L "ঐ" (U+098C) as well as VOCALIC LL "ঐ" (U+09E1) and the allographic -kāra forms of the latter two symbols - VOWEL SIGN VOCALIC L "ঐ" (U+09E2) and VOWEL SIGN VOCALIC LL "ঐ" (U+09E3). All such characters are excluded, which complies with the Conservatism principle as laid down in the Root Zone LGR procedure. However, in Bangla, the -kāra corresponding to VOCALIC RR "ঐ" (U+09E0) which is VOWEL SIGN VOCALIC RR "ঐ" (U+09C4) is still in active use in certain limited borrowed or Sanskritic words, and are, therefore, retained.

#### 4.2.1.4 No Stress Markers of Classical Sanskrit and Vedic

Stress markers for classical Sanskrit will not be included. This is also in consonance with the Letter principle as laid down in the Root Zone LGR procedure.

#### 4.2.1.5 ABNF

The Augmented Backus-Naur Formalism (ABNF) is described in Section 5.4.1 and Appendix (Section 10.1).

## 5. Repertoire

The Bangla Writing System is represented in UNICODE using the Bengali (Bangla) script name as enumerated in ISO 15924 corresponding to languages such as Asamiyā (Assamese), Bangla (Bengali) and Maṇipuri. The BENGALI block used for Bangla-Asamiyā-Maṇipuri in the UNICODE has 93 entries. This section details the code-point repertoire that the Neo-Brāhmī Generation Panel [NBGP] proposes to be included in the Bangla LGR.

It may be mentioned here that the Government of Assam has submitted a proposal to Bureau of Indian Standards (BIS) on 26<sup>th</sup> February 2016 for dis-unification of Bangla and Asamiyā Scripts. The BIS in its 8th Meeting of Indian Language Technologies and Products Sectional Committee, LITD 20, held on 23rd Aug 2017, decided to refer the proposal for recognition of Assamese script in ISO/IEC 10646 to ISO. Until the UNICODE Consortium takes any further action, it will be assumed that the Code Point Repertoire under Table 11 will be valid for all the three languages as above.

For each of the code points, language references have been given in the last column titled "Reference" under Table 8 titled the "Code Point Repertoire". For entire coverage of Bangla code points, references of Bangla, Asamiyā (Assamese), Maṇipuri (Meitei), and Bishnupriya are given. Kokborok, written in Bangla script, is not known to have introduced many new complications, except for one particular character. Though only a few representative languages under EGIDS Scale 1-4 have been chosen for referencing, they together cover all the code-points required for all the languages that NBGP has considered as given under Bangla Unicode Points (as given in UNICODE 6.3).

However, before the details are presented, it is ideal to look at the Bangla Code Point Chart from Maximal Starting Repertoire [MSR] Version 3. It may be noted that the shapes of the reference glyphs given below in the code charts are based on one of the many fonts designed, and are not prescriptive, because there could be some variations in actual fonts – both UNICODE-compatible and True-Type ones. Consider the following Code point table:

	098	099	09A	09B	09C	09D	09E	09F
0	৭ (098)	ঐ (099)	ঔ (09A)	ৱ (09B)	ী (09C)		ঋ (09E)	৺ (09F)
1	ঁ (098)		ড (09A)		ৗ (09C)		ৠ (09E)	ৡ (09F)
2	ং (098)		ঢ (09A)	ল (09B)	৒ (09C)		ৢ (09E)	ৣ (09F)
3	ঃ (098)	ও (099)	ণ (09A)		৚ (09C)		ৣ (09E)	৤ (09F)
4		ঔ (099)	ত (09A)		৛ (09C)			৥ (09F)
5	অ (098)	ক (099)	খ (09A)					২ (09F)
6	আ (098)	খ (099)	দ (09A)	শ (09B)			০ (09E)	৩ (09F)
7	ই (098)	গ (099)	ঘ (09A)	ষ (09B)	ৌ (09C)	ৌ (09D)	১ (09E)	। (09F)
8	ঐ (098)	ঋ (099)	ন (09A)	স (09B)	ৈ (09C)		২ (09E)	ৣ (09F)
9	উ (098)	ঊ (099)		হ (09B)			৩ (09E)	০ (09F)
A	ঊ (09A)	চ (099)	প (09A)				৪ (09E)	৷ (09F)
B	ঋ (09B)	ৠ (099)	ফ (09A)		ৌ (09C)		৫ (09E)	৶ (09F)
C	ৠ (09C)	জ (099)	ব (09A)	় (09B)	ৌ (09C)	ড় (09D)	৬ (09E)	
D		ঝ (099)	ভ (09A)	৞ (09B)	্ (09C)	ঢ (09D)	৭ (09E)	
E		ঞ (099)	য (09A)	া (09B)	ৎ (09C)		৮ (09E)	
F	এ (098)	ট (099)	ষ (09A)	ি (09B)		য় (09D)	৯ (09E)	

#### Colour convention:

All characters that are included in the [MSR]  
- Yellow background

PVALID in IDNA2008 but excluded from the  
[MSR] - Pinkish background

Not PVALID in IDNA2008, or are ineligible  
for the root zone (digits, hyphen) - White background

Figure 1: Bangla Code Page from [MSR] for  
Bangla- Asamiyā -Maṇipuri

Given the Bangla Unicode Block as in Figure 1, for the code points those are included in the MSR, the following symbols will need a separate treatment:

- ঈ U+09CE Bangla Letter Khaṇḍa-Ta
- ৺ U+09F0 Asamiyā -Bangla Letter Ra With Middle Diagonal
- ৡ U+09F1 Asamiyā -Bangla Letter Ra With Lower Diagonal

## 5.1 Code Point Repertoire Inclusion

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
1.	U+0981	ঁ	BENGALI SIGN CANDRABINDU	Candra-bindu	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
2.	U+0982	ং	BENGALI SIGN ANUSVARA	Onushshar (Anusvāra)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
3.	U+0983	ঃ	BENGALI SIGN VISARGA	<i>Biśarga</i> (Visarga)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
4.	U+0985	অ	BENGALI LETTER A	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
5.	U+0986	আ	BENGALI LETTER AA	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
6.	U+0987	ই	BENGALI LETTER I	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
7.	U+0988	ঐ	BENGALI LETTER II	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
8.	U+0989	ঊ	BENGALI LETTER U	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
9.	U+098A	ঊ	BENGALI LETTER UU	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
10.	U+098B	ঋ	BENGALI LETTER VOCALIC R	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
11.	U+098F	এ	BENGALI LETTER E	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
12.	U+0990	ঐ	BENGALI LETTER AI	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
13.	U+0993	ও	BENGALI LETTER O	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
14.	U+0994	ঔ	BENGALI LETTER AU	Vowel	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
15.	U+0995	ক	BENGALI LETTER KA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
16.	U+0996	খ	BENGALI LETTER KHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
17.	U+0997	গ	BENGALI LETTER GA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
18.	U+0998	ঘ	BENGALI LETTER GHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
19.	U+0999	ঙ	BENGALI LETTER NGA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
20.	U+099A	চ	BENGALI LETTER CA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
21.	U+099B	ছ	BENGALI LETTER CHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
22.	U+099C	জ	BENGALI LETTER JA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
23.	U+099D	ঝ	BENGALI LETTER JHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
24.	U+099E	ঞ	BENGALI LETTER NYA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
25.	U+099F	ট	BENGALI LETTER TTA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
26.	U+09A0	ঠ	BENGALI LETTER TTHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
27.	U+09A1	ড	BENGALI LETTER DDA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
28.	09A1 09BC (U+09DC)	ৱ	Normalized form of BENGALI LETTER RRA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]  <i>09DC is the preferred code point, however it is not available for LGR as per the standards governing this LGR development</i>
29.	U+09A2	৳	BENGALI LETTER DDHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
30.	09A2 09BC (U+09DD)	৴	Normalized form of BENGALI LETTER RHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]  <i>09DD is the preferred code point, however it is not available for LGR as per the standards governing this LGR development</i>
31.	U+09A3	৵	BENGALI LETTER NNA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
32.	U+09A4	৶	BENGALI LETTER TA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
33.	U+09A5	থ	BENGALI LETTER THA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
34.	U+09A6	দ	BENGALI LETTER DA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
35.	U+09A7	ধ	BENGALI LETTER DHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
36.	U+09A8	ন	BENGALI LETTER NA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
37.	U+09AA	প	BENGALI LETTER PA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
38.	U+09AB	ফ	BENGALI LETTER PHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
39.	U+09AC	ব	BENGALI LETTER BA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
40.	U+09AD	ভ	BENGALI LETTER BHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
41.	U+09AE	ম	BENGALI LETTER MA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
42.	U+09AF	য	BENGALI LETTER YA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
43.	09AF 09BC (U+09DF )	য়	Normalized form of BENGALI LETTER YYA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese,	[112], [122], [125]  <i>09DF is the preferred code point, however it is not available for LGR as per the standards governing this LGR development</i>
44.	U+09B0	র	BENGALI LETTER RA	Consonant	1 Bangla, 2 Maṇipuri	[112], [122]
45.	U+09B2	ল	BENGALI LETTER LA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
46.	U+09B6	শ	BENGALI LETTER SHA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
47.	U+09B7	ষ	BENGALI LETTER SSA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
48.	U+09B8	স	BENGALI LETTER SA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
49.	U+09B9	হ	BENGALI LETTER HA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
50.	U+09BE	া	BENGALI VOWEL SIGN AA	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
51.	U+09BF	ি	BENGALI VOWEL SIGN I	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
52.	U+09C0	ী	BENGALI VOWEL SIGN II	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
53.	U+09C1	ু	BENGALI VOWEL SIGN U	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Gly ph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
54.	U+09C2	ঊ	BENGALI VOWEL SIGN UU	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
55.	U+09C3	ঋ	BENGALI VOWEL SIGN VOCALIC R	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
56.	U+09C4	৐	BENGALI VOWEL SIGN VOCALIC RR	Kāra (Mātrā)	1 Bangla, 2 Assamese	[112], [125]
57.	U+09C7	এ	BENGALI VOWEL SIGN E	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
58.	U+09C8	ঐ	BENGALI VOWEL SIGN AI	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
59.	U+09CB	৓	BENGALI VOWEL SIGN O	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
60.	U+09CC	৔	BENGALI VOWEL SIGN AU	Kāra (Mātrā)	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]

No.	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	References and Comment
61.	U+09CD	্	BENGALI SIGN VIRAMA	Hasanta (=Halant)/ Virāma (=Dāri)	1 Bangla, 2 Assamese 2 Maṇipuri	[112], [122], [125]
62.	U+09CE	১	BENGALI LETTER KHANDA TA	Consonant	1 Bangla, 2 Maṇipuri, 2 Assamese	[112], [122], [125]
63.	U+09F0	৲	BENGALI LETTER RA WITH MIDDLE DIAGONAL	Consonant	2 Assamese	[125]
64.	U+09F1	৳	BENGALI LETTER RA WITH LOWER DIAGONAL	Consonant	2 Maṇipuri 2 Assamese	[122],[125]

Table 8: Bangla Code-Point Repertoire

Apart from the above individual code-points, the Neo-Brāhmī Generation Panel also proposes some specific sequences which enable conditional inclusion of the "Bangla LETTER A and E" followed by Bangla SIGN VIRAMA and Bangla LETTER YA again followed by Bangla VOWEL SIGN AA in the repertoire for enabling inclusion of /æ/ sound as in English 'bat', 'cat' etc.

Sr. No.	Unicode Code Points	Sequence	Character Names	Example languages using the code-point (Not exhaustive list)	Reference

Sr. No.	Unicode Code Points	Sequence	Character Names	Example languages using the code-point (Not exhaustive list)	Reference
S1.	0985 09CD 09AF 09BE	অ্যা	BENGALI LETTER A BENGALI SIGN VIRAMA BENGALI LETTER YA BENGALI VOWEL SIGN AA	Bangla, Assamese	[112], [122]
S2.	098F 09CD 09AF 09BE	এ্যা	BENGALI LETTER E BENGALI SIGN VIRAMA BENGALI LETTER YA BENGALI VOWEL SIGN AA	Bangla	[112]

Table 9: Sequences

## 5.2 Code Point Repertoire Exclusion

There are some characters of the Bangla script that find place in the Unicode but have not been included in the repertoire in the LGR proposal. The reason for excluding ঙ (U+098C) and ৌ (U+09D7) is that they are rare and obsolete characters.

Sr. No.	Code Points	Glyph	Character Names	Note
1.	U+098C	ঙ	BENGALI LETTER VOCALIC L	Limited or declining use
2.	U+09D7	ৌ	BENGALI AU LENGTH MARK	Limited or declining use

Table 10: Excluded Code Points

### 5.3 Code point not used alone

BENGALI SIGN NUKTA U+09BC (See 3.3.6) is excluded from repertoire since it will never be used alone. It will be used as sequence in three special characters in normalized form for ড়, ঢ়, ঞ.

Unicode Code Point	Glyph	Character Name	Reason for exclusion
U+09BC	়	BENGALI SIGN NUKTA	Never used alone. Only used together with U+09A1 ড, U+09A2 ঢ, U+09AF ঞ as to form ড়, ঢ়, ঞ respectively

Table 10b: Excluded Code Points

### 5.4 The Basis of Present IDN

The present LGR has also benefited from the earlier work on IDN for Bangla (different versions) done for .भारत or .ভারত drafted between 20.11.2009 and 18.07.2013.

#### 5.4.1 The ABNF Variables

The Augmented Backus-Naur Formalism (ABNF) began with the following variables:

- C → Consonant
- V → Vowel
- M → kâra (Mātrā)
- B → Anusvāra (/onuffār /)
- D → Candrabindu
- X → Visarga (/biʃorgo)
- H → Hasanta /Virāma
- Z → Khaṇḍa Ta

The Augmented Backus-Naur Formalism (ABNF) will use the following Operators:

Sr. Number	Operator	Function
1	“   ”	Alternative
2	“ [ ] ”	Optional
3	“ * ”	Variable Repetition
4	“ ( ) ”	Sequence Group

Table 11: The ABNF Formalism

In what follows, the Vowel Sequence and the Consonant Sequence pertinent to Bangla are given to facilitate understanding.

### 5.4.2 The Vowel Sequence

In what follows, the Vowel Sequence and the Consonant Sequence pertinent to Bangla are given. To facilitate understanding of other Brahmi script users, equivalents in Devanāgarī are provided, wherever necessary.

A vowel sequence is made up of a single vowel. It may be followed but not necessarily (optionally) by an Anusvāra /onʊʃfār/ (B), Candrabindu (D) or a Visarga /biʃɔrgo/ (X). The number of D, B or X which can follow a V in Bangla may not be restricted to one. Going by the rules illustrated in the document it is clear that formations such as VDD, VBB and VXX are invalid orthographic units. However, it is valid and possible to have formations or sequences such as anusvāra followed by a chandrabindu on one hand and visarga followed by a chandrabindu on the other as in হাঁচা ‘hāenchā’ and হাঁঃ ‘hāen’ হাঁঃ respectively.

The possibility of a Visarga or Anusvāra (/onʊʃfār/) following a Candrabindu exists in Bangla. Vowel can optionally be followed by a combination of Hasanta / Virāma [H], Consonant [C] to form a Ya-phalā. “Ya-phalā is a presentation form of U+09AF Bangla letter য় or ‘ya’. Represented by the sequence < U+09CD, i.e. ্ BENGALI SIGN VIRAMA, Bangla SIGN Hasanta or VIRĀMA, U+09AF - য় BENGALI LETTER YA>, Ya-phalā has a special form: ঞ. Again, when combined with U+09BE া BENGALI VOWEL SIGN AA, (i.e. ‘aa’(ā)), it is used for transcribing [æ] as in the “a” in the English word “bat” written in Bangla as ব্যাটে.

A Vowel-sequence admits the following combinations:

#### 5.4.2.1 A Single Vowel

Examples:    V                    অ            ঐ

#### 5.4.2.2 A Vowel with Conditions

A Vowel can optionally be followed by Anusvāra [B] or Candrabindu [D] or Visarga [X] or Candrabindu+ Anusvāra [DB] or Candrabindu+ Visarga [DX] or combination of Hasanta (or Virama) [H] followed by Consonant [C] followed by kāra (Mātrā) [M].

Examples:

VB	अं	अं
VD	अँ	अँ
VX	अः	अः
VDB	अँ	अँं
VDX	अँ	अँः
VHCM	अया/एया	

#### 5.4.2.3 VHCM Sequence

A VHCM sequence can optionally be followed by Anusvāra [B] or Candrabindu [D] or Visarga [X] or Candrabindu+ Anusvāra [DB] or Candrabindu+ *Visarga* [DX].

Examples:

VHCMB	अयां/एयां
VHCMD	अयाँ/एयाँ
VHCMX	अयाः/एयाः
VHCMDB	अयाँं/एयाँं
VHCMDX	अयाँः/एयाँः

#### 5.4.3 The Consonant Sequence

##### 5.4.3.1 A Single Consonant (C)

Example: C क क

##### 5.4.3.2 A Consonant with Conditions

A Consonant optionally followed by dependent vowel sign / kāra (Mātrā) [M] or Anusvāra [B] or Candrabindu [D] or Visarga [X] or Hasanta (also known as Virāma) [H] or Candrabindu+ Anusvāra [DB] or Candrabindu+ *Visarga* [DX]

Example:

CM	कि/ कृ	कि/ कृ
CB	कं	कं

CD	कँ	कँ
CX	कः	कः
CH	क्	क् (Pure consonant)
CDB	कँं	कँं
CDX	कँः	कँः

#### 5.4.3.3 CM Sequence

A CM sequence can be optionally followed by B, D, X, DB or DX.

Example:

CMB	कीं/ कूं	कीं/ कूं
CMD	काँ	काँ
CMX	वीः	वीः
CMDB	काँं	काँं
CMDX	काँः	काँः

#### 5.4.3.4 Sequence of Consonants

A sequence of consonants (up to 4) joined by Hasanta (also known as Virama).

\*3(CH)C

Example:

CHC	ञ	→	न+्+ त	न+ त
CHCHC	ञ	→	न+्+ त+्+ र	न+ त्+ र
CHCHCHC	ञ	→	न+्+ त+्+ र+्+ य	न+ त्+ र्+ य

#### 5.4.3.5 Subsets:

While considering its subsets, as a representative example, we will consider the combination CHC only, however the same is equally applicable to CHCHC and CHCHCHC.

**[A].** The combination may be followed by M, B, D, X, DB or DX.

Example:

CHCM	क्की	→ क् क ी	क्की → क् क ी
CHCB	क्कं	→ क् क ं	क्कं → क् क ं
CHCD	क्कँ	→ क् क ँ	क्कँ → क् क ँ

CHCX	ক্ৰঃ → ক ্ ক ঃ	ক্কঃ → ক ্ ক ঃ
CHCDB	ক্ৰঁ ং → ক ্ ক ঁ ং	ক্কঁ ং → ক ্ ক ঁ ং
CHCDX	ক্ৰঃ → ক ্ ক ঁ ঃ	ক্কঁ ঃ → ক ্ ক ঁ ঃ

[B]. \*3(CH)CM may further be followed by a B, D, X, DB or DX

Example:

CHCMB	ক্ৰীং → ক ্ ক ী ং	ক্কীং → ক ্ ক ী ং
	ক্ৰ্ং → ক ্ ক ্ ং	ক্কৃং → ক ্ ক ৃ ং
CHCMD	ক্ৰাঁ → ক ্ ক া ঁ	ক্কাঁ → ক ্ ক া ঁ
CHCMX	ক্ৰীঃ → ক ্ ক ী ঃ	ক্কীঃ → ক ্ ক ী ঃ
CHC MDB	ক্ৰাঁং → ক ্ ক া ঁ ং	ক্কাঁং → ক ্ ক া ঁ ং
CHCMDX	ক্ৰাঁঃ → ক ্ ক া ঁ ঃ	ক্কাঁঃ → ক ্ ক া ঁ ঃ

#### 5.4.4 The Khaṇḍa-Ta sequence

##### 5.4.4.1 A single 'Khaṇḍa'-Ta (Z)

Example: Z ৎ = ত্

##### 5.4.4.2 A Khaṇḍa Ta Combination<sup>10</sup>

A Khaṇḍa Ta can be preceded by a consonant and Hasanta (also known as Virāma)

[CH]Z

Example:

ৰ + ্ + ত্ = ৰ্ as in ভৰ্ৎসনা (bhartsanā) "scolding"

Note: The conditions in this context of KHANDA TA are that the C should be either RA U+09B0 (ৰ) (used in Bangla) or RA U+09F0 (ৰ্) (used in Assamese).

##### 5.4.5 Special Cases S and P

Two special cases involving Sequences (referred to as S and P in Table 16 under Section 7) could be described briefly here. Let us take up S in the first instance. It is noteworthy that there are two instances in Bangla where Hasanta (U+09CD) is preceded by a full

<sup>10</sup> Refer to Rule P in Section 7, Table 16.

vowel (U+0985 অ - BANGLA LETTER A and U+098F এ - BANGLA LETTER E). For rendering *Ya-phalā* followed by অ and ঐ, it is necessary to type U+09CD plus U+09AF ya preceded by the said vowels. This is a purely ligatural entity and the addition of *Ya-phalā* and *ā-kāra* is used to elicit the /æ/ sound as in English ‘bat’, ‘fat’ etc. The Brāhmī script, by nature does not have Hasanta after a vowel. Hasanta is generally described as ‘vowel killer’, although it actually indicates absence of a vowel after the marked consonant. Only the consonants can have the Hasanta marked. But as we see here, Bangla ends up with a deviant feature in the orthography here in which Hasanta comes immediately after a vowel in ligatures আ and ঐ (Cf Unicode 10.0 p. 473 [100]).

Another case refers to the formation of *repha* and *ra-phalā* in the said script and mentioned in the table above as P. Owing to co-occurrence with HASANTA, RA either loses its own implicit vowel (REPHA), or suppresses the implicit vowel of the preceding consonant (RA-PHALĀ). For instance, repha = *ra* + Hasanta + C (e.g. র্ধ i.e. *ra* + Hasanta + *ka*, as in অর্ক *arka* “the sun”); *ra-phalā* = C + Hasanta + *ra* (e.g. ক্ধ i.e. *ka* + Hasanta + *ra*, as in চক্র *cakra* “cycle”). The point is in both the cases the slot for *ra* could be Bangla *ra* র (U+09B0) or the Assamese *ra* ৰ (U+09F0), followed/ preceded by the common Hasanta (U+09CD), whereas the shapes of *repha* and *ra-phalā* in both the cases remain the same.

## 6. Variants

This section talks about the variants in the Bangla script. The NBGP categorizes these confusingly variants in two groups.

**Group 1:** Confusing due to pure visual similarity.

**Group 2:** Confusing due to deviation from normally perceived character formations by larger linguistic community.

For Group 1, any identical code points are defined as variants. The confusable, but not identical, cases are not proposed, as there is another panel (String similarity assessment panel) entrusted to deal with such cases. However, cases which belong to Group 2 are proposed to be considered as variants. These cases are not of mere visual similarity as they involve some deviations from the widely accepted norms of Bangla Akshar formations. These can cause confusion even to a careful observer and hence being proposed as variants.

The variants are generated in a script when two or more forms are formed with different storage or code points. In Bangla the *e-kāra*, *ā-kāra* and the *o-kāra* have different code points. One can type *o* with a consonant at one go and the same by typing *e-kāra* and *ā-kāra* as two separate keys getting the same results. A reader cannot differentiate between the two *ko* (কো), one typed with a single key and the other one typed with two different keys. Moreover, this will not be considered as a case of variant because a *kāra* followed by a *kāra* is not allowed.

### 6.1 In Script Variants

However, we propose two cases of true in-script variants in Bangla script.

#### CASE I:

As far as true variants in Bangla are concerned, we may draw our attention to cases wherein Hasanta with (U+09A5) থ (*tha*) appears as conjunct with (U+09B8) স (*sa*) and (U+09A8) ন (*na*).

1. স + Hasanta + থ (U+09B8 + U+09CD + U+09A5) versus  
স + Hasanta + ঠ (U+09B8 + U+09CD + U+09B9)
2. ন + Hasanta + থ (U+09A8 + U+09CD + U+09A5) versus  
ন + Hasanta + ঠ (U+09A8 + U+09CD + U+09B9)

The above combinations, if written in traditional orthography, could be little confusing, where the থ (tha) in conjunct appears like a হ (ha). The conjunct could be in the initial, medial or final positions (as shown below in e.g. no 1). It could be typed wrong as well, thinking it was a হ (ha) U+09B9, increasing the chances of risks in label writing and identification.

Examples:

1. স্থ and স্থ (as in স্থান sthāna, স্থূল sthūla, স্বাস্থ্য svāsthya, অস্থায়ী asthāyī)
2. হ্র and হ্র (as in গ্রন্থ grantha)

The fonts which represent traditional Bangla writing system could tend to create this problem. Therefore, these may be taken as cases of variants in Bangla.

## CASE II:

Another interesting example of variant is encountered in *ra + Hasanta* and *Hasanta + ra* combinations in writing labels in the Bangla script (for languages such as Bangla, Assamese and Maṇipuri). The variant cases arise in typing '**repha**' (involving *ra + Hasanta*) and '**ra-phala**' (involving *Hasanta + ra*).

'Repha' could be formed by two sequences (mainly because both Assamese and Bangla find place in the same UNICODE points, and 'B\_RA' as well as 'A\_RA' refer to the same phonetic element). Here, the final ligatures look the same, and will be as follows:

- (1) B\_RA + H + C
- (2) A\_RA + H + C

Where

- B\_RA = U+09B0 BENGALI LETTER RA (র) or  
A\_RA = U+09F0 BENGALI LETTER RA WITH MIDDLE DIAGONAL (৳)  
H = U+09CD BENGALI SIGN VIRAMA (্)  
C = any consonant (theoretically)

Example:

Sequence1 (Using Bangla RA)	Ligature 1	Sequence2 (Using Assamese RA)	Ligature 2
U+09B0 (র) U+09CD (্) U+0995 (ক)	ৰ্ক	U+09F0 (৳) U+09CD (্) U+0995 (ক)	ৰ্ক
U+09B0 (র) U+09CD (্) U+09A0 (ঠ)	ৰ্ঠ	U+09F0 (৳) U+09CD (্) U+09A0 (ঠ)	ৰ্ঠ

Table 12: Example of Repha

Note: As Bangla and Assamese ক্ৰ and ৰ্ৰ look exactly the same, the resultant combinations with 'Repha' look identical. Addition of 'Repha' does not make any difference.

‘Ra-phalā’ could be formed by two sequences on similar grounds, and the final ligatures would look the same

(1) C1 + H + B\_RA

(2) C1 + H + A\_RA

Where

C1 = any consonants except Khaṇḍa-ta

Example:

Sequence1 (Using Bangla RA)	Ligature 1	Sequence2 (Using Assamese RA)	Ligature 2
U+0995 (ক্ৰ) U+09CD (়্ৰ) U+09B0 (ৱ)	ক্ৰৱ	U+0995 (ক্ৰ) U+09CD (়্ৰ) U+09F0 (ৰ্ৰ)	ক্ৰৱ
U+09A8 (ন্ৰ) U+09CD (়্ৰ) U+09B0 (ৱ)	ন্ৰৱ	U+09A8 (ন্ৰ) U+09CD (়্ৰ) U+09F0 (ৰ্ৰ)	ন্ৰৱ

Table 13: Example of Ra-phalā

As the Assamese and Bangla Repha and Ra-phalā conjunct forms look the same, this could cause confusability to the end-users. Hence, the repha and ra-phalā cases need to be defined as variants.

NBGP concluded to define ৱ and ৰ্ৰ as variant code points, where only one variant set between ৱ and ৰ্ৰ could cover all cases. But this will create blocked variant labels, e.g. if someone registers “ৱ ৱ ৱ” the variant label “ৰ্ৰৰৰৰ” will be generated as variant and will be blocked and vice versa. However, it is only blocked at the label level, if someone else needs to register other labels e.g. ৰৰ or ৰৰৰৰ, it is still possible.

After the public comment, the NBGP reviewed the disposition for ৱ and ৰ্ৰ variants. These code points are used equally. Therefore, for the usability, the NBGP decided that ৱ and ৰ্ৰ are variant “allocatable”. In addition, these code points 09B0 and 09F0 should not be used in the same label, therefore the no-mix rule should be implemented.

## 6.2 Cross Script Variants

A crisp cross script study for Bangla has been done with respect to sister scripts such as Devanāgarī, Gurmukhī and Odia<sup>11</sup> (formerly Oriya) keeping in mind the visual and technical confusions they may cause as labels on the web domain. Moreover, there is no in-script variant in Bangla as far as the orthography is concerned. The following characters are being proposed by the NBPG as variants. Although there are certain characters which are somewhat similar they but have not been included here. They have been provided in the Appendix (10.2) for reference.

### 1. Bangla and Nāgarī /Devanāgarī Script

Bangla	Devanāgarī
ঋ U+09AE	ऋ U+092E
ि U+09BF	ि U+093F

Table 14 - Bangla and Devanāgarī cross-script variant code point

### 2. Bangla and Gurmukhi Script

Bangla	Gurmukhī
ঋ U+09AE	ऋ U+0A38
ि U+09BF	ि U+0A3F

Table 15 - Bangla and Gurmukhī cross-script variant code point

## 7. Whole Label Evaluation Rules (WLE)

This section provides the WLEs that are required by all the languages mentioned in section 3.2 when written in Bangla<sup>12</sup> Script. The rules have been drafted in such a way that they can be easily translated into the LGR specifications.

<sup>11</sup> Unicode uses Oriya for the script, although Odia is now the official term used.

<sup>12</sup> As used by the Unicode, denoting and including both Assamese and Maṇipuri.

Below are the symbols used in the WLE rules, for each of the "Indic Syllabic Category" as mentioned in the table provided in Code point repertoire (Section 5.1).

C	→	Consonant
M	→	Kāra (Mātrā)
V	→	Vowel
B	→	Anusvāra
D	→	Candrabindu
X	→	Visarga
H	→	Hasanta
Z	→	Khaṇḍa Ta
S	→	<p>S1, S2 (from Table 9)</p> <p>or</p> <p>(a/e) <i>Ya-phalā</i> (V1 H C1 M1)  where  V1 is either 0985 (অ - BENGALI LETTER A)  or 098F (ঐ - BENGALI LETTER E)  H is 09CD (় - BENGALI SIGN VIRAMA)  C1 is - 09AF (য - BENGALI LETTER YA)  M1 is - 09BE (া - BENGALI VOWEL SIGN AA)</p> <p>S1 and S2 are valid, even they are not allowed by the other context rules.</p>
P	→	<p>Ra-Hasanta (C2 H)  where  C2 is either 09B0 (৳ - BENGALI LETTER RA)  or 09F0 (৳ - ASSAMESE LETTER RA/  Unicode name: BENGALI LETTER RA  WITH MIDDLE DIAGONAL)  H is 09CD (় - BENGALI SIGN VIRAMA)</p>

Table 16 - Symbols used in WLE rules

It is also perhaps ideal to mention here that in Bangla, the consonant letters (or graphemes) are physically joined to form “clusters” that could theoretically conjoin from two to four consonants and combine to create new shapes. Dash and Chaudhuri (1998) state that there are “nearly 380 unique consonant...clusters” out of which Bi-consonantal combinations are 290, three-letter combinations account for another 80 and the rarer ones with four letters number 10 more [136, Pg 4]. More details of such combinations could be seen in Pabitra Sarkar (1993) [135].

## 7.1 Final Set of WLE Rules

The prevalent patterns in Bangla, and various restrictions, below are the specific WLE rules that need to be implemented.

1. C is a set of C and CN where CN is the set of normalized forms of {ড়, ঢ়, ঝ}.
2. H: must be preceded by C  
Example: ক্
3. M: must be preceded by C  
Example: কা
4. D: must be preceded by either of V, C, or M  
Example: আঁ, খঁ, থাঁ, ছাঁ
5. X: must be preceded by either of V, C, M or D  
Example: উঃ, থঃ, বঃ, াঃ, দ্বুঃ
6. B: must be preceded by either of V, C, M or D  
Example: আঃ, ইঃ, কঃ
7. Z: must be preceded by V, C, M, D, B, X or P  
Example: ইং, কং, াং, াঁং, ঞং, ঞ্ং (S is not listed, because S ends with M, Z may also follow S)”.  
Details in 7.1.1 Case of V preceded by H
8. V: CANNOT be preceded by H
9. S: CANNOT be preceded by H
10. 09B0(ঐ) and 09F0(ঐ) CANNOT be mixed  
Details in 6.1 CASE II

Now let us elaborate each rule with examples from the script keeping in mind the Bangla, Assamese and Maṇipuri communities. Some combinations of characters may seem unrealistic or rare in usage but there is no harm in adding such ligatures because it is possible to create them by any user easily but may not be attested combinations.

### 7.1.1 Case of V Preceded by H:

There could be cases involving multi-word domains where V may need to be allowed to follow an H

e.g. ব্যাংক অফ ইন্ডিয়া /bæŋk ʌv indiə / (U+09AC U+09CD U+09AF U+09BE U+0999 U+09CD U+0995 U+0985 U+09AB U+09CD U+0987 U+09A8 U+09CD U+09A1 U+09BF U+09DF U+09BE) (meaning: *Bank of India*)

This is the case where two different words are joined together first of which ends with an H (অফ) and the second word begins with a V (ইন্ডিয়া). Some sections of the linguistic community require the explicit presence of H for full representation of the sound intended. However, by and large, the form of the first word without an H (U+09CD) is considered enough for full representation of the sound intended for the first word.

This is a unique situation necessitated by the lack of hyphen, space or the Zero Width Non-joiner character in the permissible set of characters in the Root zone repertoire. Otherwise, V is never required to be allowed to follow an H. Permitting this may create a perceptive similarity between two labels (with and without H) for majority of the linguistic communities hence this is explicitly prohibited by the NBGP.

In future if required, depending on the prevailing requirements from the community, the future NBGP may consider revisiting this rule.

### 7.2 Additional Examples from Bangla ABNF:

Below are a few examples which help one understand some of the rules ABNF puts in place. These are just given for reference purposes and are not meant to be comprehensive.

1. H, M, B, D or X cannot occur in the beginning of a Bangla word. Example:

্ক	্ক
াক	াক
ংক	ংক
ঁক	ঁক
ংক	ংক

As can be seen such combination will result automatically in a “golu” or a dotted circle marking it as an invalid formation. This is an intrinsic property of the

Indian language syllable and is quasi-automatically applied wherever supported by the OS.

2. H is not permitted after V, B, D, X, M, S

Example:

अ्	अ्
अं	कं
कं	कं
कः	कः
कि	कि

3. Number of B, D or X permitted after Consonant or Vowel or a kāra (Mātrā) is restricted to one thus the following combinations are invalidated.

Example:

कः	कं
कं	कं
कः	कः
कं	कं
कीः	कीः
अः	अं
अं	अं
अः	अः

4. Number of M permitted after Consonant is restricted to one.

Example:

की	की
----	----

5. M is not permitted after V.

Example:

इ/औ	ई/ई
-----	-----

6. The combinations of Anusvāra + *Visarga* as well as *Visarga* + Anusvāra are not permissible.

Example:

कः कः  
कः कः

## 8. Contributors

### 8.1 Experts from India

Professor Udaya Narayana Singh, Chair-Professor of Linguistics & Dean, Faculty of Arts, Amity University Haryana, Gurgaon; Pachgaon, Manesar PIN 122431 (Haryana), India.

Professor Pabitra Sarkar, formerly Vice-Chancellor, Rabindra Bharati University, Kolkata.

Dr Atiur Rahman Khan, Principal Technical Officer, GIST Group, C-DAC, Pune, PIN 411008 (Maharashtra), India.

Mr Rajib Chakraborty, Linguist, Society for Natural Language Technology Research (SNLTR), Module 114 & 130, SDF Building, Salt Lake, Sector-V, Kolkata-700091 (West Bengal), India.

Mr Akshat Joshi, Project Engineer, GIST Group, C-DAC, Pune, PIN 411008 (Maharashtra), India.

Ms Moumita Chowdhury, Senior Technical Officer, GIST Group, C-DAC, Pune, PIN 411008 (Maharashtra), India.

Mr Chandrakanta Murasingh, Agartala, Tripura.

Some other NBGP members.

### 8.2 Contributors from Bangladesh

Janab Mustafa Jabbar, Honorable Minister, Ministry of Posts, Telecommunications & Information Technology, Govt of Bangladesh

Prof Shamsuzzaman Khan, Former Director-General, Bangla Academy, Dhaka

Prof Rafiqul Islam, National Professor of Humanities, Dhaka.

Prof Swarochis Sarkar, Director, Institute of Bangladesh Studies, Rajshahi University, Rajshahi, Bangladesh

Prof Jinnat Imtiaz Ali, Director-General, International Mother Language Institute, Dhaka

Mr Mohammad Mamun Or Rashid , Department of Bangla, Jahangirnagar University & Member, Bangladesh Computer Council

Prof Maniruzzaman, formerly Professor, Chittagong University, Chattagram, Bangladesh

Mr Shyam Sunder Sikder, Secretary, Secretary, Post & Telecommunications Division Govt of Bangladesh

Mr Md. Mustafa Kamal, Former Director General, Bangladesh Telecommunications Regulatory Authority, Government of Bangladesh, Dhaka

Brigadier General Md Mahfuzul Karim Majumder, Director-General, Engineering & Operations Division, Bangladesh Telecommunications Regulatory Authority, Government of Bangladesh, Dhaka

Md. Ziarul Islam, Programmer, Posts & Telecommunications Division, Government of Bangladesh, Dhaka

Prof Syed Shahriyar Rahman, Department of Linguistics, University of Dhaka

Dr Mizanur Rahman, Director (In-Charge), Translation, Text Book and International Relations Division, Bangla Academy, Dhaka

Dr Aparesh Bandyopadhyay, Director, Bangla Academy, Dhaka

Mr Md Mobarak Hossain, Director, Bangla Academy, Dhaka

Dr Jalal Ahmed, Director, Bangla Academy, Dhaka

Mr Jahangir Hossain, Internet Society Bangladesh (ICANN ALS)

Janab Sarwar Mostafa Choudhury, Bangladesh Computer Council, Dhaka

Janab Md Rashid Wasif, Bangladesh Computer Council, Dhaka

Janab Istiaque Arif, Senior Assistant Director, Bangladesh Telecommunications Regulatory Authority, Dhaka

Ms. Afifa Abbas, Information Security and Governance Lead Engineer at Banglalink, and ICANN Fellow.

Mr Mohammad Abdul Haque, Secretary General, Bangladesh Internet Governance Forum

Mr Imran Hossen, CEO, EyeSoft and key member of Bangladesh Association of Software & Information Services (BASIS).

Ms Shahida Khatun, Director, Folklore, Museum & Archive Division, Bangla Academy, Dhaka

Mr Syed Ashik Rehman, CEO, Bengal Media Corporation, Dhaka

Mr Haseeb Rahman, CEO, Professionals' Systems, Dhaka

## 9. References

- [100] Unicode Consortium. 2017. Unicode Standard 10.0. Mountain View CA.
- [101] Bandyopadhyay, Chittaranjan. 1981. *Dui Shataker Bangla Mudran o Prakashan*. Kolkata: Ananda Publishers.
- [102] Banerji, R.D. 1919. *The Origin of the Bengali Script*. Kolkata. New Delhi; Asian Educational Services; 2003 reprint.
- [103] Chatterji, S.K. 1926. *The Origin and Development of the Bengali Language*. Calcutta: Calcutta University Press.
- [104] ----- . 1939. *Bhasha-prakash Bangala Vyakaran (A Grammar of the Bengali Language)*, Calcutta: University of Calcutta.
- [105] Hai, Muhammad Abdul. 1964. *Dhvani Vijnan O Bangla Dhvani-tattwa (Phonetics and Bengali Phonology)*, Dhaka: Bangla Academy.
- [106] Jha, Subhadra. 1958. *The Formation of Maithili*. London: Luzac & Co.
- [107] Kostic, Djordje; Das, Rhea S. 1972. *A Short Outline of Bengali Phonetics*, Calcutta: Statistical Publishing Company.
- [108] Majumdar, R.C. 1971. *History of Ancient Bengal*, Calcutta: G. Bhardwaj.
- [109] Mazumdar, Bijaychandra. 1920/2000. *The History of the Bengali Language* (Repr. Calcutta, 1920. ed.). New Delhi: Asian Educational Services.
- [110] Pandey, Anshuman. 2001. Proposal to Encode the Tirhuta Script in ISO/IEC 10646.

- [111] Pal, Palash Baran. 2001. *Dhwanimala Barnamala*. Kolkata: Papyrus.
- [112] -----, 2007. 'Bangla Harapher Panch Parba'. In Swapan Chakraborty, ed. *Mudraner Sanskriti O Bangla Boi*. Kolkata: Ababhas.
- [113] Ross, Fiona. 1999. *The Printed Bengali Character and its Evolution*. London: Curzon.
- [114] Shastri, Mahamahopadhyay Hara Prasad. 1916. *Hājār Bacharēr Purāṇa Bāṅgālā Bhāṣāy Baudhdha Gān ō Dōhā*. Calcutta: Bangiya Sahitya Parishat.
- [115] Singh, Udaya Narayana (Jointly Maniruzzaman). 1983. *Diglossia in Bangladesh and language planning*. Calcutta: Gyan Bharati. 214 pp.
- [116] -----, 1987. *A Bibliography of Bengali Linguistics*. Mysore: CIIL. xii+316 pp.
- [117] -----, 2017. (with Rajib Chakraborty, Bidisha Bhattacharjee & Arimardan Kumar Tripathy) *Languages and Cultures on the Margin: Guidelines for Fieldwork on Endangered Languages*. Mimeo. Centre for Endangered Languages, Visva-Bharati.
- [118] -----, 1980. Scriptal choice and spelling reform: An essay in language and planning. *Journal of the M.S. University of Baroda*, Social Science Number, 29.2 : 173-186. A modified version reprinted E. Annamalai, Bjorn Jernudd and Joan Rubin, eds. *Language Planning: Proceedings of an Institute*. Mysore: CIIL. 405-417.
- [119] Sripantha. 1996. *Jakhan Chapakhana Elo*. Kolkata: Paschim-Banga Bangla Academy.
- [120] Sur, Atul. 1986. *Bangla Mudraner Dusho Bachar*. Kolkata: Jijnasa.
- [121] Script Behaviour for Bengali, Version 1.1, TDIL and C-DAC Pune.
- [122] Bora, Mahendra. 1981. *The Evolution of Assamese Script*. Jorhat: Assam Sahitya Sabha.
- [123] Proposal to Encode the Tirhuta Script in ISO/IEC 10646, <http://www.unicode.org/L2/L2011/11175r-tirhuta.pdf> accessed on 25.11.2017
- [124] Ethnologue, Assamese in the Language Cloud, <https://www.ethnologue.com/cloud/asm> accessed on 25.11.2017
- [125] Bengali alphabet for Manipuri, found in Ethnologue, Manipuri (Meeteilon/Meithei), <https://www.omniglot.com/writing/manipuri.htm> accessed on 20.10.2019
- [126] Wikipedia, Bengali alphabet, [https://en.wikipedia.org/wiki/Bengali\\_alphabet](https://en.wikipedia.org/wiki/Bengali_alphabet) accessed on 25.11.2017
- [129] Omniglot, Sylheti, <http://www.omniglot.com/writing/syloti.htm> accessed on 10.5.2018
- [130] Wikipedia, Bishnupriya Manipuri language, [https://en.wikipedia.org/wiki/Bishnupriya\\_Manipuri\\_language](https://en.wikipedia.org/wiki/Bishnupriya_Manipuri_language) accessed on 25.11.2017

- [131] The EMILLE/CIIL Corpus, <http://metashare.elda.org/repository/browse/the-emilleciil-corpus/abdd35c8de6f11e2b1e400259011f6ea6bce74d38dbb42d881da76c64a6adb20/> accessed on 10.5.2018
- [132] The EMILLE/CIIL Corpus, [http://catalog.elra.info/product\\_info.php?products\\_id=696](http://catalog.elra.info/product_info.php?products_id=696) accessed on 10.5.2018
- [133] Bangla Language & Script, [https://www.isical.ac.in/~rc\\_bangla/bangla.html](https://www.isical.ac.in/~rc_bangla/bangla.html) accessed on 10.5.2018
- [134] Sarkar, Pabitra. 1992. *Bangla Banan Sanskar: Samasya o Sambhabana*. Kolkata: Chirayata Prakashan.
- [135] Sarkar, Pabitra. 1993. Bangla Bhashar Yuktabyanjan. *Bhasha* 1.1: 23-45.
- [136] Dash, Niladri Shekhar and B.B.Choudhuri. 1998. Bangla Script: A Structural Study. *Linguistics Today* 1.2: 1-28. Also available at [https://www.academia.edu/9967428/Bangla\\_Script\\_A\\_Structural\\_Study](https://www.academia.edu/9967428/Bangla_Script_A_Structural_Study)
- [137] Dani, Ahmed Hasan. (1957) 'Srihaṭṭa-Nāgarī Lipir Utpatti o Bikāś.' Bangla Academy Patrika (Dhaka), Vol 1.2. (Bhadra-Agrahayan, 1364 Bangabda Number).pg 1.
- [138] Wikipedia, Sylheti Nagari, [https://en.wikipedia.org/wiki/Sylheti\\_Nagari](https://en.wikipedia.org/wiki/Sylheti_Nagari) accessed on 19.5.2018
- [139] Furui, Ryosuke. (2015). 'Variegated Adaptations: State Formation in Bengal from the Fifth to Seventh Century', in Bhairabi Prasad Sahu & Hermann Kulke, eds. *Interrogating Political Systems: Integrative Processes and States in Pre-Modern India*. Chapter 9. Pp 255-73. New Delhi: Manohar.
- [140] Ferguson, Chares A. and Munier Chowdhury. (1960) 'Phones of Bengali', *Language*, Vol. 36, No. 1, pp. 22-59.
- [141] Shahidullah, Muhammad. (2007) *Buddhist Mystic Songs*. Dhaka: Mowla Brothers.
- [142] Ray, Punya Sloka. (1966) *Bengali Language Handbook*. Washington.
- [143] Hai, Muhammad Abdul. (1960) *A phonetic and phonological study of nasals and nasalization in Bengali*. Dhaka: University of Dhaka.

- [144] Unicode Consortium, Proposal Summary Form to Accompany Submissions for Additions to the Repertoire of ISO/IEC 10646 / UNICODE, <https://www.unicode.org/L2/L2002/02387r-syloti-form.pdf> accessed on May 21, 2018
- [145] Wikipedia, Ol Chiki (Unicode block), [https://en.wikipedia.org/wiki/Ol\\_Chiki\\_\(Unicode\\_block\)](https://en.wikipedia.org/wiki/Ol_Chiki_(Unicode_block)) accessed on May 21, 2018
- [146] Bangla Script, [http://www.bangladesh2000.com/bd/bangla\\_script.html](http://www.bangladesh2000.com/bd/bangla_script.html) accessed on May 21, 2018
- [147] Bhattacharya, Ashutosh ed. (1942) *Gopichandrer Gan*, Calcutta: Calcutta University.
- [149] Das, Sisir Kumar. (1975) *Sahibs and Munshis: An Account of the College of Fort William*. Calcutta.
- [150] Islam, Rafiqul, Pabitra Sarkar, Mahbubul Haq & Rajib Chakraborty (eds.). (2014) *Bangla Academy Promito Bangla Byabaharik Byakaran (A Functional Grammar of Standard Bangla)*. Dhaka: Bangla Academy.
- [151] Sarkar, Pabitra. [2013] 'Bangla Spelling Reform: the Long and Short of It'. *Bangla Journal* 19: 215-232.
- [152] Bangla Academy. (2012) *Bangla Academy Promito Bangla Bananer Niyam (Standard Bangla Spelling as adopted by Bangla Academy)*. Dhaka: Bangla Academy.
- [153] Sarkar, Pabitra & Rajib Chakraborty. 2018. "What has happened So Far In terms of Script Reforms". Paper presented at the Face to Face meeting jointly held by the Bangla Academy, Dhaka & ICANN at Bangla Academy, Dhaka on 10.07.2018.
- [154] The Unicode Consortium. 2018. *The Unicode® Standard Version 11.0 – Core Specification*. Chapter 12, P. 473.

## 10. Appendix- I

### 10.1 Augmented Backus-Naur Formalism (ABNF)

The Augmented Backus-Naur Formalism (ABNF) is generic in nature and when applied to a specific language/script, certain restriction rules apply. In other words, in a given language some of the Formalism structures do not necessarily apply. To take care of such cases restriction rules are set in place. These restrictions will help to fine-tune the ABNF.

In case of Bangla<sup>13</sup> in particular the following rules apply:

1. *Khaṇḍa ta* (ঙ) is NOT allowed at the beginning of an IDN label. The same applies to ঞ and the velar nasal ঙ in the Bangla Scheme of five-fold ‘*varga*’ (as defined under Table 5). Moreover, Bangla does not allow *ya* (য) in the beginning of a word either but we can cite a couple of native examples, for example, the word য়াৰ্ৰডো (yæbbɔRo) from the poem ‘Lichuchor’ written by Kazi Nazrul Islam. However, there are instances of it being used in names, mostly of foreign origin such as Yaqub which may be written with *ya* (য) in the beginning as in য়াকুব). In very recent times, while transliterating some Chinese and Japanese names in Bangla, one does come across the possibility of *Khaṇḍa ta* (ঙ) followed by *sa* (স) in the beginning of a word, for example ত্সেরিং (Tsering).
2. CH can come with Khaṇḍa Ta in only the case where C is *ra* (র) (09B0).  
঳ as in ভৰ্গসনা
3. Only following combinations with VHCM will be allowed.
  - অ্যা (together pronounced as æ) as in অ্যাসিড (acid)
  - এ্যা (together also pronounced as æ) as in এ্যাসিড, এ্যাসোসিয়েশন (acid, association)

### 10.2 ‘Sylheti Nāgarī lipi’ or ‘Siloṭi’

This version of Bangla script resembles the ‘Kaithī’ script (ISO 12954) used by the Accountants (perhaps by the Kāyastha community) in Eastern Uttar Pradesh and Bihar – widely in use during the 1880s. There were several other names of Sylheti

---

<sup>13</sup> This section specifically takes up issues of restrictions pertaining to Bangla (Bengali) language. Assamese and Maṇipuri have not been covered in this section.

Nāgarī or Siloti (129) – such as ‘Jālālābāda Nāgarī’, ‘Fula (flower) Nāgarī’, ‘Muslim Nāgarī’, or ‘Muhāmmad Nāgarī’. It is said that Shāh Jālāla had brought the script with him in 13th-14th Century in Sylhet (138), although some suggested that it was an invention by the Afghan rulers of Sylhet (137). Some ascribe the credit to the Buddhist Bhikkhus from Nepal. Purely for historical reasons, the details of the script with 32 symbols are reproduced here (138):

Siloti	Bengali	Unicode (Hex)	Siloti	Bengali	Unicode (Hex)	Siloti	Bengali	Unicode (Hex)
𑒀	অ	A800	𑒁	ঈ	A811	𑒀 + 𑒁 = 𑒂	অ + ঈ = আ	A823
𑒂	ই	A801	𑒃	উ	A812	𑒀 + 𑒃 = 𑒄	ক + ঊ = কি	A824
𑒀 + 𑒅 = 𑒆		A802	𑒄	ঢ	A813	𑒀 + 𑒅 = 𑒆	ক + ঋ = কু	A825
𑒇	ঊ	A803	𑒅	ভ	A814	𑒀 + 𑒇 = 𑒈	ক + ঌ = কে	A826
𑒈	এ	A804	𑒆	থ	A815	𑒀 + 𑒉 = 𑒊	ক + ঒ = কো	A827
𑒉	ও	A805	𑒇	দ	A816	𑒋		A828
𑒀 + 𑒋 = 𑒌	ক	A806	𑒈	ধ	A817	𑒌		A829
𑒍	ক	A807	𑒉	ন	A818	𑒍		A82A
𑒎	খ	A808	𑒊	প	A819	𑒎		A82B
𑒏	গ	A809	𑒋	ফ	A81A	𑒏		
𑒐	ঘ	A80A	𑒌	ব	A81B	𑒀 + 𑒐 = 𑒑	ক + ত = ক্ত	
𑒀 + 𑒑 = 𑒒	ক + ঞ = কং	A80B	𑒍	ভ	A81C	𑒀 + 𑒑 = 𑒑	ক + স = ক্স	
𑒓	চ	A80C	𑒎	ম	A81D	𑒀 + 𑒑 = 𑒑	ক + ল = ক্ল	
𑒔	ছ	A80D	𑒏	র	A81E	𑒀 + 𑒑 = 𑒑	ক + র = কলা = ক্র	
𑒕	জ	A80E	𑒐	ল	A81F	𑒀 + 𑒑 = 𑒑	ক + ক = ক্ক	
𑒖	ঝ	A80F	𑒑	ড	A820			
𑒗	ট	A810	𑒒	স	A821			
			𑒓	হ	A822			

Table 17 – The Script Table of Sylheti Nāgarī or Siloti

### 10.3 Confusable code points

The following code points were analysed and concluded that they are either (a) distinguishable or (b) confusable but not enough to be defined as variant code points.

#### 10.3.1 Bangla and Nāgarī or Devanāgarī

Bangla	Devanāgarī	NBGP Decision
◌ः U+0983	◌ः U+0903	Confusable
ॐ U+0993	ॐ U+0909	Confusable
घ U+0998	घ U+0918	Confusable
ँ U+0981	ँ U+0945	Confusable

Table 18: Bangla and Devanāgarī confusable code points

### 10.3.2 Bangla and Gurmukhi

Bangla	Gurmukhi	NBGP decision
ਬ U+0998	ਬ U+0A2C	Confusable
ੈ U+0981	ੈ U+0A71	Confusable

Table 19: Bangla and Gurmukhi confusable code points

Bangla	Gurmukhi	NBGP decision
ੳ U+0993	ੳ U+0A24	Distinguishable
ਯ U+09B6	ਯ U+0A05	Distinguishable
ਯ U+09AE	ਯ U+0A2E	Distinguishable
ਯਾ U+09AC and U+09BE	ਯਾ U+0A17	Distinguishable

Table 20 – Bangla and Gurmukhi distinguishable code points

### 10.3.3 Bangla and Oriya (Odia)

Bangla	Oriya (Odia)	NBGP Decision
ੳ U+0993	ੳ U+0B13	Confusable

Table 21 – Bangla and Oriya distinguishable code points

Bangla	Oriya (Odia)	NBGP Decision
ਬ U+0998	ਬ U+0B38	Distinguishable

Table 22 – Bangla and Oriya distinguishable code points

## 11. Appendix -II

### Bengali consonants and their allographs

Consonants	Phonetic Value	Allographs	
		Clusters	Transparent Form (Bangla Akademi font)
প	/p/	প্ত (প্+ত), প্ন (প্+ন), প্প (প্+প), প্য (প্+য), প্র (প্+র), প্ল (প্+ল), প্স (প্+স)  স্প/স্প (স্+প), ল্প (ল্+প)	
ফ	/p <sup>h</sup> /	ফ্র (ফ্+র), ফ্ল (ফ্+ল)  স্ফ/স্ফ (স্+ফ)	
ব	/b/	ভ্র (ব্+জ), ব্দ (ব্+দ), ব্ধ (ব্+ধ), ব্ব (ব্+ব), ব্য (ব্+য), ব্র (ব্+র), ব্ল (ব্+ল), বভ (ব্+ভ)  স্ব (স্+ব), হ্ব (হ্+ব)	ব্ধ (ব্+ধ)  হ্ব (হ্+ব)
ভ	/b <sup>h</sup> /	ভ্য (ভ্+য), ভ্র (ভ্+র), ভ্ল (ভ্+ল)	
ত	/t/	ত্ব (ত্+ত), ত্য (ত্+ত্+য), ত্ব (ত্+ত্+ব), ত্থ (ত্+থ), ত্ত্ব (ত্+ন), ত্য (ত্+য), ত্ম (ত্+ম), ত্ম্য (ত্+ম্+য), ত্ব (ত্+ব), ত্র (ত্+র)  প্ত (প্+ত), ত্ত (ক্+ত), ত্ত্ব (ক্+ত্+ব), ত্ত্ব (ন্+ত), ত্ত্ব (ন্+ত্+ন্+য), ত্ত্ব (স্+ত্+র) There is a marked form of ত্+্=ৎ, ত্ (ন্+ত্/ৎ)	ত্ত্ব (ক্+ত)

Consonants	Phonetic Value	Allographs	
		Clusters	Transparent Form (Bangla Akademi font)
থ	/t <sup>h</sup> /	থ্য (থ+য), থ্র (থ+র) স্থ (স্+থ), থ্ (ত্+থ), থ্ (ন্+থ)	স্থ (ন্+থ), স্থ (স্+থ)
দ	/d/	দ্র (দ+গ), দ্র (দ+ঘ), দ্র (দ+দ), দ্র (দ+ধ), দ্য (দ+য), দ্র (দ+ব), দ্র (দ+ভ), দ্র (দ+র) দ্র (ব+দ), দ্র (ন্+দ), দ্র (ন্+দ+র), দ্র (র+দ+র)	দ্র (দ+গ), দ্র (দ+ধ)
ধ	/d <sup>h</sup> /	ধ্র (ধ+ন), ধ্র (ধ+ম), ধ্য (ধ+য), ধ্র (ধ+র) ধ্র (গ্+ধ), ধ্র (দ+ধ), ধ্র (ব্+ধ), ধ্র (ন্+ধ)	ধ্র (গ্+ধ), ধ্র (দ+ধ), ধ্র (ব্+ধ), ধ্র (ন্+ধ)
ট	/t/	ট্র (ট+ট), ট্র (ট+য), ট্র (ট+ব), ট্র (ট+র) ক্ট (ক্+ট), ট্ট (স্+ট)	
ঠ	/t <sup>h</sup> /	ঠ্য (ঠ+য) ঠ্র (গ্+ঠ), ঠ্র (স্+ঠ)	
ড	/d/	ড্র (ড+ড), ড্র (ড+য), ড্র (ড+র)	
ঢ	/d <sup>h</sup> /	ঢ্র (ঢ+য) ঢ্র (গ্+ঢ)	
ঢ	/tʃ/	ঢ্র (ঢ+ঢ), ঢ্র (ঢ+ঢ), ঢ্র (ঢ+ঢ+র), ঢ্র (ঢ+ঞ), ঢ্র (ঢ+য) ঞ (ঞ+ঢ), ঢ্র (শ্+ঢ)	ঞ (ঞ+ঢ)

Consonants	Phonetic Value	Allographs	
		Clusters	Transparent Form (Bangla Akademi font)
ছ	/tʃʰ/	ছ (ছ+র)  চ্ছ (ঢ+ছ), ঞ (ঞ+ছ), শ্ছ (শ+ছ)	ঞ্ছ (ঞ+ছ)
জ	/dʒ/	জ (জ+জ), জ্জ (জ+জ+ব), জ্জা (জ+ঝা), জ্জা (জ+ঞ), জ্যা (জ+য), জ্র (জ+র)  জ্ঞ (ঞ+জ)	জ্ঞ (ঞ+জ)
ঝ	/dʒʱ/	(not privileged enough to have clusters as a first member)  জ্ঝ (জ+ঝা), ঞ্ঝ (ঞ+ঝা)	
ক	/k/	ক (ক+ক), কৈ (ক+ট), ক্ত/ক্ত (ক+ত), ক্ত্র (ক+ত+র), ক্ত্ব (ক+ত+ব), ক্ক (ক+ন), ক্ক (ক+ব), ক্ক (ক+ম), ক্য (ক+য), ক্র (ক+র), ক্ষ (ক+ষ), ক্ক (ক+ষ+ণ), ক্ক (ক+ষ+ম), ক্ক (ক+ষ+ব), ক্ক (ক+ষ+য), ক্ক (ক+স)  ক্খ (ঙ+ক), ক্ক (স+ক+র)	ক্ক (ক+ত), ক্ক (ক+ত+র), ক্ক (ক+ত+ব), ক্ক (ক+র)  ক্ক (ঙ+ক), ক্ক (স+ক+র)
খ	/kʰ/	(not privileged enough to have clusters as a first member)  ক্খ (ঙ+খ)	

Consonants	Phonetic Value	Allographs	
		Clusters	Transparent Form (Bangla Akademi font)
গ	/g/	গ (গ্+গ), গ্ (গ্+দ), গ্ (গ্+ধ), গ্ (গ্+ন), গ্ (গ্+ব), গ্ (গ্+ম), গ্য (গ্+য), গ্র (গ্+র), গ্ন (গ্+ল)  ঙ (ঙ্+গ), ঙ্ (র্+ঙ্+গ)	গ্ (গ্+ধ)  জা (ঙ্+গ), জা (র্+ঙ্+গ)
ঘ	/g <sup>h</sup> /	ঘ (ঘ্+ন), ঘ্য (ঘ্+য), ঘ্র (ঘ্+র)  ঞ (ঙ্+ঘ)	
ঞ	This letter does not have any particular phonetic value, but mostly pronounced as /n/.	ঞ (ঞ্+চ), ঞ্ (ঞ্+ছ), ঞ (ঞ্+জ), ঞ্ (ঞ্+ঝ)  ঞ (জ্+ঞ),	ঞ (ঞ্+চ), ঞ্ (ঞ্+ছ), ঞ (ঞ্+জ), ঞ্ (ঞ্+ঝ)
ণ	/n/	ণ্ট (ণ্+ট), ণ্ (ণ্+ঠ), ণ (ণ্+ড), ণ্ (ণ্+ড্+র), ণ (ণ্+ঢ), ণ (ণ্+ণ), ণ্য (ণ্+য), ণ্ (ণ্+ব)  ক্ষ (ক্+স্+ণ), ঞ (স্+ণ), হ্ (হ্+ণ)	ণ (ণ্+ড), ণ্ (ণ্+ড্+র)  ন্ন (স্+ণ)
ঙ/ং	/ŋ/	ঙ (ঙ্+ক), ঙ্ (ঙ্+ক্+র), ঙ্ (ঙ্+খ), ঙ (ঙ্+গ), ঙ্ (ঙ্+ঘ), ঙ্ (ঙ্+ক্+ষ), (In some contexts ঙ is replaced by ং )  কং, অং	ঙক (ঙ্+ক), জা (ঙ্+গ), ঙ্ (ঙ্+ঘ)

Consonants	Phonetic Value	Allographs	
		Clusters	Transparent Form (Bangla Akademi font)
ম	/m/	ম্ম (ম্+ল), ম্প (ম্+প), ম্প্র (ম্+প্+র), ম্ম (ম্+ভ), ম্ম (ম্+ভ্+র), ম্ম (ম্+ম), ম্ম (ম্+র),  ম্ম (ভ্+ম), ম্ম (ধ্+ম), ম্ম (হ্+ম), ম্ম (ক্+স্+ম)	ম্ম (হ্+ম)
ন	/n/	ন্ট (ন্+ট), ন্ট্র (ন্+ট্+র), ন্ঠ (ন্+ঠ), ন্ড (ন্+ড), ন্ট্র (ন্+ড্+র), ন্ত (ন্+ত), ন্ম (ন্+ত্+র), ন্ম (ন্+ত্+র+য), ন্ধ (ন্+থ), ন্দ (ন্+দ), ন্দ্র (ন্+দ্+র), ন্ম (ন্+ধ), ন্ম (ন্+ধ্+র), ন্ম (ন্+দ+ব), ন্ম (ন্+ন), ন্ম (ন্+ম), ন্ম (ন্+য), ন্ম (ন্+স)  ন্ম (হ্+ন)	ন্ম (ন্+থ), ন্ম (ন্+ধ), ন্ম (ন্+ধ্+র)
শ	/ʃ/	শ্চ (শ্+চ), শ্ছ (শ্+ছ), শ্ম (শ্+ন), শ্ম (শ্+ম), শ্র (শ্+র), শ্ম (শ্+ল), শ্য (শ্+য)	
ষ	/ʃ/	ষ্ক (ষ্+ক), ষ্ট (ষ্+ট), ষ্ঠ (ষ্+ঠ), ষ (ষ্+গ), ষ্প (ষ্+প), ষ্প্র (ষ্+প্+র), ষ্ফ (ষ্+ফ), ষ্ট (ষ্+ট), ষ্ট্র (ষ্+ট্+র), ষ্ঠ (ষ্+ঠ), ষ (ষ্+গ), ষ্য (ষ্+য)  ষ্ক (ক্+ষ), ষ্ক (ক্+ষ্+গ), ষ্ম (ক্+ষ্+ম)	ষ্ম (ষ্+গ)

Consonants	Phonetic Value	Allographs	
		Clusters	Transparent Form (Bangla Akademi font)
স	/s/ & /ʃ/	স্ক/স্ক (স্+ক), স্ট (স্+ট), স্প (স্+প), স্ফ (স্+ফ), স্ত (স্+ত), শ্ব (স্+থ), স্টে (স্+টে), স্ক (স্+ক), শ্ব (স্+থ), স্য (স্+য), স্র (স্+র), স্ল (স্+ল)  ক্স (ক্+স)	স্ম (স্+থ)
হ	/h/	হ্গ (হ্+গ), হ্ণ (হ্+ন), হ্ম (হ্+ম), হ্য (হ্+য), হ্র (হ্+র), হ্ল (হ্+ল)	হ্ম (হ্+ম)
ড়	/ɽ/	ড়গ (ড়্+গ)	
ঢ়	/ɽʰ/	(not privileged enough to have clusters)	
য	/dʒ/ The secondary symbol (allograph) jɔ-phalā has two phonetic values. When added to the initial consonant in a word, it is a vowel /æ/ (as in শ্যামল, র্যাপার, etc.). But after a non-initial consonant, it just doubles it in pronunciation (as in কার্য, ধার্য, etc.). The র্+য combination has two physical manifestations—র্য and র্য.	ক্য (ক্+য), স্য (স্+য), র্য (র্+য) [Just র্য is never used in Bangla orthography. র্যা is, but then its last two symbols, Ya-phalā ā-kāra, constitute a vowel sign, representing the vowel অ্যা.]	

Consonants	Phonetic Value	Allographs	
		Clusters	Transparent Form (Bangla Akademi font)
র	/r/	Two manifestations— i. রেফ /rep <sup>h</sup> / as the first member of a cluster, e.g., প, র, র্, র্, র্, র্ (র+ধ+ব) (earlier র্=র+দ+ধ+ব, a four-term cluster), etc. (placed over the following consonant) ii. র-ফলা /rɔ-p <sup>h</sup> ɔla/ as the second/third member of a cluster, e.g., র্, র্, etc. (placed under the consonant it follows)	
ল	/l/	ল (ল+গ), ল (ল+প), ল (ল+ব), ল (ল+ম), লট (ল+ট), লড (ল+ড), লক (ল+ক), ল (ল+গ), লদ (ল+দ), ল্য (ল+য)  ল (গ+ল), ল (ভ+ল), ল (ম+ল)	
ঃ	/h/ word finally, word medially it doubles the pronunciation of the following consonant.	অঃ, কঃ	
ঁ	/~/	অঁ, বঁ	