

Language Support

Typesetting Bengali in T_EX

Anshuman Pandey

1 Introduction

The Bengali (or *Bāṅglā* বাংলা) script is one of the thirteen primary scripts used throughout India. Like other Indic scripts, the Bengali is derived from the ancient Brahmi script. The script is intimately tied with the Bengali language, which according to the latest data from *Ethnologue*, is currently the fourth most spoken language in the world with roughly 189 million speakers [2]. The language is spoken mainly in the Indian province of West Bengal and in Bangladesh. Bengali has been the medium for many notable artists, of whom the famous *literateur* Rabindranath Tagore and the great film-maker Satyajit Ray are the best known.

2 The Script

Of all the scripts derived from Brahmi, Bengali is most closely related historically to Devanagari. The two scripts share a comparable inventory of consonant, vowel, and conjunct characters, however, aside from superficial form and design, the primary difference is the phonetic value assigned to certain characters of the Bengali script.¹

Like all Brahmi-based scripts, Bengali is technically an alpha-syllabic script. This system is based on the unit of the “graphic” syllable, or *akṣara*, which by definition always ends in a vowel. Each basic consonant character in Bengali is understood to represent the consonant modified by the inherent vowel *a*, eg. ক = *ka*.

When a consonant is modified by any other vowel, the syllable is written using a diacritic form of the vowel. For example, the syllable *kā* কা is composed of the consonant *ka* ক and the vowel *ā* া (diacritic form: া). Syllables consisting of only a vowel, or with a vowel in word-initial position, are written with the full form of the vowel.

A “graphic” syllable consisting of a sequence of consonants is written using a specific conjunct form, or *yuktākṣar* যুক্তাক্ষর. For instance, the consonant cluster *kka* (ক + ক) is written as a single grapheme, ক্ক, not laterally as কক. The latter form

¹ For further details on Bengali phonology please consult Suniti Kumar Chatterji’s *Origin and Development of the Bengali Language*, Allen & Unwin: London, 1970–72, reprint of 1926 ed.

represents *kaka*, not *kka*. Such conjuncts are relatively easy to decipher because the conjunct retains some semblance to the individual elements it comprises. However, with others decipherment is a little more daunting because the form gives no indication of its constituents: *kṣa* (ক্ + ষ) ক্ষ, *ṅga* (ঙ + ঙ) ঙ্গ, and *tra* (ত্ + র) ত্র.

The subscript character called *hasanta* (০্) is used to indicate an elision of the inherent *a* from a consonant. For example, ক্ is *k*, not *ka*. In some instances, generally when dealing with poor or limited types, the *hasanta* is used in modifying consonants to produce conjuncts. The modified consonant is then written laterally with the following consonant to produce a simplified ligature. Therefore, ক্ may theoretically be written as ক্‌ক্, but such practice is rare as traditional Bengali orthography places great importance on the proper formation of letters.

3 Short History of Bengali Typesetting

The development of modern modes of printing and typesetting Bengali coincided with the assumption of government of Bengal by the British East India Company in 1772. Seeing the need to educate its officials in the vernacular, the British sought to provide a means by which to expedite such instruction. The task was taken up by Nathaniel Brassey Halhed who wrote a book titled *A Grammar of the Bengal Language*. Printed in 1778 at Hoogly (near Calcutta), this was the first book containing Bengali characters printed with movable types. These movable types were cast by Sir Charles Wilkins.

Wilkins, considered the pioneer of Indic typesetting, not only instituted the mechanical printing of Bengali, but even trained native technicians and motivated them to apply their skills to other Indic scripts as well. One of Wilkins' students, a man by the name of Panchanana Karmakar, eventually brought forth a large inventory of Bengali types which led to further advancements in the printing of Bengali [5].

4 Bengali Fonts and Packages for T_EX

The typesetting of Bengali gained another big boost when it was introduced to electronic typesetting and publishing. Initially, due to the complex nature of Bengali conjuncts and the intricate design elements of the basic characters themselves, it was rather difficult to find a computer font containing a complete set of conjuncts for the script, and/or a typeface which was aesthetic in its display of the intricate and complex glyphs.

T_EX was first introduced to Bengali in 1992 when Avinash Chopde added a support module for the HP Softfont 'SonarGaon' (*sgaon*) to his *itrans* package. 'SonarGaon' was designed by Anisur Rahman in 1990, and was available in three weights: normal, slanted, and compressed. Although it was a decent Bengali font for the time, it was less than complete in that it lacked several standard ligatures.

To amend this deficiency, Muhammad Ali Masrour developed a package which supplemented the support for 'SonarGoan' in *itrans*. Called *arosgn*, or "Aro SonarGaon" ("more SonarGaon"), this package extended the 'SonarGaon' font by providing in a METAFONT source the glyphs missing from the parent font and a set of macros with which to access these new extensions.

The support for 'SonarGaon' and *arosgn* was later dropped from *itrans* partially because the font was retracted from the public domain. The other factor was the development by Shrikrishna Patil of a user-defined Type-1 PostScript font called "ItxBengali". This font was adapted for use with *itrans* and replaced 'SonarGoan'.²

Last, but not least, is a program called 'Bengali Writer' developed by Abhijit Das (Barda).³ 'Bengali Writer' is a text editor for the X11 windows system which allows the user to type Bengali documents and save them in T_EX format. Documents saved in T_EX format are then to be used with T_EX in conjunction with the 'Bengali Writer T_EX Interface' (*bwti*) package.

The *bwti* package consists of a beautiful Bengali METAFONT and a set of macros which facilitate the inputting of Bengali text. The input system is not extremely fluid and has low readability. It appears to be a verbatim reflection of the font encoding, and conjuncts are defined as macros. For example, with *bwti* a word like *digbijaya* দিগ্বিজয় must be input as `idi\gb jy` instead of `digbija.ya` as with the *bengali* package.

5 The bengali Package

The *Bengali for T_EX* (*bengali*) package is housed on CTAN in the `language/bengali/pandey/` directory. This package differs from those described in the above section in that it:

1. provides a means aside from *itrans* and *bwti* of typesetting Bengali
2. provides a simple, stand-alone preprocessor interface

² These and other Indic language packages are described in the article "An Overview of Indic Fonts for T_EX" [4].

³ For more information on 'Bengali Writer' and related utilities see <http://www2.csa.iisc.ernet.in/~abhij/bwti/>.

3. implements the ‘Velthuis’ transliteration scheme (see Section 5.2 for details)
4. provides a simple, macro-based method of delimiting Bengali text, similar in form and style to other Indic script packages like *Devanagari for T_EX*
5. complies with the New Font Selection Scheme (NFSS)
6. incorporates the latest version of the ‘Bengali’ METAFONT developed by Abhijit Das (Barda)

5.1 The ‘Bengali’ Font

Das designed his ‘Bengali’ METAFONT after studying the various types employed in the modern printing of Bengali books in India. Das emphasized that he did not model the font after the design and form of any particular typeface, nor did he employ any instruments in its development, but relied heavily on crude approximations made by the naked eye. Upon close inspection, the characters of Das’s METAFONT adhere to the traditional Bengali orthographic style found in many printed books, and rival many modern Bengali typefaces of the highest quality.

The beautiful ‘Bengali’ (bn) font is currently available in two shapes: normal (bnr) and slanted (bnsl). It contains an almost complete inventory of Bengali vowels, consonants, numerals, diacritics, and punctuation marks. The font does lack three rare characters: \bar{r} (long vocalic ‘r’), \bar{l} (vocalic ‘l’, resembling the numeral 9: ঙ), and \bar{i} (long vocalic ‘i’).⁴

Nonetheless, the font contains the entire repertoire of traditional Bengali conjunct and ligature forms, and also a few extra ligatures used in writing loan words. In addition to the conjunct forms given in Table 2, Bengali has a few special consonant-vowel ligatures formed from the vowel *u* উ, and in one case the vowel *r* র. These are illustrated in the chart below:

গ + উ	gu	গু	হ + উ	hu	হু
র + উ	ru	রু	ত + র + উ	tru	ত্ৰু
র + উ	ruu	রু	ন + ত + উ	ntu	নুতু
শ + উ	"su	শু	ল + গ + উ	lgu	ল্গু
হ + র	h.r	হ	স + ত + উ	stu	স্তু

⁴ The existence of these characters in the traditional Bengali syllabary is debatable. Although *The Unicode Standard, Version 2.0* does not place the long forms of these characters in the traditional ordering, it does reserve the positions they would occupy were they to be included in the proper arrangement. Nor does it list them under the heading ‘Independent Vowels’, but places them instead under ‘Generic Additions.’ Unicode also gives the diacritic form of these characters [7]. However, the ISO draft standard 15919 for the transliteration of Indic scripts does show the two long forms in the traditional arrangement [6].

With a few exceptions — namely *kra* ক্রা, *tra* ত্রা, and *bhra* ব্রা, and variants of these — Table 2 does not show conjuncts formed with *ra* রা or *ya* য়া, as most ligatures containing these two elements are produced in a relatively static vertical or lateral fashion. In the case of *ra*, the *ra-phala* (ৱ), or the ligature form of *ra* is placed beneath the consonant: প্ৰ pra and স্ৰ sra. In the case of *ya*, the *ya-phala* (৳) is placed after the consonant: প্য pya and স্য sya.

Additionally, when *ra* রা is the initial element in a consonant conjunct, it is written as a superscript diacritic called *repha*. For instance, the word *karma* is written কৰ্ম karma, with the *r* above the *ma*.

As there are no character primitives in the bn font, producing conjuncts containing non-traditional phonemes is a problem. A few of these have been accounted for, but there may be cases which have been overlooked. One solution is to use the *hasanta* to explicitly produce such characters. If the preprocessor detects a conjunct it does not recognize it will ‘create’ one by joining the full sizes with a *hasanta*. The result is not pretty, but it is a solution!

A few subtle kerning adjustments were needed to correctly align the placement of vowel diacritics below consonants. The preprocessor manages such kernings and therefore manual adjustments are not needed.

Since the font contains several complex ligatures which will produce rather unsatisfactory output at small sizes such as 8pt, font magnifications below 10pt are not supported. Therefore attempts to use bn in footnotes and with such NFSS commands as `\small` will result in a message from T_EX complaining about a missing metric file. Perhaps a solution is possible in the near future.

5.2 Transliterated Input

Bengali text is entered in transliteration. Each character in the Bengali syllabary has been assigned a corresponding value based on the Roman alphabet, or a combination of signs from the Roman alphabet. The author has adapted the ‘Velthuis’ transliteration scheme for the bengali package.

The ‘Velthuis’ scheme was developed in 1990 by Frans Velthuis as a means of providing transliterated input for his *Devanagari for T_EX* package, or *devnag*. As the Bengali syllabary resembles the Devanagari, the ‘Velthuis’ scheme is perfectly suitable to transliterate Bengali (and, for that matter, any other Indic script). It was necessary to add a few extensions to the scheme, otherwise it has remained largely unchanged from the original. The scheme as modified for Bengali is given in Table 1.

STOPS		
	C	V
ক	<i>ka</i>	<i>ka</i>
খ	<i>kha</i>	<i>kha</i>
গ	<i>ga</i>	<i>ga</i>
ঘ	<i>gha</i>	<i>gha</i>
ঙ	<i>ṅa</i>	<i>"na</i>
চ	<i>ca</i>	<i>ca</i>
ছ	<i>cha</i>	<i>cha</i>
জ	<i>ja</i>	<i>ja</i>
ঝ	<i>jha</i>	<i>jha</i>
ঞ	<i>ṅa</i>	<i>~na</i>
ট	<i>ṭa</i>	<i>.ta</i>
ঠ	<i>ṭha</i>	<i>.tha</i>
ড	<i>ḍa</i>	<i>.da</i>
ড়	<i>ṛa</i>	<i>Ra</i>
ঢ	<i>ḍha</i>	<i>.dha</i>
ঢ়	<i>ṛha</i>	<i>Rha</i>
ণ	<i>ṇa</i>	<i>.na</i>
ত	<i>ta</i>	<i>ta</i>
থ	<i>tha</i>	<i>tha</i>
দ	<i>da</i>	<i>da</i>
ধ	<i>dha</i>	<i>dha</i>
ন	<i>na</i>	<i>na</i>
প	<i>pa</i>	<i>pa</i>
ফ	<i>pha</i>	<i>pha</i>
ব	<i>ba</i>	<i>ba</i>
ভ	<i>bha</i>	<i>bha</i>
ম	<i>ma</i>	<i>ma</i>

SONORANTS		
	C	V
য	<i>ya</i>	<i>ya</i>
য়	<i>yā</i>	<i>.ya</i>
র	<i>ra</i>	<i>ra</i>
ৱ	<i>ra</i>	<i>~ra</i>
ল	<i>la</i>	<i>la</i>
ব	<i>ba</i>	<i>ba</i>
ৱ	<i>va</i>	<i>va</i>
শ	<i>śa</i>	<i>"sa</i>
ষ	<i>ṣa</i>	<i>.sa</i>
স	<i>sa</i>	<i>sa</i>
হ	<i>ha</i>	<i>ha</i>

VOWELS			
		C	V
অ		<i>a</i>	<i>a</i>
আ	া	<i>ā</i>	<i>aa</i>
ই	ি	<i>i</i>	<i>i</i>
ঈ	ী	<i>ī</i>	<i>ii</i>
উ	ু	<i>u</i>	<i>u</i>
ঊ	ূ	<i>ū</i>	<i>uu</i>
ঋ	ৠ	<i>r̄</i>	<i>.r</i>
এ	ে	<i>e</i>	<i>e</i>
ঐ	ৈ	<i>ai</i>	<i>ai</i>
ও	ো	<i>o</i>	<i>o</i>
ঔ	ৌ	<i>au</i>	<i>au</i>

NUMERALS		
	C	V
০	<i>0</i>	<i>0</i>
১	<i>1</i>	<i>1</i>
২	<i>2</i>	<i>2</i>
৩	<i>3</i>	<i>3</i>
৪	<i>4</i>	<i>4</i>
৫	<i>5</i>	<i>5</i>
৬	<i>6</i>	<i>6</i>
৭	<i>7</i>	<i>7</i>
৮	<i>8</i>	<i>8</i>
৯	<i>9</i>	<i>9</i>

SPECIAL CHARACTERS			
	C	V	
◌̣	<i>ṁ</i>	<i>.m</i>	<i>anusvāra</i>
◌̤	<i>ṁ</i>	<i>~m</i>	<i>candrabindu</i>
◌̥	<i>ḥ</i>	<i>.h</i>	<i>visarga</i>
◌̇		<i>&</i>	<i>hasanta</i>
◌̈	<i>t</i>	<i>T</i>	<i>khaṇḍa ta</i>
◌̣̣	<i>.</i>	<i> </i>	<i>dāṁṛi</i>

C CSX+ 8-bit scheme
V 'Velthuis' 7-bit scheme

Table 1: Inventory of Bengali Characters

ক + ক	kk	কক	ত + থ	tth	তথ	ম + ত + র	mbhr	মতর
ক + ট	k.t	কট	ত + ম	tn	তম	ম + ম	mm	মম
ক + ত	kt	কত	ত + ব	tb	তব	ম + গ	ml	মগ
ক + ম	km	কম	ত + ম	tm	তম	গ + ক	lk	গক
ক + র	kr	কর	ত + র	tr	ত্র	গ + গ	lg	গগ
ক + ল	kl	কল	থ + ব	thb	থব	গ + ট	l.t	গট
ক + ব	kb	কব	দ + গ	dg	দগ	গ + ড	l.d	গড
ক + য	k.s	কস	দ + য	dgh	দঘ	গ + প	lp	গপ
ক + য + গ	k.s.n	কসন	দ + দ	dd	দদ	গ + ব	lb	গব
ক + য + ম	k.sm	কসম	দ + ধ	ddh	দধ	গ + ম	lm	গম
ক + স	ks	কস	দ + ব	db	দব	গ + ল	ll	গল
গ + ধ	gdh	গধ	দ + ত	dbh	দত	শ + চ	"sc	শচ
গ + ম	gn	গম	দ + ত + র	dbhr	দতর	শ + জ	"sch	শজ
গ + ম	gm	গম	দ + ম	dm	দম	শ + ম	"sn	শম
গ + ল	lm	গল	ধ + ম	dhn	ধম	শ + ম	"sm	শম
গ + ব	gb	গব	ধ + ব	dhb	ধব	শ + গ	"sl	শগ
ঘ + ম	ghn	ঘম	ন + ত	nt	নত	শ + ব	"sb	শব
ঙ + ক	"nk	ংক	ন + ত + ব	ntb	নতব	য + ক	.sk	য়ক
ঙ + ক + য	"nk.s	ংকস	ন + ত + র	ntr	নতর	য + ক + র	.skr	য়কর
ঙ + থ	"nkh	ংখ	ন + থ	nth	নথ	য + ট	.s.t	য়ট
ঙ + গ	"ng	ংগ	ন + দ	nd	নদ	য + ঠ	.s.th	য়ঠ
ঙ + ঘ	"ngh	ংঘ	ন + দ + ব	ndb	নদব	য + গ	.s.n	য়গ
ঙ + ম	"ngm	ংম	ন + ধ	ndh	নধ	য + প	.sp	য়প
চ + চ	cc	চচ	ন + ম	nn	নম	য + ফ	.sph	য়ফ
চ + ছ	cch	চছ	ন + ব	nb	নব	য + ম	.sm	য়ম
চ + ছ + ব	cchb	চছব	ন + ম	nm	নম	স + ক	sk	সক
চ + ঙ	c~n	চং	ন + স	ns	নস	স + ক + র	skr	সকর
জ + জ	jj	জজ	প + ট	p.t	পট	স + ক + ল	skl	সকল
জ + জ + ব	jjb	জজব	প + ত	pt	প্ত	স + থ	skh	সথ
জ + ঝ	jjh	জঝ	প + ম	pn	পম	স + ট	s.t	সট
জ + ঙ	j~n	জং	প + প	pp	প্প	স + ত	st	সত
জ + ব	jb	জব	প + ল	pl	প্ল	স + ত + র	str	সতর
ঞ + চ	~nc	ংচ	প + স	ps	প্স	স + থ	sth	সথ
ঞ + ছ	~nch	ংছ	ফ + ল	phl	ফল	স + ম	sn	সম
ঞ + জ	~nj	ংজ	ব + জ	bj	বজ	স + প	sp	সপ
ঞ + ঝ	~njh	ংঝ	ব + দ	bd	বদ	স + প + ল	spl	সপল
ট + ট	.t.t	টট	ব + ধ	bdh	বধ	স + ফ	sph	সফ
ট + ব	.tb	টব	ব + ব	bb	বব	স + ব	sb	সব
ড + ড	.d.d	ডড	ব + ল	bl	বল	স + ম	sm	সম
ণ + ট	.n.t	ণট	ত + র	bhr	ত্র	স + গ	sl	সগ
ণ + ঠ	.n.th	ণঠ	ত + ল	bhl	তল	হ + গ	h.n	হগ
ণ + ড	.n.d	ণড	ম + ম	mn	মম	হ + ম	hn	হম
ণ + গ	.n.n	ণন	ম + প	mp	ম্প	হ + ম	hm	হম
ণ + ম	.nm	ণম	ম + ফ	mph	মফ	হ + ব	hb	হব
ত + ত	tt	তত	ম + ব	mb	মব	হ + ল	hl	হল
ত + ত + ব	ttb	ততব	ম + ত	mbh	মত	ড + গ	Rg	ডগ

Table 2: Supported Bengali conjuncts

কে লইবে মোর কার্য, কহে সন্ধ্যা রবি শুনিয়া জগৎ রহে নিরুত্তর ছবি । মাটির প্রদীপ ছিল, সে কহিল, স্বামি আমার যেটুকু সাধ্য করিব তা আমি । – রবিন্দ্রনাথ ঠাকুর	<pre>{\bn ke la{}ibe mor kaarya, kahe sandhyaa rabi "suni.yaa jagaT rahe niruttar chabi maa.tir pradiip chila, se kahila, sbaami aamaar ye.tuku saadhya kariba taa aami -- rabindranaath .thaakur}</pre>
---	--

Figure 1: A Poem by Rabindranath Tagore

Bengali does not distinguish between *ba* and *va*, and collapses both into *ba*. For this reason, as a general rule, all words containing *va* should be transliterated as *ba*. However, when parsing conjuncts the preprocessor will accept *va* in place of *ba*, so the word স্বামি may be transliterated as either *svaami* or *sbaami*. If the preprocessor detects *va* outside of a conjunct, it will assume that the Assamese *va* is intended, and will produce an undesired output.

Another significant point is the use of `{}` to break the lexical scan. This is important when two short vowels are encoded in succession. The brackets will prevent the preprocessor from interpreting the two short vowels as a diphthong. For example, compare লইবে `la{}ibe` with লৈবে `laibe`.

The above two points are illustrated in Figure 1. As indicated in the figure, the transliterated input text is placed within the scope delimited by `{\bn` and `}`. This allows the preprocessor to locate and properly translate the input to the appropriate font character codes.

5.3 The Macros

The initialization for the `bengali` package is managed by the `beng` style file. The style defines:

1. the macro `\bn` as the delimiter for Bengali text. The definition of `\bn` sets the current font to the `bn` family, initializes `bn` for use with NFSS, and provides an appropriate `\baselineskip` so there is adequate clearance between lines for super- and subscript characters.
2. a `bengali` counter which may be used to produce Bengali numerals for page numbering and enumerated environments.

An idiosyncrasy of the `\bn` macro is that a curly brace must immediately precede it and that a space must immediately follow it: `{\bn_`. Otherwise, the preprocessor will return an error stating it was unable to locate a valid Bengali delimiter.

The document containing Bengali text is to be considered as any other `TEX` file. As Bengali text is delimited by `{\bn text }`, any other `TEX` macros and packages may be used in the document. However,

these Bengali documents should have the extension `.bn` in the file name. This lets the preprocessor know that the file is associated with it.

Additionally, any shape- or size-changing commands such as `\large` or `\slshape` should follow the `\bn` macro. Otherwise, the declaration of the default font as `bn` will be over-written by the NFSS macro's default declaration of the Computer Modern fonts. The result will be a jumble of characters, not the intended Bengali.

5.4 The Preprocessor

Once the Bengali text has been transliterated, the file is run through a program called a preprocessor. A preprocessor is the ideal method for enabling the typesetting of such complex scripts as Bengali because it presents the user with a simple interface for character input. The user has only to be concerned with the transliteration, because the preprocessor will manage the conversion of the input text into character codes with which `TEX` is familiar.

A preprocessor called `beng` has been developed for use with the `bengali` package. It is a small program written in C and based in function on the preprocessor for Charles Wikner's `sanskrit` package. The syntax for its use is:

```
beng infile [.bn] [outfile [.tex]]
```

where *infile* is the name of the Bengali document. The `.bn` extension is optional, as is the target output file. By default the preprocessor will name the output file the same as the input, but with the extension `.tex`.

Running the preprocessor without any arguments invokes the interactive mode. The input and output filenames must then be manually entered at the prompts. The version number and other information can be obtained from the preprocessor by invoking it with `-h`. The author's email address is also given in case any problems arise with the use of the program.

5.5 A Note on Hyphenation

There is no tradition of hyphenating words in Indian orthography. In manuscripts, the use of hyphens is non-existent, and in early printed materials, hyphenation was applied arbitrarily between any *akṣara*-s when the end of a line was reached. Nowadays, printers are more keenly aware of word-breaks at the end of lines, and attempt to maintain syntactic sense when applying hyphenation.

Unfortunately, such “intelligent” hyphenation cannot be produced with the `bengali` package. When the preprocessor converts transliterated input into internal character codes, the output deviates substantially from what `TeX` would consider a ‘word’. However, if by chance a word is broken at the end of a line, `TeX` will produce a hyphen because the `bn` font possesses a hyphen character at the standard position. A hyphen may otherwise be encoded within a given text simply by typing the hyphen character: `-`.

6 Support for Assamese

The script used in the far eastern Indian province of Assam is nearly identical to the Bengali script. The Assamese (or *Asamīya* অসমীয়া) script (also the name of the associated language) differs from the Bengali in the design of two consonant characters, although the correspondence between pronunciation and script is also different in a number of respects between Assamese and Bengali. These two characters are *ra* র and *va* ব.

These characters were not part of the original `bn` font developed by Das. The author has created them based on the design found in Halhed’s Bengali grammar [3] and the descriptions given by Banerji [1].

It is interesting that these two characters are used in Halhed’s grammar. Their appearance in the book implies that they were commonly used in Bengali orthography, and that the switch from র to ব and from ব to র must have occurred within the past 200 years. Also, it is probably during this time that these earlier forms were restricted to the Assamese script.

A group led by Jugal Kalita is believed to be working on an Assamese package.⁵ However, no indication of progress has been posted at the project’s site, and as of yet, nothing has been released.

⁵ Details are available at <http://www.acsu.buffalo.edu/~talukdar/assam/language/assamlang.html>.

7 What’s Next?

7.1 Refining the font

In addition to creating a suitable boldface, Das and the author have discussed replacing the explicit conjunct glyphs with character primitives. One advantage of using primitives is the opening up of several positions for other characters. The disadvantage is replicating traditional ligature forms with primitives: ligatures can be formed laterally or by juxtaposition, but this method serves an injustice to the aesthetic of the Bengali script.

Depending on the number of character positions available after the font has been revised, new characters may be introduced to the repertoire. Possible additions may be the characters \bar{r} , \bar{l} , and \bar{v} , along with their diacritic forms (described in further detail in Section 5.1). Other possible inclusions may be the currency signs given in Unicode.

Taco Hoekwater has generously offered to produce Type-1 versions of the `bn` font. Once the font is stabilized, the conversions will be performed and made available as part of the `bengali` package.

7.2 Uniting `babel` and `bengali`

Jun Takashima has developed hyphenation patterns for Sanskrit and Kannada to be used in conjunction with Johannes Braams’s `babel` package. As the lexica of both Kannada and Bengali are heavily influenced by Sanskrit, it is feasible that either the Sanskrit or Kannada hyphenation pattern may be adapted for use with the Bengali language.

Such an adaptation would require an input encoding scheme compatible with the `babel` convention. One idea could be to modify the `beng` preprocessor to convert the ‘Velthuis’ transliterated input to this `babel`-compliant scheme.

7.3 Printing Dates in Bengali

A new method for printing dates in Bengali needs to be developed. This new method would print the date according to both the Western and traditional Bengali calendars. Both styles are commonly used throughout Bengali-speaking regions of the world.

Of the two, the Western style is easier to implement. As is expected, the `\today` macro does not produce the correct result in the Bengali environment. To overcome this, Masroor, who also developed the `aroggn` package, wrote a `LaTeX` package called `bngtoday`, intended for use with `itrans`.⁶ This package provides the macro `\BanglaToday` which gives the current Western date transliterated into Bengali. For example, June is simply *Jūn* জুন.

⁶ On CTAN at `language/bengali/bngtoday.sty`

The traditional Bengali calendar is quite different from the Western and therefore requires a bit more work for correct implementation. Due to the arrangement of the Bengali calendar, Bengali and Western months overlap. Thus, a given Western month may be known by two different names in Bengali. For example, June may either be *Jyais̥tha* জ্যৈষ্ঠ or *Āsār* আসাদ়, depending on whether the first or second half of the month is being referred to. Furthermore, জ্যৈষ্ঠ may also refer to the last half of May, and আসাদ় to the first half of July.

Additionally, in the Bengali calendar the first four days of the month have special names, and ordinal numbers used for days of the month also have distinct forms. The new package would automate the calculations needed to produce the date according to these conventions.

7.4 Implementing CSX+ in bengali

CSX+, or Classical Sanskrit eXtended+, is an 8-bit encoding scheme which parallels the convention adopted for the ISO Committee Draft 15919 [6]. Contrary to what its name indicates, CSX+ uniformly supports all Indic scripts. The CSX+ scheme is modelled after IBM Code Page 437 and occupies characters in the Upper ASCII block. Table 1 shows the CSX+ scheme as it pertains to Bengali.

Currently the `beng` preprocessor only recognizes 7-bit input in the form of the ‘Velthuis’ transliteration scheme. The author intends to extend the preprocessor to recognize the 8-bit CSX+ encoding scheme as well.

8 Notes

The author was informed of a book by Fiona G. E. Ross titled *The Printed Bengali Character* (Curzon: Richmond, 1999) which provides lucid and detailed information on Bengali orthography and typography. The author was strongly encouraged to review Ross’s work, but unfortunately, a copy of the book could not be acquired in time. It is hoped that this does not diminish the force of the article.

References

- [1] Banerji, R. D. *The Origin of the Bengali Script*. Nababharat Publishers: Calcutta, 1973. Reprint of 1973 1st ed.
- [2] Grimes, Barbara F. [ed]. *Ethnologue: Languages of the World*. Summer Institute of Linguistics, Inc. Dallas, Texas, 1996. [Note: Information on most widely spoken languages may be found at <http://www.sil.org/ethnologue/>. The list was last updated in February 1999.]
- [3] Halhead, Nathaniel Brassey. *A Grammar of the Bengal Language*. Printed at Hoogly, Bengal, 1778.
- [4] Pandey, Anshuman. “An Overview of Indic Fonts for T_EX”, *TUGboat*, **19**(2), 1998. pp. 115–120.
- [5] Priolkar, Anant K. *The Printing Press in India: Its Beginnings and Early Development* [Being a quatercentenary commemoration study of the advent of printing in India (in 1556)]. Marathi Samshodhana Mandala: Bombay, 1958.
- [6] Stone, Anthony [ed]. *ISO Committee Draft 15919: Transliteration of Devanagari and Related Scripts into Latin Characters*. Available at http://ourworld.compuserve.com/homepages/stone_catend/trdcd1c.htm.
- [7] The Unicode Consortium. *The Unicode Standard, Version 2.0*. Addison-Wesley Developers Press: Reading, Massachusetts, 1997.

◇ Anshuman Pandey
 University of Washington
 Department of Asian Languages
 and Literature
 225 Gowen Hall, Box 353521
 Seattle, WA 98195
apandey@u.washington.edu
<http://weber.u.washington.edu/~apandey/>