

Documentation in support of proposal for encoding Syloti Nagri in the BMP - N2591

Title: Documentation in support of proposal for encoding Syloti Nagri in the BMP
Doc. Type: Expert contribution
Source: James Lloyd-Williams (Sylheti Translation and Research), Sue Lloyd-Williams (Sylheti Translation and Research), Peter Constable (SIL International)
Date: 2002-11-01
References: WG2 Nxxxx (= US L2/02-387)
Distribution: WG2 members, UTC members

Abstract

The Syloti Nagri is a lesser-known Brahmi-derived script used for writing the Sylheti language. Because this script is not well known, and since very little information regarding the script has been published (particularly in English), we have prepared this document to provide complete background information regarding this script for the benefit of WG2 and UTC.

This document is in three parts: The first provides general contextual and background information regarding the script, including information on the Sylheti language, the history of the script and its usage, its relationship to other Brahmi-derived scripts, and on the current user community. The second part gives detailed information about the script relevant for encoding in the Universal Character Set. The third part gives additional information of potential use to implementers.

Most of the information provided here is provided by James and Sue Lloyd-Williams of Sylheti Translation and Research (STAR) in London. Further information regarding STAR as well as the Sylheti language and Syloti Nagri script is available online.¹

¹ See <http://www.sylheti.org.uk/>.

Part I: Background Information

In Part I, we present background information that puts the Syloti Nagri script into the context of the people and language with which it is associated, its relationship to other scripts, and its history of usage.

I.1 The Sylheti language

I.1.1 The spoken language

Sylheti is an Indo-European language spoken in the Barak Valley region of northeast Bangladesh and southeast Assam (India). It has commonly been regarded as a dialect of Bengali, with which it shares a high proportion of vocabulary (Spratt and Spratt (1987) report 70% shared vocabulary, while Chalmers (1996) reports at least 80%). However, even words counted as the same—e.g. the Sylheti ‘haf’ (“snake”) and ‘aiz’ (“today”) vs. the corresponding Standard Bengali ‘shap’, ‘aj’—are pronounced so differently as to render the two mutually unintelligible. Indeed, Bengali agencies in Bangladesh dealing with Sylhetis often have to engage interpreters.

Up to the end of British rule in 1947, the religious mix of the Sylhet region, then part of Assam, was approximately equally split between Hindus and Muslims, but since then due mainly to migration into India the part of the Sylhet region that is in Bangladesh is now over 80% Muslim. This is important to mention here, because Muslims speak a significantly different form of Sylheti to Hindus. Firstly, Muslims use a large proportion of words and phrases borrowed from Persian and Arabic, and secondly, pronunciation is often different such as the ‘k’ which Muslims usually pronounce as a rough fricative but with Hindus is usually hard. From now on we will focus on the *Muslim* Sylhetis and their language, because the Syloti Nagri script seems to have been used almost exclusively by Muslims, and the organisations that are interested in its modern use are mainly Muslim.

The Sylheti language is related to both Assamese and Bengali, but is distinct from both. Modern Sylheti appears to be identical to the language described by Grierson (1928) under the name ‘Bengali of Cachar’ and listed as language number 548; Cachar was the name of the region of Assam bordering Sylhet District to the east, now referred to as the Barak Valley. Sylheti shares many features of rural East Bengali dialects generally, and retains many words and forms which in Standard Bengali are restricted to poetry or are obsolete.

I.1.2 The name of the language

There is of course no ‘correct’ way of spelling a word transcribed from a different language and writing system. Whilst *Sylheti* is now the commonest English spelling of the language name after the accepted British spelling *Sylhet* of the geographic region, the Roman transcription of the Standard Bengali spelling of the name is *Sileti*. Some Sylhetis spell this in Bengali as equivalent to *Siloti*, which they claim is more original and authentic. In the 19th century, the British tea-planters in the area referred to the language as *Sylhetia*.² In Assam, the language is still referred to as *Srihattiya*, the name used in ancient literature. We have found that many Sylhetis, even those who do not understand Standard Bengali, call their language *Bangla*, *Bangela* or *Bangala* (i.e., *Bengali*), but others who are aware of Sylheti's distinctiveness do refer to their language as *Sylheti*. When writing English, these latter people use a range of spellings such as *Sylhetee*, *Sylety*, *Siloty*, etc. The

² Grierson (1903).

English spelling of the name of the script preferred by national writers and researchers is *Syloti Nagri*.

In this paper, we have called the language *Sylheti* and its writing system *Syloti Nagri*, as discussed further below. We have retained the spelling *Siloti Nagri* where we are showing the Roman transliteration of text in Bengali or Syloti Nagri scripts.

I.1.3 Number of speakers

A proper linguistic survey has not been carried out, at least in recent times. Published figures are at best guesswork. The figures of 5 million given by Spratt and Spratt (1987) and 7 million by Chalmers (1996) refer to Sylhetis in Bangladesh only. Our own rough estimate is 10 million, based partly on information from a number of Sylheti community leaders and writers:

(a) Sylhet Division of Bangladesh recorded 6.7 million in the 1991 census. Assuming a 2% per annum growth since then brings the population to 8.0 million in 2000, of whom, say, 7 million are Sylheti speakers.

(b) Assam, particularly the Barak Valley region around Karimganj and Silchar, may have another 2 million Sylheti speakers. Prior to 1947, Sylhet District then in the State of Assam consisted of five administrative sub-districts of which four make up the present-day Sylhet Division of Bangladesh, and the fifth, Karimganj, now falls inside Assam in India. The Barak Valley region experienced heavy immigration from Sylhet during the 19th and early 20th centuries.

(c) The 'diaspora' has at least another 1 million people: over 300,000 in the UK, possibly 500,000; USA/Canada, 150,000; Middle East, 300,000; Malaysia, Singapore, Australia, Rome and Berlin, 150,000; cities of India, especially Calcutta, at least 150,000.

I.1.4 History of migration into Sylhet

Sylhet is mentioned in written records as early as the 7th century when it was subject to the Hindu kingdom of Kamrup, but remained sparsely populated up to the 15th century as about half of the region consisted of a huge shallow lake now silted up (the petty kingdoms of the day even maintained navies), the rest being mainly jungle. Muslim rule began in 1303 with the conquest by Shah Jalal, a Sufi mystic from Yemen. Many of his companions were from Persia and Afghanistan.

The conquest of northern India by the Chaghtai Turks in the 16th century led to a further wave of immigration, as the previous conquerors of northern India who were of mainly Afghan and other Central Asian origin fled east, at first into Orissa and soon after, as Orissa in turn fell, into Sylhet:

Another very large stream of the defeated Afghans flowed from Orisa [sic], by way of the coast and East Bengal, into the province of Sylhet and greatly augmented the existing rural people of the same faith who had been converted into Islam centuries ago by Shahjalal and other missionaries of the Crescent.³

During the great upheavals of the period, people throughout North India other than the Afghans just referred to fled to Sylhet. Many Sylheti families, both Hindu and Muslim, trace their ancestry to migrants from Kashmir, UP and Bihar; among these is Sylhet's best-known mystic poet Hason Raja (1854-1922) whose ancestors migrated from Ray-Bareilly.

³ Sarkar (1948), as quoted by Qadir (1999).

As was the case with the Normans who lost their Viking tongue within a few decades of invading northern France in the 10th century, the mainly Persian-speaking migrants seem quickly to have adopted the local East Bengali dialect, particularly as the new settlers, fleeing large distances in a wartime situation, would have been almost exclusively male and therefore would have had to marry local women. Even so, censuses of Sylhet during British rule still show numbers of people, albeit small, claiming Persian as their mother tongue.

I.2 The Sylheti script—past and present

I.2.1 Name of the script

In this paper, we refer to the script as *Syloti Nagri*, in accordance with the expressed preference of the Sylot Academy. We have retained the spelling *Siloti Nagri* where we are showing the Roman transliteration of the name from Bengali or Syloti Nagri scripts.

There are, however, many alternate names that have been used for this script. For example, Bhuiya (2000) refers to the script as *Jalalavadi Nagri*, “Jalalavad” being an alternative name for Sylhet Town.

I.2.2 Origins

The exact origin of the Syloti Nagri script is unknown, and our opinions are therefore based on circumstantial evidence.

Firstly, Syloti Nagri is not derived from Bengali script, or even from Devanagari (used for writing Hindi) which is superficially more similar. However, Syloti Nagri is clearly a member of the group which includes all the major scripts in use today in northern India (other than Urdu which uses a modified Arabic script), such as Gurmukhi (Punjabi), Gujarati, Devanagari, Bengali and Oriya.

There are various opinions, summarised by Chalmers in the following words:

Various theories have been put forward. The [Syloti-] Nagari script most probably came to Sylhet with the Muslim Saint Shah Jalal [in 1303 AD] and the followers he gained in the more western parts of India. Alternatively, it may have come with the Brahmins (Hindu priests) whom Sylhet’s Hindu kings traditionally imported from North Bihar in India, with Brahmins from Gujarat, with the medieval Afghan rulers of Bengal, or with seventeenth century Bihari soldiers.⁴

While the source of the script obviously comes from Brahmic origins, it is clear that the users of this script, whose cultural roots were Perso-Arabic, were not constrained in their development and use of the script by conventions associated with Brahmic traditions. This can be seen in certain characteristics of the script: a small vowel repertoire without long / short contrast, the adaptation of a semi-vowel character to write either a semi-vowel or a vowel, a lack of a halant / virama (introduced only relatively recently, initially for purely didactic reasons), and the formation of conjuncts being affected more by writing convenience than by phonological constraints. (These aspects of Syloti Nagri script are described in detail in sections II.1.2, II.1.5, II.2 and Appendix 1.)

⁴ Chalmers (1996).

1.2.3 Relationship to other scripts

In the opinion of Qadir (1999) and of Professor Clifford Wright of SOAS (personal communication), Syloti Nagri is a form of Kaithi, a script (or family of scripts) which belongs to the main group of North Indian scripts.

Wright says:

The answer that I arrive at from your specimens [of Syloti Nagri manuscripts] seems so clear-cut, namely that all your scribes are simply writing the script that is listed in the [Linguistic Survey of India] chart as Magahi [a form of Kaithi]. But being Bengalis, they have replaced curves with angles. And being Muslims, there is some influence from Urdu spelling. I don't think poring over Kaithi usage further west, the basis of Gurmukhi, etc., would serve any purpose.

The word *Kaithi* is probably derived from *Kasthiya*, the caste of professional secular scribes spread over much of North India. Gurmukhi (Panjabi) and Gujarati are the westernmost examples of *Kasthiya*-derived scripts. The Magahi, Maithili and Bhojpuri scripts of Bihar (which are just different styles of the same Kaithi script) are claimed to be the easternmost examples, though really Syloti Nagri should receive that distinction.

Devanagari, Bengali and Oriya were all scripts mainly reserved for writing sacred texts in Sanskrit. The *Kasthiya* scripts, by contrast, though of a common origin, were primarily for secular use:

Having thus given a brief description of the Deva-nagari alphabet, I proceed to describe how the Kaithi arose side by side with it. The Deva-nagari, although an admirable alphabet in some respects, was long ago found to be too cumbrous for the common affairs of life. The pen in each letter has to be lifted three times, once for the horizontal line, once for the vertical, and at least once for the essential part of the letter. In short, it was not sufficiently cursive. There were two ways of giving it a cursive character, one of which was to preserve the frame-work, and alter the shape so that the whole could be written without lifting the pen from the paper. This was the course adopted in the Bengali and Oriya written hands. The other course was to discard as much as possible the horizontal and perpendicular lines, leaving only the essential part of the letter, and as much of the rest as could be combined by one stroke of the pen, without materially altering the shape of the letter. This was the course adopted in the Modh, Gujarati, and Kaithi alphabets. All the alphabets of this latter class show a great similarity of character. Gujarati, the most western, differs little from Kaithi, the most eastern; and a Tirhutia patwari [of eastern Bihar] finds little difficulty in reading a Gujarati book.⁵

The history of migration into the Sylhet region makes the arrival of a form of the Kaithi script from somewhere west of the region highly plausible, though precisely how or when this occurred is not clear.

Having examined all the common North Indian scripts, we have decided to show a comparison of characters in Bengali, Gujarati and Kaithi (Magahi variety) scripts in Appendix 1. Gujarati script has been chosen for this purpose as being the most westerly, and hence geographically distant, from Syloti Nagri. It is interesting that it is the only script we have found with a similar 'L' to Syloti Nagri's. Kaithi (Magahi variety) is chosen as it is the nearest geographically, and also the nearest in form that we have observed. Finally, Bengali script is included as being the other, and now the dominant, script in use in the Sylhet region. Some of the most popular works such as *Halot-un-Nobi* had editions printed in both the Bengali and the Syloti Nagri scripts. Bengali is included for these

⁵ Grierson (1899).

reasons, therefore, not for any similarity with Syloti Nagri; in fact, Bengali is perhaps the least similar to Syloti Nagri of all the North Indian scripts.

I.2.4 Uniqueness of Syloti Nagri

The North Indian scripts generally share common characteristics, and are believed to have a common origin, yet are today recognised as separate scripts. The position is analogous to the European scripts where the Roman, Greek and Cyrillic alphabets, though related, are regarded as distinct. We believe that an examination of the tables in Appendix 1, which compare Syloti Nagri with other related scripts, as well as the discussion of the behaviours of the script in Part II: will justify our assertion that Syloti Nagri, whatever its exact relationship to Kaithi and Gujarati may be, needs to be encoded as a separate script in its own right.

Kaithi, which used to be used for writing Bihari languages, is the nearest contender to being considered the ‘same’ script as Syloti Nagri. But Syloti Nagri has a smaller character set and a significantly different spelling system. For example, the Syloti Nagri symbol that is cognate with the Kaithi consonant /w/ is used as the independent form for the vowel /o/. A linguist currently working in Bihar on Bihari languages informs us that she has seen some old manuscript samples in the Kaithi script, but that it is not in use today, Devanagari being used instead. European missionaries produced a Kaithi font in the late 19th century, but it seems only to have been used for Christian literature and fell out of use long ago. Grierson (1899) says that there is no indigenous Bihari literature in the Kaithi script; this is marked contrast to Syloti Nagri, as will be seen in §I.3 below.

I.3 Existing Syloti Nagri documents

I.3.1 Sources and collections of older books and manuscripts

The District Gazetteer for Sylhet (Rizvi 1975) lists 44 different literary works in the Syloti Nagri script. The earliest (undated) is said to be a translation by Sanjay Laur of the Hindu epic “Mahabharat”, followed “in the early fifteenth century” by Bhabananda’s “Haribangsha”.⁶ This dating may be incorrect, however, as the only surviving fragments of “Haribangsha” in Syloti Nagri are from copies printed in the 19th century.

Professor Asaddor Ali has over 50 works in his private collection in Sylhet Town, and claims to know of the existence of a total of 140 surviving works, the earliest dating from 1549.⁷

Mohammed Sadique lists 47 separate books (67 copies) in his private collection in Stockholm, Sweden, of which 18 are manuscript and 29 are printed.

The Mukshada Library in Calcutta, formerly a private library but now under the care of Calcutta University, holds 27 books in the Syloti Nagri script, mainly printed.

Our own collection at the office of STAR, London, contains many items, including the following:

- “Halot-un-Nobi” (Account of the Prophet), by Sadeq Ali (1874); edition printed c.1920, 275 pp.
- “Mohobbot Nama—Yusuf Zulikha” (The Love Story of Joseph and Zulaikha), adapted from the Persian of Jami (1475 AD) by Sadeq Ali (1874); edition printed 1918, 95 pp.

⁶ Rizvi (1975).

⁷ Speech by Prof Asaddar Ali at the 8th National History Convention of the Bangladesh History Association, quoted in article by Farid Ahmed Reza in *Notun Din* Bengali newsweekly, page 25, London 9th June 2000.

- copies of several hundred fragments of manuscripts collected from villages in Sylhet (18th/19th C?)
- “Nur Poricoy” (Introduction to Light), by Afzul Shah (undated); manuscript, 139 pp.
- “Nur Nosihot” (Teachings of Light), by Syed Shahnur (1820 AD); manuscript, 420 pp.
- “Iuropio Mohajuddher Kobita—2 kondo” (Poems of the European Great War, Vol. 2), collected by Sheikh Abdul Wahid al Adiri ol Hanofi (undated); printed c.1950, 16 pp.
- “Bhed Johur” (Mystery Revealed), by Pir Mozir Uddin (1863-1933); manuscript, 130 pp.
- “Bhed Johur” (Mystery Revealed), by Pir Mozir Uddin (later expanded edition); manuscript, 193 pp.
- “Puthi Sahadote Buzurgan” (Traditions of the Elders), by Pir Mozir Uddin Ahmed; manuscript, 842 pp.
- “Moshkil Toran” (Rescue from Trouble), songs of the Sufi Pir Sitalong Shah (1806-1899); manuscript, 604 pp.
- “Jongo Nama” (Account of the War), the history of Hasan and Hussein and the massacre at Karbala, by Wahed Ali; printed, 460 pp. (first 28 pages missing).
- “Puthi Shohorchorit” (A Manual of Civilized Behaviour), by Shah Asod Ali Pir (1813-1905); printed 1878, 104 pp.
- “Silet Nagri Pohela Kitab o Doikhurar Rag” (A Sylhet Nagri Primer and Doikhura's Songs), by Md Abdul Lotif, including songs by Munshi Mobin Uddin 'Doikhura'; printed 1930, 14 pp.
- “Baula Dukhkhrit” (The sad Minstrel), collection of songs by Munshi Moshahed Ali; printed 1937, 15 pp.

Scanned images of some of the items in our collection are available online.⁸ It appears that none of these have been published in recent times, nor have any been translated.

I.3.2 Recent documents

Digital type is making possible the reintroduction of publishing in Syloti Nagri script, though this is still hindered by the lack of standard implementations. There are authors in the Sylhet region still creating manuscripts by hand. “Trust” S.M.S. Choudhury in Phulbari, Sylhet, Bangladesh have produced primers and have been reprinting older Sylheti texts using fonts that they have created. STAR and the Sylheti Language Development Project, Birmingham have been producing primers and reading materials using fonts created by STAR. Various Bengali-language newspapers in the U.K. have expressed an interest in printing articles using Syloti Nagri script, once adequate computer implementations are available.

I.4 Printing in Syloti Nagri

I.4.1 History of printing

The first Syloti Nagri print font was made in woodblock in about 1870 by Abdul Karim, who learned the printing trade in Europe in the 1860s. Later, metal typefaces were made using the same design. There were soon a total of at least 7 presses flourishing in Sylhet, Sunamganj, Shillong and Calcutta, printing books and at least one newspaper. There seems to have been a decline in reading in Syloti Nagri, leading to the presses closing down by the early 1970's. Professor Erhasuzzaman, a

⁸ See <http://www.sylheti.org.uk/>.

teacher in Comilla, Bangladesh, reportedly holds a collection of type rescued from one of these presses.

I.4.2 Computer fonts

The first Syloti Nagri computer font, “S.M.S Nagri”, was commissioned by Jalil Chowdhury in 1994 and made by one of his relatives. It has not been accepted by the Sylot Academy, Sylhet, because the shapes of some of the characters differ substantially both from those seen in manuscript samples and from the woodblock font created by Adbul Karim (see above). We have spoken with Jalil Chowdhury, and it seems that his font was designed using a limited range of poorly photocopied samples. Thus, for example, the “apostrophe” in the character ঐ /h/ is missing, making his ‘h’ identical to ঐ /p/.

An excellent DOS-based bitmapped computer font has been made by Mr. Roger Gwynn of Minchinhampton, UK, based on Karim’s design. At the time of writing, it is not known if he has yet made it available to others.

“Surma”, STAR’s first font, was produced in 1997 from scanned samples of Karim’s woodblock type, and was designed primarily for academic purposes. At the time, we were not aware of the existence of the two fonts mentioned above. A fundamental drawback of Karim’s design was that certain pairs of letters were too similar and could easily be confused. Having subsequently obtained a range of manuscripts in different styles, and samples of Karim’s later metal type, the “New Surma” Syloti Nagri font was produced in 1999, designed to overcome the reading problems encountered with the first font, and including a complete glyph set including conjunct characters and punctuation.

I.5 Current user community

I.5.1 Current literacy

As discussed in §I.1.3 above, the number of Sylheti speakers world-wide is around 10 million. Syloti Nagri was taught in schools in Sylhet until its incorporation into the new nation of Pakistan in 1947. Since then, literacy in Syloti Nagri has been in decline and, apart from a modern revival which is just beginning and will be discussed below, is restricted to academics, to those who learned to read before 1948, and to some in villages where children have learned informally from parents who own books in Syloti Nagri.

Literacy in Sylhet Division of Bangladesh was recorded at 28.2% in the 1991 census. Currently, of those who are literate, the vast majority are literate in Bengali script; there are probably no more than a few thousand who are literate in Syloti Nagri.

Bengali-script literacy in Britain is restricted to first-generation immigrants and is therefore much lower than in Sylhet, maybe as low as 5% of the total community. Literacy in English is almost non-existent in the first generation, and still poor among those born in Britain, so that acquiring literacy in Sylheti, which the entire community continues to speak, may be an attractive option for many.

I.5.2 Literacy revival in Sylhet and Britain

Photocopies of a handwritten reading primer of 16 pages, titled “Nagri Siloti” have been circulating within Sylhet for some years. These were developed by Prof. Erhasuzzaman (but some copies appear under a pseudonym), PO Box 46, Sylhet. He currently teaches at Victoria College, Comilla.

More recently, a printed reading primer “Siloti Poyla Kitab—Syloty First Book: Learn Syloty in 15 Minutes!” by Jalil Chawdhuri, was published in November 1998 by “Trust” Shah Mohibus Samad Choudhury. It is available in bookshops in Sylhet Town as well as in Bangladeshi bookshops in Britain. “Trust” Shah Mohibus Samad Choudhury has also published a selection of religious poetic texts in Siloti Nagri.

Sylheti community leaders in the city of Birmingham, which has Britain’s largest Sylheti community outside London, have set up the Sylheti Language Development Project. The first stage of the project, being the production of a series of reading primers in the Syloti Nagri script and the training of 25 Sylheti adults as literacy teachers, has been successfully completed. The second stage is projected to teach Syloti Nagri literacy in the Birmingham area, and to republish some popular Sylheti books beginning with ‘Yusuf Zulikha’ and ‘Halat-un-Nobi’, both of which has been fully retyped using our font and are awaiting publication. New books—humour, poetry, novels, children's books—are also in progress or are planned, in order to serve the new readership. Plans are being made by the Sylheti Language Development Project to sponsor Syloti Nagri literacy projects in primary schools in Sylhet Division of Bangladesh.

The Birmingham project has generated much interest, and has been reported in Bengali language newspapers both in Britain and in Bangladesh. Since 1998, there have been over 100 articles in Bengali-language newspapers and magazines in Bangladesh, the U.K. and Assam on Syloti Nagri issues, citing our work with approval. One U.K. Bengali newspaper editor has told us he plans to launch a newspaper in Sylheti using our font, and other Bengali newspapers in Britain and in Sylhet plan to include at least one page in Syloti Nagri as soon as standard computer implementations are available.

A joint project between STAR and Destiny Associates Ltd., Sylhet is underway in Sylhet Town in Bangladesh, the aim of which is to digitise the entire existing corpus of Syloti Nagri manuscripts. Most of the collections held by STAR and by Prof. Asaddar Ali have already been digitised.

In summary, although literacy in Syloti Nagri has undergone a decline, a revival is underway and gaining momentum. There is, therefore, a living community of users in need of standard computer implementations of Syloti Nagri script.

Part II: Details of the Syloti Nagri Script

II.1 Inventory of characters

II.1.1 Consonant letters

There are 27 consonant letters which represent syllables comprising the consonantal sound plus the inherent vowel /o/. These are shown in the following table, with Romanised character names.

(Doubled consonants in the Romanised names indicate retroflex characters.)

𑖀	Syloti Nagri ko	𑖄	Syloti Nagri do
𑖁	Syloti Nagri kho	𑖅	Syloti Nagri dho
𑖂	Syloti Nagri go	𑖆	Syloti Nagri no
𑖃	Syloti Nagri gho	𑖇	Syloti Nagri po
𑖄	Syloti Nagri co	𑖈	Syloti Nagri pho
𑖅	Syloti Nagri cho	𑖉	Syloti Nagri bo
𑖆	Syloti Nagri jo	𑖊	Syloti Nagri bho
𑖇	Syloti Nagri jho	𑖋	Syloti Nagri mo
𑖈	Syloti Nagri tto	𑖌	Syloti Nagri ro
𑖉	Syloti Nagri ttho	𑖍	Syloti Nagri lo
𑖊	Syloti Nagri ddo	𑖎	Syloti Nagri rro
𑖋	Syloti Nagri ddho	𑖏	Syloti Nagri so
𑖌	Syloti Nagri to	𑖐	Syloti Nagri ho
𑖍	Syloti Nagri tho		

Table 1: Syloti Nagri consonant letters

II.1.2 Vowel letters and dependent vowel signs

There are 5 independent vowel letters used to write syllables that consist of a vowel only or that begin with a vowel:

𑌵	Syloti Nagri independent a
𑌶	Syloti Nagri independent i
𑌷	Syloti Nagri independent u
𑌸	Syloti Nagri independent e
𑌹	Syloti Nagri independent o

Table 2: Syloti Nagri independent vowel letters

There are five dependent vowel symbols which do not stand alone but are attached to a consonant symbol to indicate a CV syllable with a different vowel from the inherent /o/. A further dependent vowel symbol, *dvisvara sign*, is also used to represent diphthongs with /i/ as the second element.

𑌺	Syloti Nagri dependent a
𑌻	Syloti Nagri dependent i
𑌼	Syloti Nagri dependent u
𑌽	Syloti Nagri dependent e
𑌾	Syloti Nagri dependent oo
𑌿	Syloti Nagri dvisvara sign

Table 3: Syloti Nagri dependent vowel signs

The dvisvara sign can attach to consonants to form the diphthong /oi/ with the inherent vowel, or it can also combine with dependent or independent vowels to form other diphthongs:

𑌵𑌶𑌷 koilam (inherent vowel + dvisvara)

𑌵𑌶𑌷𑌿 khailam (dependent vowel + dvisvara)

𑌵𑌶𑌷𑌿𑌸 buidhi (dependent vowel + dvisvara)

𑌵𑌶𑌷𑌿𑌹 oino (independent vowel + dvisvara)

𑌵𑌶𑌷𑌿𑌵 aibar (independent vowel + dvisvara)

Note that alternate spellings of diphthongs using the independent vowel /i/ are quite common:

𑌵𑌶𑌷𑌿 = 𑌵𑌶𑌷𑌿𑌶 koilam

𑌵𑌶𑌷𑌿𑌿 = 𑌵𑌶𑌷𑌿𑌿𑌶 khailam

𑌵𑌶𑌷𑌿𑌸 = 𑌵𑌶𑌷𑌿𑌸𑌶 buidhi

𑌵𑌶𑌷𑌿𑌹 = 𑌵𑌶𑌷𑌿𑌹𑌶 oino

ঐবার = ঐহবার aibar

Indeed, the dvisvara sign is only commonly used with the inherent vowel, and even then the alternative use of the equivalent, independent vowel /i/ is quite widespread.

It should be further noted that, in syllables with dvisvara and the dependent vowel /a/, spellings with the dvisvara positioned above the consonant and spellings with the dvisvara above the vowel are both attested:

ঐৗৗৗ = ঐৗৗৗ = ঐৗৗৗ khailam

The correct encoded representation for this diphthong follows the phonological ordering: < Syloti Nagri dependent a, Syloti Nagri dvisvara sign >. If the alternate presentation were desired, this could be implemented as a font variation (e.g. using a font feature mechanism), though we do not anticipate any real user requirement for such functionality. Such an implementation would maintain consistent spelling. On the other hand, in palaeographic studies, if there were a need to record this distinction in digital transcriptions, this could be done by using different orderings (in effect, different spellings) in the encoded representation:

ঐৗৗৗ = < ..., Syloti Nagri dependent a, Syloti Nagri dvisvara sign, ... >

ঐৗৗৗ = < ..., Syloti Nagri dvisvara sign, Syloti Nagri dependent a, ... >

It should be noted that this assumes either that the two combining marks are in the same canonical combining class, or that at least one is in class 0. (If they were in different, non-zero canonical combining classes, the two representations would be canonically equivalent.) It is recommended in any case that font implementations produce distinct renderings for these two orderings of these combining marks. Also, we will reiterate that the second sequence above should be considered extraordinary, and that normal spelling should always place the dvisvara after the vowel.

It should be noted that the vowel sign oo is not at all attested in historic use. It is a recent innovation in response to the desire by some authors to differentiate a word-final long o vowel in imperative forms of certain verbs.⁹ The design we have shown here was suggested to these authors since it could be implemented using existing fonts (as a sequence of the dependent a followed by the overstriking dependent e), and since there were similar characters in other North Indic scripts providing a precedent, and was well received. We propose this for addition to the Universal Character Set in response to the interest of current users.

II.1.3 Consonant signs

The Syloti Nagri character *anusvara* is included in the “alphabet” (consonantary), but it is written as a dependent sign, not as a standalone letter.

◌̣ Syloti Nagri anusvara sign

Table 4: Syloti Nagri consonant signs

⁹ A nearly identical innovation has occurred in Bengali writing within the past thirty years. Thus, the traditional imperative form কর (koro) is now usually spelled করো (korō), even though the pronunciation has not changed.

𐰽 rong

वाङ्मय bangla

अंगुर angur

ગુનવર્ધીએ ડામ્મ મીના ॥ વા વામ્મન હળીન મેના ❀
 જ્ઞાનાવેન માને જંગમનાપે ❀
 મહત્તી નમ્મી ભાત્ત માદ્મનુ ॥ માદ્મન નામ્મનુ માન્ને નુદ્મનુ
 મહત્તે માદ્મે જમ્મી મીત્તજાપે માપે ❀

This way of writing has potential for confusion with a commonly-used punctuation mark. In fact, in the sample shown in Figure 1, a glyph for the punctuation mark (blue highlight) has been mistakenly used for an anusvara (the correct glyph is shown with the red highlight). For this reason, we consider the spacing type form to be a compromise to deal with the limitations of technologies, and that the non-spacing anusvara is to be preferred for modern digital type.

शं shong

शंगे shonge

बांग्ला = बांग्ला bangla

बाँत्ता = < ..., Syloti Nagri dependent a, Syloti Nagri anusvara sign, ... >

वाँना = < ..., Syloti Nagri anusvara sign, Syloti Nagri dependent a, ... >

It should be noted that this assumes either that the two combining marks are in the same canonical combining class, or that at least one is in class 0. (If they were in different, non-zero canonical combining classes, the two representations would be canonically equivalent.) It is recommended in any case that font implementations produce distinct renderings for these two orderings of these combining marks. Also, we will reiterate that the second sequence above should be considered extraordinary, and that normal spelling should always place the anusvara after the vowel.

A potential co-location problem exists between anusvara and the dependent vowels i and e (i-kar and e-kar). Words containing syllables with i-kar + anusvara or e-kar + anusvara are rare. In manuscripts, various solutions for relative positioning of anusvara with these dependent vowels are used.

For e-kar + anusvara, no one solution for relative positioning is predominant. The following samples show several variations:

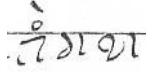


Figure 2: E-kar + anusvara: anusvara written below

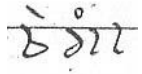


Figure 3: E-kar + anusvara: anusvara over following consonant

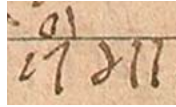


Figure 4: E-kar + anusvara: anusvara left of e-kar

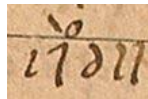


Figure 5: E-kar + anusvara: anusvara right of e-kar

Note that Figure 2 and Figure 3 show samples by the same author on two pages of the same manuscript, but with different relative positioning of anusvara and e-kar. Figure 4 and Figure 5 show samples of the very same word written by the same author on two consecutive pages of the same manuscript, yet with different relative positioning of anusvara and e-kar. For modern implementations, we consider it preferable to position e-kar to the left of the anusvara, with both over the initial consonant.

For i-kar + anusvara, the predominant solution we have encountered is to position the anusvara to the left of the i-kar, though we have also encountered instances in which the author wrote the anusvara over the following consonant. The following samples are from the same manuscript:

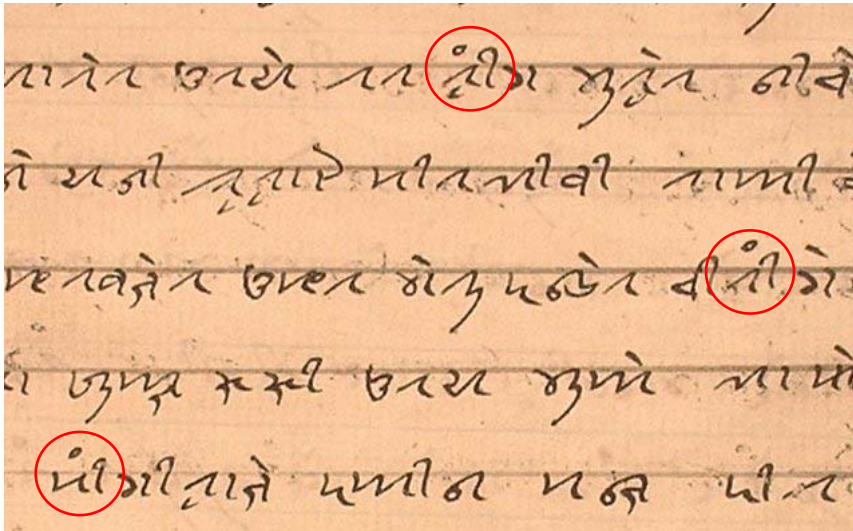


Figure 6: I-kar + anusvara: anusvara left of i-kar

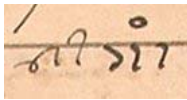


Figure 7: I-kar + anusvara: anusvara over following consonant in “bhaṅga”

For modern implementations, we consider it preferable to position anusvara to the left of i-kar, over the initial consonant.

Regardless of the relative positioning of anusvara and dependent vowels that is used, we maintain that spelling in digital texts should consistently have anusvara as syllable-final. Variant positioning could be implemented, if desired, as a font variation (e.g. using a font feature mechanism), though we do not anticipate any real user requirement for such functionality. Such an implementation would maintain consistent spelling. Because the number of variations with e-kar are more than two, and because the marked i-kar variation is so extraordinary in placing the anusvara over the following consonant, we cannot recommend any mechanism for recording these positional variations in plain text in terms of alternate character sequences.

In one poetic manuscript, the author shows a strong preference for using a conjunct न्ग = न + ग (n + g) rather than the anusvara.

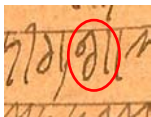


Figure 8: Conjunct n-g used in lieu of anusvara

This is the exception, however. For modern usage, it is not essential for fonts to support the n-g conjunct. (For further discussion of conjuncts, see §II.2.)

II.1.4 Punctuation and digits

There are five distinctive Syloti Nagri punctuation symbols used in poetic manuscripts and printed poetic texts:

✱	Syloti Nagri ful
◦	Syloti Nagri poetry mark 1
ঃ	Syloti Nagri poetry mark 2
৞	Syloti Nagri poetry mark 3
ঃঃ	Syloti Nagri poetry mark 4

Table 5: Syloti Nagri poetic punctuation marks

They are used in poetry to indicate various metrical units. The Devanagari double danda (U+0965) is also very commonly used in Sylheti poetry. Several different metres are used in Sylheti poetry, and different combinations of punctuation are used according to the type of metre. These punctuation marks are illustrated in the following samples.

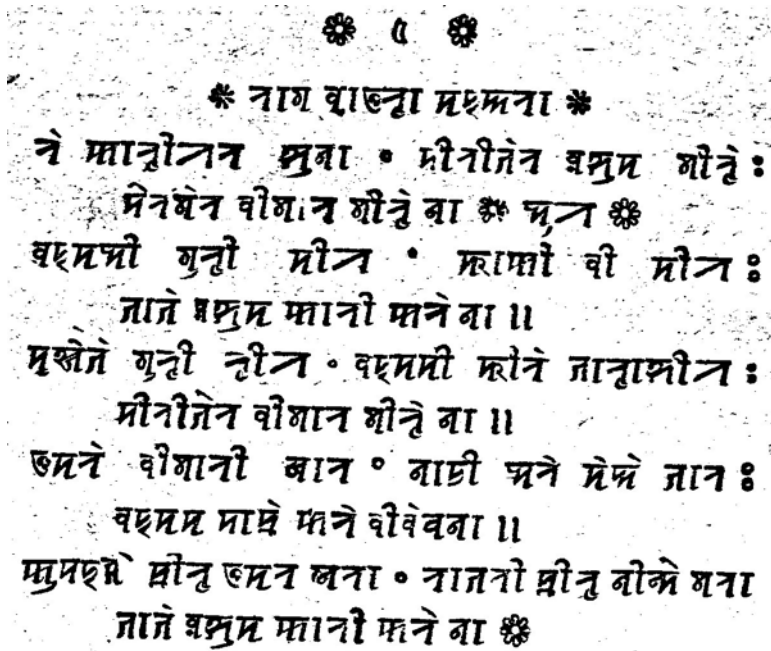


Figure 9: Sylheti tripodi poetry using ful, poetry marks 1 and 2, and double danda

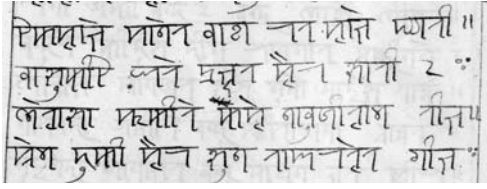


Figure 10: Sylheti poetry using poetry mark 3 and double danda

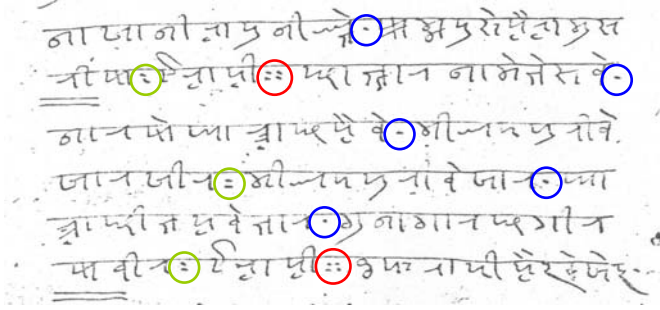


Figure 11: Sylheti poetry using poetry marks 1, 2 and 4

In Figure 11, poetry mark 1 is highlighted by a blue circle, poetry mark 2 is highlighted by a green circle, and poetry mark 4 is highlighted by a red circle.

In addition to its use as punctuation, the ful is also used as a flourish. This can be seen in Figure 9, where the ful occurs on either side of the page number at the top of the page.

The single danda appears to have been introduced in relatively recent printed texts. (At least, we have not encountered it in older documents.) It is not used for the most commonly-used metres, but is used in more complex metres. The single danda is also used as a full stop and for other purposes in modern prose.

Some texts use two styles of ful to delimit different elements, as seen in the following sample:

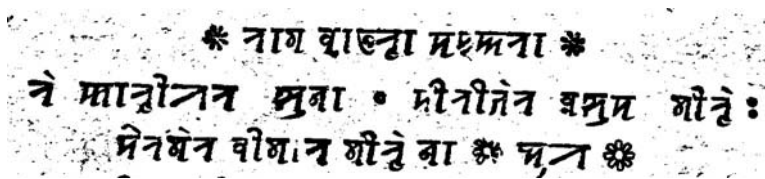


Figure 12: Distinctive fuls used for different text elements

In Figure 12, one style of ful is used as a flourish around the title, while another style of ful is used to demarcate the label *dhua* ('refrain'). We consider this a font difference only, and do not see any need to encode distinct characters.

The ful has also been used in Bengali poetry, as seen in the following sample:

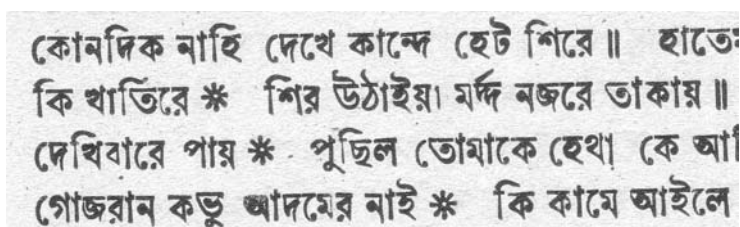


Figure 13: Ful used in Bengali poetry

Given its use in both scripts, it would be reasonable to consider encoding it as part of a Syloti Nagri block of the Universal Character Set (UCS), or as part of the Bengali block, or in some other neutral block. It is most strongly associated with Syloti Nagri script, however, thus we consider it preferable to encode it as part of a Syloti Nagri block of the UCS, and for the Syloti Nagri character to be used in Bengali documents if needed.

The ful may have been used like a Latin full stop. Historic prose literature is almost non-existent, however, and evidence from poetic texts is inconclusive. In recent literacy efforts, there have been attempts to use ful as a full stop, but these have not been well received. It appears to be preferable to use a Devanagari single danda (U+0964), as described above.

Although the ful is not used as punctuation in modern prose, we recommend that fonts always include a glyph for the ful, even if intended only for modern-text usage, as it is considered a distinctive element of Syloti Nagri writing and is still used as a flourish. The poetry marks are not essential for modern usage, though some authors still like to use them in poetry. We recommend that fonts always include poetry marks 1, 2 and 3; poetry mark 4 is not essential for modern use. We also recommend that fonts always include the single and double danda. Various Latin punctuation (full stop, comma, colon, semicolon, question mark, exclamation mark, dashes, quotation marks and parentheses) are also important for modern usage.

Syloti Nagri uses Bengali digits (U+09E6..U+09EF) and the Bengali rupee sign (U+09F3). The question as to whether there are digits specific to Syloti Nagri script is discussed in section II.1.6 below.

To accommodate the needs of speakers in the Sylheti diaspora, fonts should perhaps include various currency symbols in addition to the Bengali rupee sign (e.g., the pound sign, U+00A3, or the dollar sign, U+0024).

II.1.5 *Hasanta* (halant / virama)

Whereas a halant is a common characteristic of North Indic scripts, a corresponding character is not attested in Syloti Nagri documents prior to the mid-20th century. The Syloti Nagri *hasanta* (halant / virama) appears to have been introduced relatively recently and has not been in universal use. Professor Asaddar Ali of the Sylot Academy is of the opinion that the *hasanta* is only found in a minority of woodblock and metal-type printed texts, all late—possibly as late as the 1950s.

The manuscript primers developed by Professor Erhasuzzaman (c.f. §I.5.2) make use of a visible *hasanta* to indicate a consonant cluster rather than using conjunct forms, apparently to avoid introducing conjunct forms at an early stage of learning.

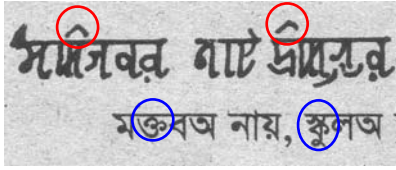


Figure 14: Hasanta used to indicate a consonant cluster (Bengali translation shown below)

In Figure 14, the Syloti Nagri hasanta is shown with red highlights, and the conjuncts in the corresponding Bengali text are shown with blue highlights.

Professor Erhasuzzaman's primers also use the hasanta to indicate dead consonants.

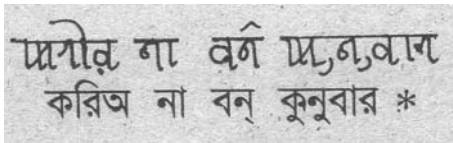


Figure 15: Hasanta used to indicate a dead consonant (Bengali translation shown below)

In current usage, a visible hasanta is still not consistently used or considered obligatory. Thus, in both historic and modern usage, dead consonants and consonant clusters are regularly written with no indication that the inherent vowel is killed. Spellings can be ambiguous with respect to pronunciation, though the appropriate reading is generally clear from context.

The hasanta has been included in all digital fonts, though there has been some variation in presentation. Whereas fonts by Roger Gwynn and STAR use the same shape as used in Professor Erhasuzzaman's primers, the font commissioned by Jalil Chowdhury uses the same shape as the Bengali hasanta.¹⁰ Also, there are slight differences with regard to positioning between the fonts developed by Roger Gwynn and those developed by STAR. Regardless of variations in presentation, we concur with Professor Asaddar Ali and his colleagues in the Sylot Academy that the hasanta should be included in the Syloti Nagri character repertoire.

The hasanta will be considered further below in relation to conjunct forms (see §II.2).

II.1.6 Characters for which there is insufficient evidence

There are various characters which have tenuous attestation. The modern handwritten primers published and distributed by Professor Erhasuzzaman (c.f. §I.5.2) describe a set of Syloti Nagri digits. In fact, these appear to be an eclectic set including Bengali digit 3, Arabic-Indic digits 1, 2, 7 and 9, Arabic (i.e. Latin) digit 8, and variations on Arabic-Indic 4 and Bengali 5. The use of Bengali digit 2 is well-attested in older Sylheti manuscripts to indicate reduplication, clearly calling into question the claim for inclusion of the corresponding Arabic-Indic digit.

We have encountered a single manuscript that includes what appear to be distinct Syloti-Nagri digits—an incomplete set used in a shopping list scrawled in the cover of a printed Syloti Nagri book. Due to the lack of clear attestation, we are not proposing a set of Syloti-Nagri digits for inclusion in the UCS at this time.

¹⁰ As mentioned in §I.4.2, Chowdhury's font has not been widely accepted as the design was based on samples of inadequate quality. We recommend that font developers use the same design for hasanta as attested in Professor Erhasuzzaman's primers.

Professor Erhasuzzaman's primers also describe four characters for indicating Arabic characters not used in Sylheti. We have seen no other instances of these, despite a heavy borrowing of Arabic words in the Syloti Nagri manuscripts. These primers also cite a candrabindu. Again, we have seen no other instances of these, nor are there nasal phonemes in Sylheti for which these might be needed.

II.2 Conjuncts

Syloti Nagri is somewhat distinct from more familiar Indic scripts in terms of its use of conjuncts. Conjuncts are not strictly correlated with the phonology being represented. They are neither necessary in contexts involving a dead consonant, nor are they limited to such contexts. It appears the convenience and practical considerations in writing have been more significant factors in conjunct formation and usage than phonology.

In terms of the phonological units involved, there are several types of conjuncts in Syloti Nagri writing, consonant-consonant conjuncts being only one.

Consonant-consonant conjuncts occur in Siloti Nagri writing, but not with high frequency. Their limited occurrence can be accounted for by the nature of the phonology of the Sylheti language, which tends to avoid consonant clusters. There is a potential for conjunct formation when a dead consonant is followed by another consonant, though, and conjuncts involving any pair of consonants are, in principle, possible.

It is not possible to state fully explicit rules as to when consonant conjuncts should or should not occur, though there are some general restrictions:

- Conjuncts are not formed across morpheme boundaries.
- Conjuncts are not generally used in Arabic and Persian loan words, although there are notable exceptions, such as **মক্কা** mokka ('Mecca') and **মহাম্মদ** mohammad ('Muhammad').
Conjuncts are common in English loan words, however, particularly those involving clusters with /r/, such as **ব্রিটিশ** brittis (pronounced /british/, 'British'), **ডিগ্রী** digri ('degree') and **ট্রেন** ttren ('train').

Certain clusters (when they do occur) are typically written with conjuncts, except across morpheme boundaries. The following are some of the more commonly encountered consonant conjunct forms:

ন /no/ + **দ** /do/ → **ন্দ** e.g., **বন্দু** bondu ('friend'), **সুন্দর** sundor ('beautiful')

ন /no/ + **ত** /to/ → **ন্ত** e.g., **পন্থা** pont ('path'), **অন্তর** ontor ('interior')

ঢ /cho/ + **ত** /tho/ → **ঢ্ত** e.g., **দুচ্ছ** duchth (pronounced /dust/, 'friend'), **বেদেচ্ছ** behechth (pronounced /behest/, 'heaven')

শ /so/ + **ত** /to/ → **স্ত** e.g., **ফীরাস্তা** phirista ('angel')

ম /mo/ + **ব** /bo/ → **ম্ব** e.g., **পেগাম্বর** pegambor ('apostle')

In addition to consonant-consonant conjuncts, vowel-consonant conjuncts also occur in Siloti Nagri writing. These are attested in historic manuscripts and modern usage, though not in wood-block and metal-type printed materials. In our research, we have most frequently encountered conjuncts involving the independent vowel **আ** /a/ as a first component, or involving consonants **ত** /to/, **র** /ro/, **ল** /lo/ or **শ** /so/ as a second component, though it seems that the possible combinations are limited

only by which characters are structured in ways that lend themselves to ligation with an adjoining character. For instance, we would not expect to encounter conjunct forms involving the independent vowel ṣ /e/ due to its structure. On the other hand, the shape of the independent vowel ā /a/ particularly lends itself to ligation with a following character. The following sample shows an a-m conjunct in which the vertical stroke of the ā /a/ is completely merged into the initial stroke of the m /mo/, forming a natural connection.

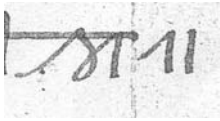


Figure 16: Vowel-consonant conjunct a-m in “amra” (‘we’)

Of course, while the consonant-consonant conjuncts discussed earlier involve a killed vowel for the letter representing the first half of the conjunct, this does not apply to vowel-consonant conjuncts.

Additional vowel-consonant conjuncts are illustrated in the following samples.

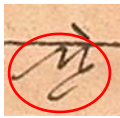


Figure 17: Vowel-consonant conjunct a-ch in “ache” (‘it is’)

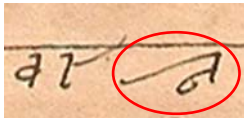


Figure 18: Vowel-consonant conjunct a-n in “boean” (‘narrative’)

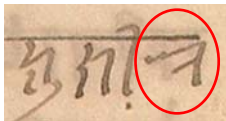


Figure 19: Vowel-consonant conjunct a-r in “husiar” (‘warning’)

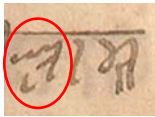


Figure 20: Vowel-consonant conjunct a-l in “alamot” (‘miracle’)

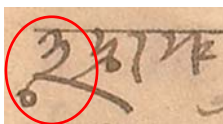


Figure 21: Vowel-consonant conjunct i-n in “inchaph” (‘justice’)

There are instances of vowel-consonant conjuncts that co-occur with the anusvara, which combines with the vowel component of the conjunct. Such a combination is illustrated in Figure 22.

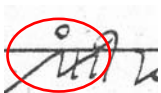


Figure 22: Vowel-consonant conjunct a-k with anusvara in “angki” (‘eye’)

In this example, the vowel ঐ /a/ is conjoined with the consonant ক /k/, but the anusvara also occurs, combining with the vowel component of the conjunct to give the pronunciation “angki”. This seems particularly unusual if one views the script from the perspective of Brahmic writing traditions, in which conjuncts are closely associated with phonology. It would not seem unusual if one were comparing with Perso-Arabic writing traditions, however, in which connection and ligation are natural correlates of fluent writing.

Another seemingly unusual use of conjuncts is found in historic manuscripts. While these conjuncts have the form of consonant-consonant conjuncts, they are used in representing CVC syllables. We shall refer to these as “false conjuncts.”

One group of false conjuncts occur in words that have cognates in Bengali, making it seem that one could account for these as the result of a phonological change that occurred in Sylheti but not in Bengali. For instance, the Sylheti word প্রীতি has a form that corresponds to the cognate Bengali word, প্রীতি /prīti/ (‘love’). Although the expected pronunciation following the conventions of the Sylheti writing system would be /priti/, the word is, in fact, pronounced /pirti/. This might be interpreted as suggesting that Sylheti spelling has retained an historic form and that pronunciation has undergone a phonological metathesis (re-ordering).

In fact, this is not likely the correct account. A different Bengali cognate suggests the more likely explanation. The Bengali word সর্প /sorppo/ (‘serpent’) is written with a *reph*. The corresponding Sylheti word is সর্প and is pronounced /sorppo/, similarly to Bengali. The conjoined /r/ in Syloti Nagri is comparable in function to the Bengali reph, though it is conjoined to the consonant of the preceding syllable rather than to the doubled consonant of the following syllable. An instance of the Sylheti word /sorppo/ taken from a manuscript is illustrated in Figure 23.

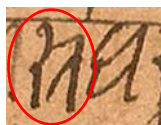


Figure 23: “False” conjunct s-r in “sorppo” (‘serpent’)

Given the nature of Sylheti phonology, which avoids initial consonant clusters, it seems likely that every consonant conjunct that has /r/ as the second element (except in loan words, particularly from English) will be a false conjunct of this type.

Among authors of poetic manuscripts, it was a common practice to use conjoined forms at the ends of lines as a device to manage line lengths, making it easier for them to justify lines and maintain visual aesthetics. These line ending conjuncts were quite typically these false conjuncts, representing CVC syllables. Conjuncts for a wide variety of syllables are attested, and it appears that the possible conjunct forms are limited only by what consonant pairs are structured in ways that make them amenable to ligation.

Such a line-ending false conjunct is illustrated in Figure 24:

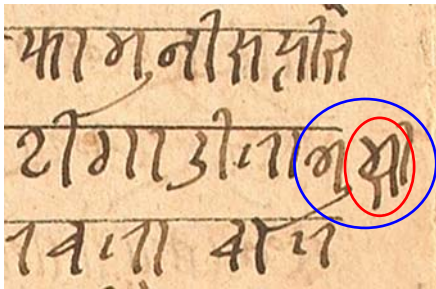


Figure 24: Line-ending “false” conjunct m-t in “bhumit” (‘in the land’)

In this figure, the red oval indicates the false m-t conjunct. It occurs in the word **ভূমি**, /bhumti/, pronounced /bumit/ and indicated with the blue oval. The pronunciation of this word is clearly indicated from the last word of the preceding line, **সহিত** /sohit/ (‘with’), with which it rhymes.

Another example appeared earlier, in Figure 20, repeated here (but highlighting a different detail):

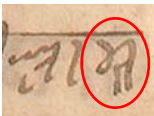


Figure 25: Line-ending “false” conjunct m-t in “alamot” (‘miracle’)

Thus, the single word in this figure contains both a vowel-consonant conjunct, a-l, and also a false conjunct, m-t.

Usage of false conjuncts to control line length is seen in the following sample, showing two successive lines ending in the rhyming words “por” and “upor”. On the first line, the author has used a false conjunct, apparently in order to fit the text into the line.

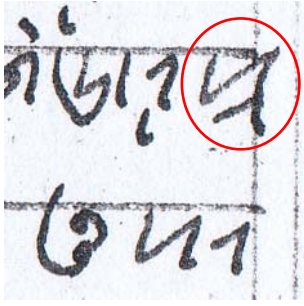


Figure 26: Rhyming lines ending in “por” (‘after’) and “upor” (‘on’), the first with a false conjunct p-r

The practice of using line-ending false conjuncts was also generalised, sometimes being applied to words not at the ends of lines.

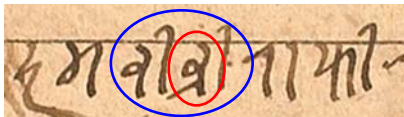


Figure 27: “False” conjunct b-r in “bibir” (‘lady’s’)

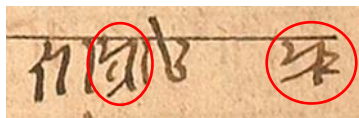


Figure 28: “False” conjuncts h-r and ph-r in “sahar uphor” (‘on the king’)

It should be noted that false conjunct forms can occur with diacritic dependent vowels. So, for example, in Figure 29, the conjunct form m-d occurs with e-kar.

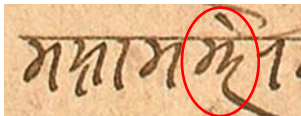


Figure 29: “False” conjunct m-d with e-kar in “mohammoder” (‘Muhammad’s’)

In this word, the e-kar is pronounced as though combining with the /d/ portion of the conjunct (the /m/ portion of the false conjunct is pronounced with the inherent vowel). In principle, the reading is ambiguous: “mohammedor” would be another possible reading for this spelling. In context, though, a reader will recognise what the intended reading is.

The following example goes one step further: two e-kar vowels combine with the conjunct form, one for each consonant component in the conjunct:

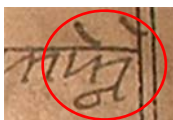


Figure 30: “False” conjunct k-n with two e-kar diacritics in “kene” (‘why’)

It should also be noted that words that occurs with false conjuncts generally can be, and usually are, written with full, non-conjunct forms. (This may not be the case, however, with words such as /sorppo/, described above.) The general optionality of these conjuncts is illustrated in Figure 31 and Figure 32, which show the same word, /iman/ ('faith'), taken from the same manuscript and written by the same scribe. The occurrence shown in Figure 31 contains a false conjunct m-n form, which the occurrence shown in Figure 32 does not.

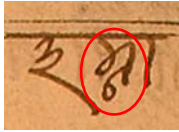


Figure 31: The word /iman/ ('faith'), written with a "false" conjunct m-n

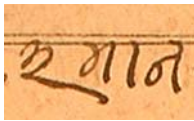


Figure 32: The word /iman/ ('faith'), written without any conjunct forms

An issue that arises in relation to the false conjuncts is how they should be represented in encoded character sequences. In particular, one of the general principles for encoding of Indic scripts is that they should be encoded in their reading ("logical") order, yet in the case of Syloti Nagri false conjuncts, the actual pronunciation is not consistent with the otherwise well-behaved conventions of Sylheti writing. We propose that Syloti Nagri false conjuncts should be encoded in the order suggested by the written form, and not by their unconventional pronunciation.

There are two other types of conjunct that occur in manuscripts, though only rarely. We have encountered two vowel-vowel conjuncts, illustrated in Figure 33 and Figure 34.

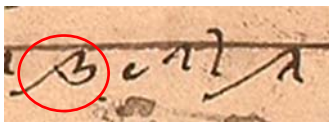


Figure 33: Vowel-vowel conjunct a-u in "auliar" ('saint's')

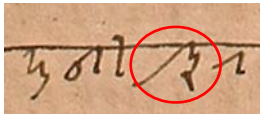


Figure 34: Vowel-vowel conjunct ai in "duniar" ('world's')

Because of the structures involved, we expect that a-e and a-o conjuncts would also be plausible, whereas all other pairings of independent vowels do not seem amenable to forming conjuncts.

Finally, we have encountered a small number of conjuncts involving a dependent vowel and consonant. For example:

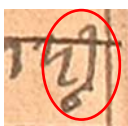


Figure 35: Conjunct dependent-í + n in “din” (‘day’)

Here, the half form of 𑂔 /n/ is attached at the bottom of the i-kar. While such forms are extremely rare, the fact that they exist at all in historic manuscripts suggests that this script has long been viewed by its users in ways that are quite atypical for Indic scripts.

II.3 Encoded representation of hasanta and conjuncts

Having reviewed the facts regarding usage of the hasanta (halant / virama) character and of conjunct forms, we must now consider how these should be represented in encoded character sequences. It is necessary to consider these together because it is typical for existing implementations of Indic scripts in Unicode to use the virama to control the occurrence of conjunct forms.

Another related factor is the inherent vowel, which may be significant for some types of processing (e.g., speech synthesis or pronunciation-based sorting). Of course, presence or absence of the inherent vowel is inversely correlated with presence or absence of virama in the typical processing model for Indic scripts.

It should be apparent from the preceding section, however, that Siloti Nagri script is atypical of Indic scripts in significant ways, particularly in relation to hasanta (virama) and in relation to conjuncts. To review:

- Hasanta was only recently introduced into the script, and even yet is used only in limited contexts.
- Conjuncts are not limited to sequences involving dead consonants; indeed, conjuncts can be formed from pairs of characters of almost any type (consonant, independent vowel, dependent vowel) and be used to represent a very wide variety of syllables.
- It is not necessary, in general, to overtly indicate dead consonants with a conjunct or explicit hasanta.

The presence or absence of overt hasanta is largely independent of the use of conjuncts. The only restriction is that an overtly-rendered hasanta cannot occur in connection with the first element of a conjunct.¹¹ Beyond that, the absence of hasanta does not imply a live consonant, and has no bearing on the occurrence of conjuncts. Similarly, the absence of a conjunct does not imply a live consonant, and has no bearing on the occurrence of hasanta.

With this in mind, we must consider possible encoding mechanisms to control occurrence of overt hasanta, and to control occurrence of conjunct forms. We see two possible encoding models. One of these is the virama model that is typical of Indic script implementations, in which mechanisms for controlling the occurrence of overt hasanta and for controlling conjunct formation would be closely

¹¹ There are plausible situations in which one may want an overt hasanta to appear indicating that the second element of a conjunct is dead. For instance, this could be used in dictionaries to indicate pronunciation in ambiguous situations, e.g. 𑂇𑂔, which has two possible readings: “bond” and “bondo”. Comparable usage in Bengali script can be found in some Bengali dictionaries.

interrelated. The other uses completely independent mechanisms for controlling the two behaviours. We will briefly describe each of these in turn.

II.3.1 Possible encoding model: viramic model

In a viramic model, a hasanta character would be present both when an overt hasanta occurs, and when conjuncts occur. Control of the two rendering alternatives, overt hasanta versus conjunct, would be provided by the presence or absence of ZERO WIDTH NON-JOINER (ZWNJ). (The use of ZWNJ to result in overt rendering of hasanta would be comparable to the encoding model used for Devanagari.)

For the most common text processes (including normal collation), it is not necessary to distinguish (non-conjoined) consonants with no vowel from (non-conjoined) consonants with inherent vowel. Thus, it might be possible in typical usage to use encoded sequences of the form $\langle \dots, C, \textit{word boundary} \dots \rangle$ or $\langle \dots, C, C, \dots \rangle$ for (non-conjoined) consonants with or without inherent vowel. If there were a need to distinguish inherent vowel from absence of vowel, however, an additional mechanism would be needed: the hasanta character would occur whenever there was a dead consonant. Control of whether the hasanta would be rendered overtly or suppressed would be provided by the absence or presence of ZERO WIDTH JOINER (ZWJ). (It seems to us that the use of ZWJ to suppress the overt rendering of hasanta would provide the closest overall similarity with the Devanagari encoding model.)

Because hasanta is used to determine conjunct formation, and because the first elements of conjuncts can be consonants, independent vowel or dependent vowels, and the syllables represented can include dependent vowels or anusvara, one implication of this model is that the Syloti Nagri virama can occur following any of these types of characters.

These mechanisms are summarised in Table 6. Note the following conventions that are used: Co = live consonant with inherent vowel, Cd = dead consonant, Vn = independent vowel, Vs = spacing dependent vowel (a-kar, i-kar), Vc = combining vowel mark (e-kar, u-kar), # = word boundary.

Note that this model is not able to distinguish between consonant-consonant conjuncts in which the first consonant element is dead and false conjuncts in which the first consonant has the inherent vowel.

Note also that false consonant-consonant conjuncts pronounced as CVC syllables with a-kar or i-kar cannot be distinguished from conjuncts representing consonant clusters followed by a-kar or i-kar, either in encoded representation or in presentation. (The alternate character sequences that might be considered for this would be the same sequencing as used for Vs + C conjuncts.) This is true of both encoding models, however.

Situation	Encoded representation	Examples (may be hypothetical)
Co# or CoC	< ...C, #... >, < ...C, C... >	𑌕𑌃𑌃 “koto” = < 𑌕, 𑌃 >
Cd# or CdC (<i>for processing where absence of vowel must be distinguish from inherent vowel</i>)	< ...C, hasanta, ZWJ, #... >, < ...C, hasanta, ZWJ, C... >	𑌕𑌃𑌃𑌃 “kantok” = < 𑌕, 𑌃, 𑌃, hasanta, ZWJ, 𑌃, 𑌕, hasanta, ZWJ >
Cd + overt hasanta + #, Cd + overt hasanta + C	< ...C, hasanta, ZWNJ, #... >, < ...C, hasanta, ZWNJ, C... >	𑌕𑌃𑌃𑌃𑌃 “kantok” = < 𑌕, 𑌃, 𑌃, hasanta, ZWNJ, 𑌃, 𑌕, hasanta, ZWNJ >
Cd + C conjunct	< ...C, hasanta, C... >	𑌕𑌃𑌃𑌃𑌃 “kantoko” = < 𑌕, 𑌃, 𑌃, hasanta, 𑌃, 𑌕 >
Vn + C conjunct	< ...Vn, hasanta, C... >	𑌕𑌃𑌃𑌃 “atiko” = < 𑌕, hasanta, 𑌃, 𑌃, 𑌕 >
Vn + Vn conjunct	< ...Vn, hasanta, Vn... >	𑌕𑌃𑌃𑌃 “aeno” = < 𑌕, hasanta, 𑌃, 𑌃 >
Vn-anusvara + C conjunct	< ...C, anusvara, hasanta, C... >	𑌕𑌃𑌃𑌃 “angki” = < 𑌕, 𑌃, hasanta, 𑌕, 𑌃 >
Vs + C conjunct	< ...Vs, hasanta, C... >	𑌕𑌃𑌃 “kir” = < 𑌕, 𑌃, hasanta, 𑌕 >
C-Vc + C false conjunct	< ...C, Vc, hasanta, C... >	𑌕𑌃𑌃 “kere” = < 𑌕, 𑌃, hasanta, 𑌕, 𑌃 >
C-Vs + C false conjunct	(indistinguishable from Cd + C conjunct followed by Vs)	
Co + C false conjunct	(same as Cd + C conjunct)	

Table 6: Encoded representations using viramic model

II.3.2 Alternate model: ZWJ, combining hasanta, inherent vowel character

The alternative encoding model for Syloti Nagri uses independent mechanisms to control conjunct formation and overt rendering of hasanta:

- Conjunct formation is controlled using ZWJ: a conjunct form is requested if ZWJ is present, and does not occur if absent.
- The hasanta is simply a combining diacritic. An overt hasanta appears if the hasanta character is present, and not otherwise. The hasanta character is not used to indicate a dead consonant for processing purposes.

- For specialised processes in which there is a need to distinguish between inherent vowel and the absence of any vowel, a normally-invisible SYLOTI NAGRI INHERENT VOWEL O character is proposed.

These mechanisms are summarised in Table 7. The same notational devices are used as in Table 6. Note the additional notation: IV = SILOTI NAGRI INHERENT VOWEL O.

Situation	Encoded representation	Examples (may be hypothetical)
Co# or CoC (<i>for processing where absence of vowel must be distinguish from inherent vowel</i>)	< ...C, IV, #... >, < ...C, IV, C... >	𑌕𑌃𑌃 “koto” = < 𑌕, IV, 𑌃, IV >
Cd# or CdC	< ...C, #... > , < ...C, C... >	𑌕𑌃𑌃𑌃 “kantok” = < 𑌕, 𑌃, 𑌃, 𑌃, 𑌃, >
Cd + overt hasanta + #, Cd + overt hasanta + C	< ...C, <i>hasanta</i> , #... > , < ...C, <i>hasanta</i> , C... >	𑌕𑌃𑌃𑌃𑌃 “kantok” = < 𑌕, 𑌃, 𑌃, <i>hasanta</i> , 𑌃, 𑌃, <i>hasanta</i> >
Cd + C conjunct	< ...C, ZWJ, C... >	𑌕𑌃𑌃𑌃𑌃 “kantoko” = < 𑌕, 𑌃, 𑌃, ZWJ, 𑌃, 𑌃 >
Vn + C conjunct	< ...Vn, ZWJ, C... >	𑌕𑌃𑌃𑌃 “atiko” = < 𑌕, ZWJ, 𑌃, 𑌃, 𑌃 >
Vn + Vn conjunct	< ...Vn, ZWJ, Vn... >	𑌕𑌃𑌃𑌃 “aeno” = < 𑌕, ZWJ, 𑌃, 𑌃 >
Vn-anusvara + C conjunct	< ...C, <i>anusvara</i> , ZWJ, C... >	𑌕𑌃𑌃𑌃 “angki” = < 𑌕, 𑌃, ZWJ, 𑌃, 𑌃 >
Vs + C conjunct	< ...Vs, ZWJ, C... >	𑌕𑌃𑌃 “kir” = < 𑌕, 𑌃, ZWJ, 𑌃 >
C-Vc + C false conjunct	< ...C, Vc, ZWJ, C... >	𑌕𑌃𑌃 “kere” = < 𑌕, 𑌃, ZWJ, 𑌃, 𑌃 >
C-Vs + C false conjunct	(<i>indistinguishable from Cd + C conjunct followed by Vs</i>)	
Co + C false conjunct	< ...C, IV, ZWJ, C... >	𑌕𑌃 “kot” = < 𑌕, IV, ZWJ, 𑌃 >

Table 7: Encoded representations using alternate mode

The use of ZWJ to determine conjunct formation makes use of previously-established semantics for the ZWJ. The fact that combining marks (e.g., anusvara or e-kar) can occur within a conjoined character sequence relies on a model similar to implementations that are already required for Arabic and other scripts.

That hasanta is merely a combining mark and is not used to indicate dead consonants for processing purposes matches well the facts of usage: it is a recent innovation and is not consistently used.

The fact that overtly-rendered hasanta should never occur in combination with the first element of a conjunct (that is, assuming this model, sequences of the sort $\langle \dots C, \textit{hasanta}, \textit{ZWJ}, C \dots \rangle$ are invalid) is not a built-in feature of this model. Similarly, the inherent vowel character should never co-occur with hasanta or any dependent vowel, and should never follow anusvara, dvisvara or any independent vowel. These restrictions would have to be taken into consideration in software implementations. Such situations could simply be treated as spelling constraints; that is, such sequences would simply be rejected by spelling checkers. Fonts and rendering subsystems could, perhaps, be implemented so as to cause marked renderings for such invalid sequences.

Note also that false consonant-consonant conjuncts pronounced as CVC syllables with a-kar or i-kar cannot be distinguished from conjuncts representing consonant clusters followed by a-kar or i-kar, either in encoded representation or in presentation. (The alternate character sequences that might be considered for this would be the same sequencing as used for Vs + C conjuncts.) This is true of both encoding models, however.

II.3.3 Proposed encoding model

While the viramic model is normally expected for Indic scripts, we find that Syloti Nagri is not typical of Indic scripts, particularly in relation to the behaviours most directly related to the viramic model: usage of hasanta and conjunct formation. The only potential benefit of using the viramic model for Syloti Nagri would be familiarity for those who have had to implement other Indic scripts. This benefit is only marginal, however, since Syloti Nagri behaviours are different. If anything, the viramic model may be misleading for implementers, leading to incomplete or incorrect implementations.

The alternate model uses mechanisms that are already generally familiar to implementers. The mechanisms are also very simple ones, far more simple than those used in the viramic model. As a result, attempts to implement this script would be far more likely to succeed.

Not only is the alternate model simpler overall, it is especially simple in the most frequent usage situations. Assuming that conjunct usage remains infrequent and that most users will not need to distinguish between dead consonants and consonants with inherent vowel, the occurrence of invisible control characters will be very low. When they do occur, control character sequences will be simpler: two-character control sequences (inherent vowel character + ZWJ) would likely occur only in rare instances in the context of paleographic studies.

An additional minor distinction between the two models is that the alternate model is able to accommodate all of the behaviours we have encountered in Syloti Nagri documents. The viramic model is able to handle most but not all behaviours: it cannot distinguish between the absence of vowel versus inherent vowel for the first consonant in a consonant-consonant conjunct. Of course, the need for this ability is quite limited.

In view of these considerations, then, we propose that Syloti Nagri script should be encoded using the simpler alternate encoding model—using ZWJ, a combining hasanta diacritic, and a SYLOTI NAGRI INHERENT VOWEL O character—and not the virama model.

II.4 Other rendering issues

II.4.1 Diacritic positioning

The dependent vowels u-kar and e-kar, and the vowel sign dvisvara generally attach to the base consonant at or near the right-hand edge. For consonants that have a vertical stem on the right-hand edge, these combining marks should attach to the extremes of the stem. Such positioning should be maintained for characters that have strokes to the right of the vertical stem:

Character sequence	Presentation
𑖥 + 𑖅	𑖥𑖅
𑖦 + 𑖅	𑖦𑖅
𑖧 + 𑖅	𑖧𑖅

Table 8: Positioning of dependent vowels u-kar and e-kar

It is also desirable, though not essential, for the left terminal of the hook at the top of the dependent vowel i-kar to attach at the same point:

Character sequence	Presentation
𑖦 + 𑖆	𑖦𑖆

Table 9: Positioning of dependent vowel i-kar

For consonants with no vertical stem, the principle of right-hand attachment should be adapted in appropriate ways. For some consonants, a suitable point of attachment for e-kar and dvisvara is the point at which the body of the consonant meets the mātrā line.

Character sequence	Presentation
𑖛 + 𑖅	𑖛𑖅
𑖛 + 𑖅	𑖛𑖅
𑖛 + 𑖅	𑖛𑖅

Table 10: Positioning of vowel signs on consonants without vertical stem

The u-kar can attach to the descending tail on 𑖛/do/.

$$𑖛 + 𑖅 = 𑖛𑖅$$

Alternately, if a swash form of 𑖛/do/ is used (as is common in manuscripts), the u-kar is typically attached to the descender as follows:

$$\text{𑜀} + \text{𑜁} = \text{𑜂}$$

The positioning of dvisvara in relation to base consonants is comparable to that of e-kar. In contrast, there is considerable variation in manuscripts for the positioning of the anusvara in relation to a base consonant. We consider consistent right-hand alignment for anusvara to be preferable (for consonants with a right-hand vertical stem; for other consonants, appropriate adjustments should be made).

The same principles of positioning for dvisvara and anusvara apply for syllables with u-kar as in syllables with no dependent vowel.

Also, as noted in §§II.1.2 and II.1.3, for syllables with consonant + a-kar + dvisvara or consonant + a-kar + anusvara, alternate positions of dvisvara and anusvara are found in manuscripts: either over the consonant, or over the a-kar. We consider it preferable to position the dvisvara and anusvara directly over the a-kar.¹²

Syllables with consonant + i-kar + dvisvara do not occur in Sylheti. As noted in §II.1.3, for syllables with consonant + i-kar + anusvara, some variation of positioning can be found in manuscripts, but we consider it preferable for modern usage to position the anusvara to the left of the ascending hook of the i-kar.¹³

Syllables with consonant + e-kar + dvisvara do not occur in Sylheti. As noted in §II.1.3, for syllables with consonant + e-kar + anusvara, several alternative positionings of anusvara relative to the e-kar occur, none of which is predominant over the others. For modern usage, we consider it preferable for the e-kar to appear to the left of the anusvara.

Preferred positioning of dvisvara and anusvara is illustrated in Table 11.

Character sequence	Presentation
𑜀 + 𑜁	𑜂
𑜃 + 𑜄 + 𑜁	𑜅
𑜆 + 𑜇	𑜈
𑜀 + 𑜉 + 𑜇	𑜊
𑜋 + 𑜌 + 𑜇	𑜍
𑜎 + 𑜏 + 𑜇	𑜐

Table 11: Positioning of dvisvara and anusvara

¹² As also discussed in §§II.1.2 and II.1.3, the alternate positioning of dvisvara and anusvara in such sequences could, perhaps, be supported using a font feature. Or, if there were a need to distinguish the alternate positions in plain text for palaeographic studies, the distinction could be encoded using alternate spellings. For common-use implementations, however, this issue can be disregarded entirely.

¹³ As discussed in §II.1.3, the alternate positioning of anusvara in relation to i-kar and e-kar could, perhaps, be supported using a font feature, but we do not recommend any mechanism to record such distinctions in plain text. For common-use implementations, this issue can be safely disregarded.

II.4.2 Reordering

Reordering is not an issue for Syloti Nagri script. Unlike many other Indic scripts, it does not have any multi-part or reordering vowel symbols, nor does it have a reordering reph.

II.5 Canonical ordering of combining marks

In this document we propose various combining marks: spacing dependent vowels (a-kar, i-kar, oo-kar), non-spacing dependent vowels (u-kar, e-kar), dvisvara, anusvara and hasanta. It is necessary to consider what canonical combining classes should be assigned to each.

The main issue for canonical combining classes is this: if combining marks can occur consecutively, should alternate orderings be considered semantically distinct or canonically equivalent? In asking this question, we need to consider current usage. It would also be prudent to consider possible adaptations in the future that might not otherwise be expected.

If different orderings should be considered distinct (not canonically equivalent), the combining marks involved must be assigned to the same class, or at least one must be assigned to class 0.

Regardless of what other combining marks they can or cannot co-occur with, it seems reasonable that for any sequences involving the spacing combining marks (a-kar, i-kar, oo-kar), different character ordering should correspond to a different rendering. For these, a combining class of 0 seems reasonable.

Anusvara can co-occur with either e-kar or u-kar. Because of its linguistic meaning, it would only be expected to follow them. It does not interact typographically with u-kar, so different orderings in encoding cannot be used to signify anything. Thus, it is preferable that they be assigned to different non-zero classes.

The anusvara does interact typographically with e-kar, however, and different visual orderings are possible. There is a possibility that this visual distinction could be utilised for some linguistic purpose in the future (e.g. if the Sylheti language underwent substantial phonological change or if the script were adopted for writing some other language with a larger vowel inventory). Thus, it seems prudent to assign anusvara and e-kar to the same non-zero class.

The dvisvara can co-occur with u-kar, but does not co-occur in Sylheti with e-kar. Because of their visual similarity, it seems unlikely that anyone might ever want to use dvisvara together with e-kar. The combination of dvisvara and anusvara is potentially usable, on a purely visual basis, and the two would interact. It seems prudent, then to assign dvisvara to the same class as anusvara.

Thus, it seems to be potentially useful to assign e-kar, anusvara and dvisvara to the same class. These could be assigned to a fixed-position class, but that offers no advantages, and some potential disadvantages, over a class of other combining marks with which they could potentially interact typographically. We propose, therefore, that these be assigned to class 230 (above).

Due to its linguistic meaning, hasanta would not be expected to co-occur with any other of the Syloti Nagri combining marks. If this were always true in practice, then the class to which it was assigned actually would not ever matter. It is difficult to predict whether its linguistic significance might ever undergo change in the future.¹⁴ It seems that there is nothing to be lost if we assign it to the same class as other combining marks with which it would typographically interact, should it

¹⁴ While unexpected, and unsupportable by Unicode, I have heard of a neo-literate community in South Asia using Devanagari script that wanted to use the virama as a vowel modifier to support the large vowel inventory of their language.

ever happen to co-occur with them. Therefore, we propose that this also be assigned to class 230 (above).

Based on reasoning applied to e-kar, we propose that u-kar should be assigned to class 220 (below).

II.6 Collation

While there are a selection of sources describing Syloti Nagri alphabetical order, there is not one established sort order. While the various sources outline an order, we are not aware of any attempt at creating a dictionary and actually putting the suggested orderings to the test. There are some aspects of ordering for which there is complete consensus, however. We will attempt to outline the details, indicating on which aspects there appears to be consensus and on which there is not.

A basic conventional sort order for Syloti Nagri can be found in primers that have been in relatively wide circulation since the mid 20th century. The general ordering they present is typical of Indic scripts.

There is one significant point of variation among authors: some place the ് /rro/ toward the end of the order, after ് /lo/ and before ് /so/; others place it earlier, at the end of the retroflex series, that is, after ് /ddho/. The same issue occurs for Bengali: both orders of the corresponding Bengali character are documented. In the case of Syloti Nagri, therefore, both orders can claim to have established precedent. The traditional Bengali ordering places the counterpart to ് /rro/ toward the end of the order, and we will assume that in the discussion that follows:

The basic alphabetic order places independent vowels before consonants. The basic order, therefore, is as follows:

Basic order	Basic order (cont'd)	Basic order (cont'd)
𑌀 independent /a	𑌁 /jo/	𑌂 /po/
𑌃 independent /i/	𑌄 /jho/	𑌅 /pho/
𑌆 independent /u/	𑌇 /tto/	𑌈 /bo/
𑌉 independent /e/	𑌊 /ttho/	𑌋 /bho/
𑌌 independent /o/	𑌍 /ddo/	𑌎 /mo/
𑌏 /ko/	𑌐 /ddho/	𑌑 /ro/
𑌒 /kho/	𑌓 /to/	𑌔 /lo/
𑌕 /go/	𑌖 /tho/	𑌗 /rro/
𑌘 /gho/	𑌙 /do/	𑌚 /so/
𑌛 /co/	𑌜 /dho/	𑌟 /ho/
𑌠 /cho/	𑌡 /no/	

Table 12: Syloti Nagri basic sort order

After considering ordering of syllable onsets, one must consider the ordering of syllable codas. In pronunciation-based ordering, it would be necessary to distinguish live consonants with inherent

vowels from dead consonants. In normal Syloti Nagri sorting, however, ordering is based purely on spelling. This means that inherent vowels are effectively ignored. It is simple enough to achieve this result simply by placing dependent vowels after consonants in the primary ordering. Thus, character sequences will sort grouped in the order

- < C, # >
- < C, Vn >
- < C, C >
- < C, Vd >

(where # = word boundary, Vn = independent vowel, Vd = dependent vowel). Thus, the first and third of these groups will have live consonants with inherent vowel intermixed with corresponding dead consonants.

The relative ordering of dependent vowels is as follows:

◌া	Syloti Nagri dependent a
◌ী	Syloti Nagri dependent i
◌ু	Syloti Nagri dependent u
◌ে	Syloti Nagri dependent e
◌ৌ	Syloti Nagri dependent oo

Table 13: Syloti Nagri dependent vowel signs

Most primers list the dvisvara immediately after the dependent vowels. This results in an ordering illustrated by the following hypothetical word forms:

রই /roi/, রদা /rok/, রা /ra/, রী /ri/, রে /re/, রৌ /roi/

We see this as problematic for actual practice in Syloti Nagri, however, since every word written with dvisvara has an alternate spelling using the independent vowel ঐ /i/. Thus, for Syloti Nagri, we recommend that dvisvara be given equal weight in primary ordering as the independent vowel ঐ /i/. (The dvisvara and independent vowel ঐ /i/ could be given different weights in secondary ordering.) The effect on the previous list would be as follows:

রই /roi/, রৌ /roi/, রদা /rok/, রা /ra/, রী /ri/, রে /re/

While we believe that modern users would find this more natural, it is admittedly possible that some familiar with the mid-20th century primers might expect a different order. Nevertheless, we consider this ordering preferable.

Another point of variation has to do with the anusvara. One alternative (mentioned in various primers and comparable to the practice of some Bengali dictionaries) places the anusvara after the dependent vowels (and after dvisvara, if it is sorted after dependent vowels). The other alternative places it after other velar consonants (filling the gap in the nasal position for that series of consonants). We prefer the latter ordering, though both have valid motivation. The first alternative results in an ordering illustrated by the following hypothetical word forms:

রদা /rok/, রন /rot/, রা /ra/, রাম /ram/, রাং /rang/, রী /ri/, রং /rong/

In contrast, the second order (which we prefer), results in the following order for the same word forms:

रक /rok/, रंग /rong/, रत /rot/, रा /ra/, रंग /rang/, राम /ram/, री /ri/

The few primers that list hasanta in the alphabet list it at the very end. If this is interpreted to mean that hasanta should be given a primary sorting weight after the dependent vowels (and also dvisvara and anusvara), then anomalies in ordering arise. We observed earlier that the distinction between live consonants with inherent vowel versus dead consonants should be ignored for sorting purposes. Giving hasanta a primary weight contradicts that, however. We conclude, therefore, that hasanta should be ignored in primary ordering. (Hasanta could be given a sorting weight in secondary sorting, if desired. We suggest that it come before consonants in secondary sorting.)

If we agree that inherent vowels and hasanta are irrelevant for primary sorting, then conjuncts do not present any additional problems: they simply sort the same as their non-conjoined expansion.

II.7 Line breaking and hyphenation

In earlier manuscript documents, spaces were not written between words. Because most documents were narrative poetry with regular meters, line-breaking was not an issue: the number of syllables to be put on each line was determined by the meter, and the text was stretched or squeezed in as needed. For certain documents, poetic lines and stanzas were indicated using punctuation rather than physical lines, and in these cases, the potential for line-breaking issues did arise. Some authors would break lines only at word boundaries. For other authors, however, lines could be broken anywhere, the sole exception being between a consonant and its dependent vowels.

In printed materials from the period 1870–1970 and in most 20th century manuscripts, spaces were written between words. This has become standard practice. Thus, for modern usage, software implementations can rely on white space to determine valid line boundaries.

Hyphenation has not been used in Syloti Nagri written tradition. There are potential situations in which hyphenation could be useful, however. We do recommend support for hyphen in Syloti Nagri implementations. Valid hyphenation points are at any “syllable” boundary, meaning before any consonant or independent vowel.

Part III: Other Implementation Issues

III.1 Input methods

Syloti Nagri does not present serious issues in relation to input methods. For instance, the inventory of essential characters are no greater in number than are commonly used for English, thus there is no great concern with fitting the inventory of characters into a limited number of keys. There is potential need for a wider variety of punctuation characters than in languages such as English, so there are minor issues to be considered by someone implementing an input method for Syloti Nagri. But any such issues are relatively minor, and do not impact the feasibility of implementing the script as proposed.

The main potential concerns for input of Syloti Nagri (and Indic scripts in general), have to do with ordering, and the entry of layout control characters.

In general, reading order and visual order coincide in Syloti Nagri. The main potential question in relation to typing order has to do with *dvisvara* and *anusvara*: as noted in §II.1.2 and §II.1.3, and also in §II.4.1, visual ordering of *dvisvara* and *anusvara* has not been consistent in Sylheti writing. There is no established typing order for these characters. As mentioned in earlier sections, however, we recommend consistent positioning of these diacritics in modern usage, and this ordering corresponds with data stored in reading order. The implication for input methods is that typing order, storage order and presentation order would all be consistent.

There is another very minor issue in relation to typing order: as described in §II.2, there are certain false conjuncts for which actual pronunciation order does not match the normal reading order. As discussed in that section, the order of encoded characters in these cases must match the order suggested by the visual reading, and not the order of actual pronunciation. Thus, in these infrequent situations, the order in which users would have to enter data would not be the same as the pronunciation order. Generally, however, typing order, storage order and presentation order would all correspond to pronunciation order.

Assuming the encoding model proposed in §II.3.3, consideration needs to be given to the entry of *hasanta* and ZERO WIDTH JOINER. In specialised usage contexts, input methods would also need to support entry of the SILOTI NAGRI INHERENT VOWEL O.

The ZERO WIDTH JOINER is used to request conjunct formation. Many users are likely to be familiar with certain implementations for Bengali that make use of a *halant* key to produce conjunct characters, so that typing *halant* between two consonants results in the appropriate conjunct appearing. In these same implementations, variant keying sequences involving this same key are used to cause the overt halant to appear. Because conjunct formation and overtly-rendered halant involve distinct characters that do not occur consecutively, it would be possible to implement behaviour for both involving a single key: some particular key, without modifiers, could be used for entry of the overt *hasanta*, and the shifted counterpart sequence could be used for entry of the ZERO WIDTH JOINER, causing a conjunct to appear (assuming entry at an appropriate location in text).

III.2 Typeface variations

III.2.1 General

Only one printed typeface has been in use for the Syloti Nagri script. Although of necessity it is very familiar to present Syloti Nagri readers, it is not ideal for general use because the following pairs of syllables are so close in shape as to be almost indistinguishable:

/co/	व	/bo/	ब
/po/	प	/do/	द
/no/	न	/jo/	ज

Table 14: Inadequately-distinguished glyph pairs in woodblock and metal type

Such problems do not occur in Syloti Nagri manuscripts, where generally there is a clear distinction between characters. For instance, the following sample shows a clear contrast between /ho/ (indicated by the blue oval) and /pu/ (indicated by the red oval).

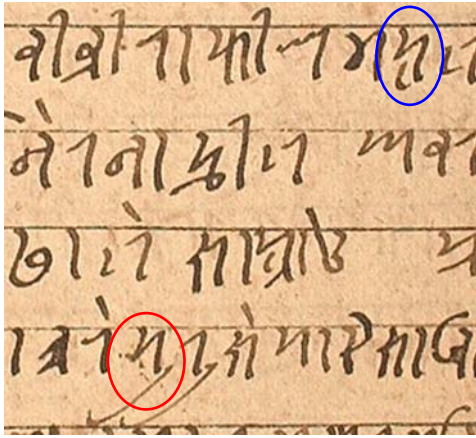


Figure 36: Contrast between /ho/ and /pu/ in manuscripts

Accordingly, the digital font used here has been based on Syloti Nagri manuscripts, modified to approach the woodblock and metal type by consultation with Syloti Nagri-literate members of the Sylot Academy. The principal differences from the manuscript forms are the reduction in the length of swashes on Consonant letters 'Do' 'Cho' and 'Jo', and on vowel letters 'I' and 'U' and the dependent vowel symbol 'U'. The attachment of the dependent vowel symbols also follows the style of the older typeface.

'co'	व	'bo'	ब
'po'	प	'do'	द
'no'	न	'jo'	ज

Table 15: Clearly-distinguished glyph pairs in New Surma font

A further consequence of familiarity with the older typeface is the expectation that Syloti Nagri words should have the horizontal mātrā line at the top, in common with most North Indian scripts except Gujarati. The mātrā line is not an essential feature of Syloti Nagri script, however, as illustrated by the following manuscript sample:

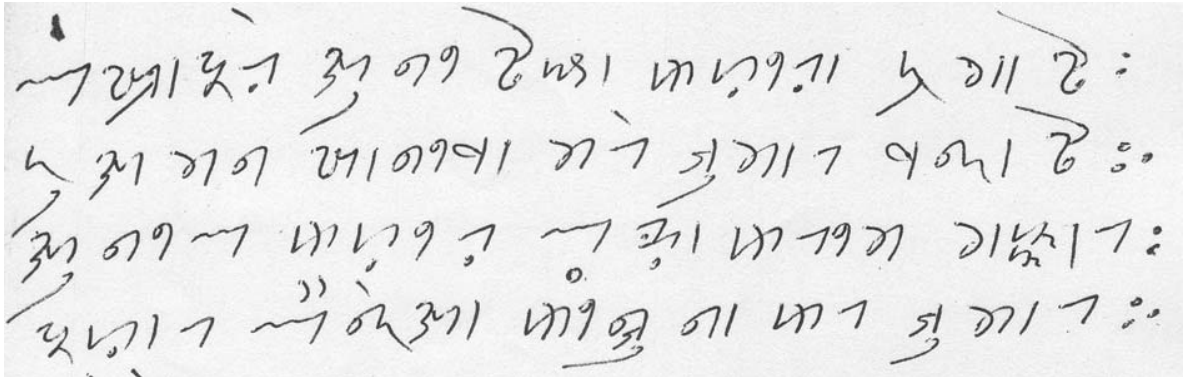


Figure 37: Syloti Nagri writing without mātrā line

In manuscripts, the usual practice has been to use paper to which a ruled grid was added beforehand. Characters were written hanging from the ruled lines. The impression of words having a mātrā line is due to the medium, therefore, and not to the writing of the scribes themselves. This is shown in the following illustration:

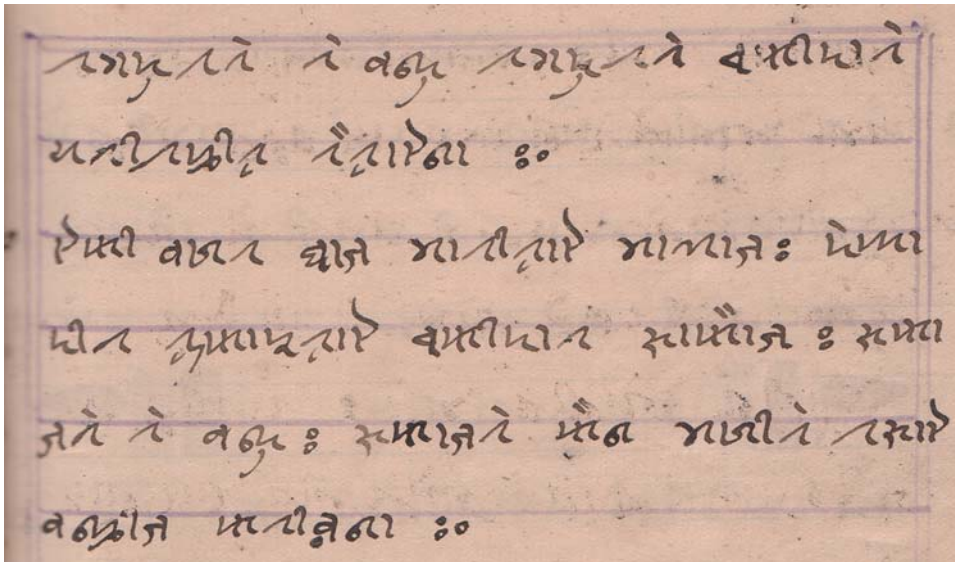


Figure 38: Syloti Nagri written on ruled pages giving the impression of mātrā line

As discussed in §I.2.3, Syloti Nagri is closely related to Gujarati and Kaithi. Since these are not written with the mātrā line, it is not surprising that the same is true in Syloti Nagri manuscripts.

Thus, we think the expectation that the mātrā line should be an integral component of modern Syloti Nagri type should be re-examined. The mātrā line is not a necessary element of Syloti Nagri type forms. It could be argued, in fact, that the mātrā line should be excluded from typefaces intended to closely match historic calligraphic traditions.

Manuscripts have usually been written with a reed pen at an angle of about 45 degrees, but there is considerable variation in slant and serif formation, giving scope for many authentic typefaces.

III.2.2 Specific letterforms

In any script, there can be significant variations in the design of characters from one scribe to another. This is the case, as expected, for Syloti Nagri. We wish to point out a small number of particular cases.

Two forms of the consonant conjunct /ko/ + /ko/ are found in manuscripts:

𑖦+𑖦 → 𑖦𑖦 or 𑖦𑖦

As is the case with conjuncts in general, neither occurs very frequently.

The greatest variation in shape is to be found in the consonant letter 'so', which appears to have been formed in two distinct ways by different Syloti Nagri scribes. Nominally, this shape for this character has three elements:



Figure 39: Stroke order for 'so'

Some scribes write this in two separate parts, the second part (strokes 2 and 3) being occasionally indistinguishable from the character 𑖦 'ro'. Other scribes use slightly different strokes to create a very similar shape: one stroke sequence is used to form the initial curl, upward diagonal and downward vertical, then afterwards adding a small downward stroke to the bottom of the diagonal:



Figure 40: Alternate stroke order for 'so'

Other scribes use a stroke order similar to the first order described above, but create differently structured type forms:

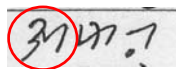


Figure 41: Alternate structure for 'so'

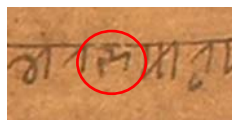


Figure 42: Alternate structure for 'so'

In older manuscripts, 'co' is found with both pointed and round shapes:

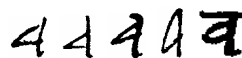


Figure 43: Variant forms for 'co'

We prefer a pointed design to provide greater contrast with ব ‘bo’. (C.f. §III.2.1.)

In older manuscripts, the independent vowel o has a pair of circles, either separate, or attached, or with one separate and one attached:

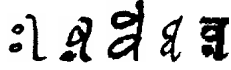


Figure 44: Variant forms for independent vowel o

A form with a dot in the place of the lower circle is also attested, which we prefer to provide greater contrast with /bo/ + /bo/ conjunct.

III.3 Document design

III.3.1 Directionality

Syloti Nagri text is written consistently left-to-right with successive lines proceeding down the page. In Sylheti written tradition, however, books have been ordered with a right-to-left page sequence (start from the right-hand cover with odd pages on the left and even pages on the right), as for Arabic. This is true of manuscripts and printed materials alike. A comparison with Arabic may be intentional, as the vast majority of Syloti Nagri literature is religious teaching with a mainly Islamic content.

For modern usage, it seems reasonable to use right-to-left page sequencing for certain documents, such as reproductions of historic, religious poetry, but to allow left-to-right page sequencing for other purposes. This corresponds to current practice for Bengali, in which most books use left-to-right page order, but particular works, such as reproductions of Bengali Islamic poetry, are published using right-to-left page order.

III.3.2 Page numbers

Page numbers do not appear to have been used originally in manuscripts. (Although all the manuscripts we have seen are now numbered using Bengali numerals, these are usually in a different hand, and generally scrawled in an approximate position on the page, in contrast to the precise layout and careful calligraphy of the manuscript.)

Woodblock and metal-type printed materials from the period 1870–1970 have page numbers in Bengali numerals which clearly use a Bengali typeface. This is especially obvious in some books which also have headers in Bengali script. We have not seen any examples of numerals in the Syloti Nagri type.

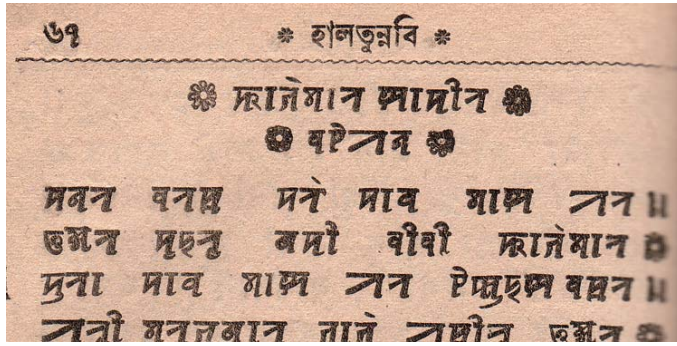


Figure 45: Printed material with Bengali page numbers and header

In the absence of page numbers, the sheet order seems to have been preserved by repeating the first word of a recto (odd-numbered) page in the bottom left hand margin of the previous verso (even-numbered) page.

Page numbering is clearly a recent innovation, but also clearly desirable for modern use. It is recommended that Bengali digits be used.

III.3.3 Page and book layout

There were several common characteristics of historic Sylheti manuscripts:

- Manuscripts typically used pre-ruled pages with a grid of usually 16 to 24 horizontal lines.
- The text is always justified, sometimes with extreme difficulty when a long line of poetry is compressed to fit a line on the page.
- The text usually ‘hangs’ from the ruled lines, so that the horizontal line appears to form part of each character, though in fact this is not so.

Printed books from the period 1870–1970 maintain the justification, but not the grid. The typeface used in these materials added a mātrā line to the characters, however, creating a somewhat similar impression as the ruled pages.

The first page of historic printed and manuscript books is not a title page as such, but has a decorative design in the top half, sometimes coloured, generally incorporating stylised Islamic style minarets and domes together with some simple Arabic calligraphy. The text begins below this in the lower half. This is illustrated in Figure 46.

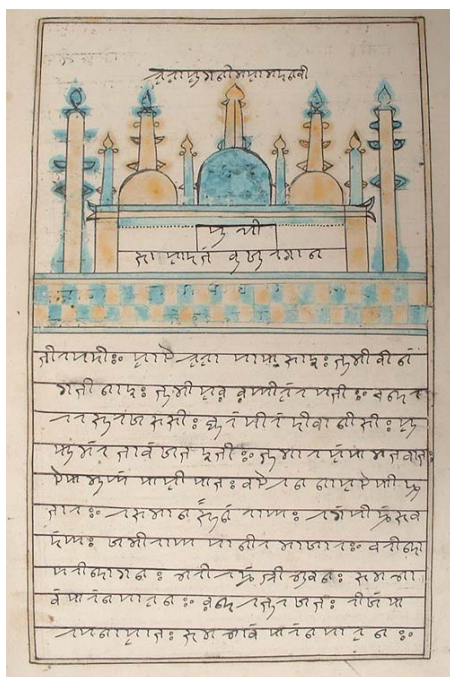


Figure 46: Opening page of a Syloti Nagri book

The author (and sometimes the date of composition) is disclosed as part of the text of the book, either near the beginning or near the end, but not necessarily on the opening page. For example, in one book, the details of the publisher and printer, including personal biography, are provided in an appendix, written in verse using the same meter as the body of the work.

Not all manuscripts have titles and section headings, but where present, they are centred and often have illustrated borders, as shown in Figure 47.

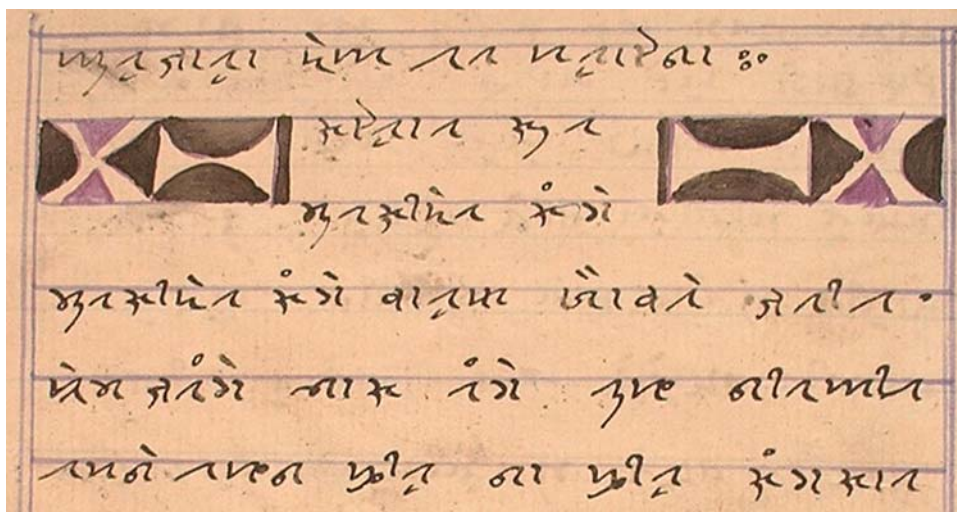


Figure 47: Manuscript with illustrated section heading

Where a section heading consists of multiple lines, these are often of reducing length. In printed books, the lines of a section heading typically are surrounded by full characters or other flourishes.

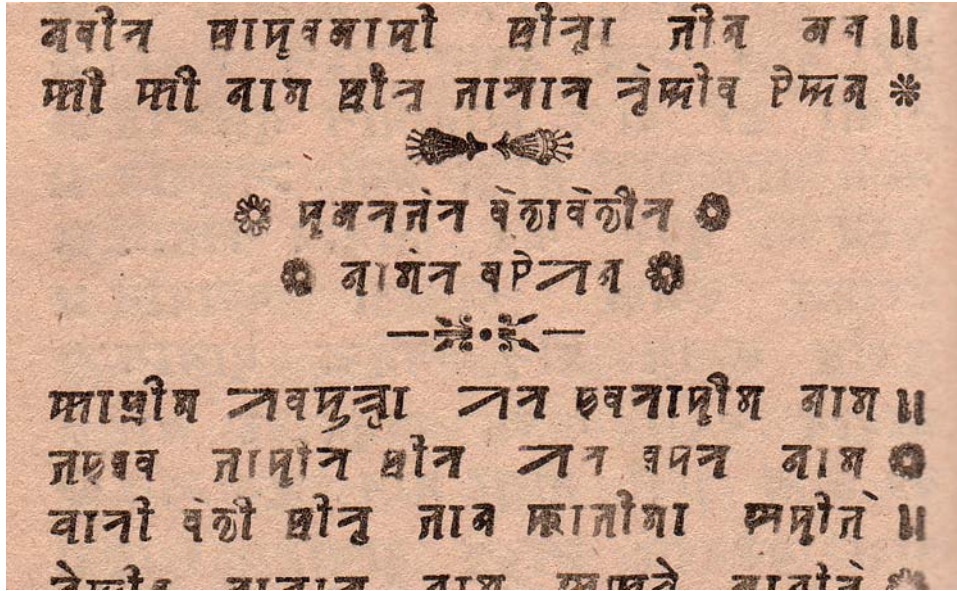


Figure 48: Section heading in printed book with flourishes

Appendices

Appendix 1. Comparison of Syloti Nagri and other North Indic scripts

This appendix provides a comparison of characters in Bengali, Gujarati and Kaithi (Magahi variety) scripts. Gujarati script has been chosen for this purpose as being the most westerly, and hence geographically distant, from Syloti Nagri. It is interesting that it is the only script we have found with a similar ‘L’ to Syloti Nagri’s. Kaithi (Magahi variety) is chosen as it is the nearest geographically, and also the nearest in form that we have observed. Finally, Bengali script is included as being the other, and now the dominant, script in use in the Sylhet region. Some of the most popular works such as *Halot-un-Nobi* had editions printed in both the Bengali and the Syloti Nagri scripts. Bengali is included for these reasons, therefore, not for any similarity with Syloti Nagri; in fact, Bengali is perhaps the least similar to Syloti Nagri of all the North Indian scripts.

A1.1 Vowels

Table A1.1 shows a comparison of vowel signs. The first column shows a generic Roman transliteration. (Note that the actual pronunciations may vary between languages that use one or another of these scripts.) Both dependent and independent vowel signs are shown. The dependent vowels are shown together with the consonant corresponding to /k/.

It will be noted that Syloti Nagri has fewer vowel signs than the other scripts shown. The blank entries for Kaithi reflect that we did not encounter these vowels as distinct signs in the small sampling of Kaithi texts to which we referred.

The following notes are provided as annotations to the table. Roman transliterations are shown between slash characters.

The full form of the Syloti Nagri ᱠ /ā/ seems to be derived from the Kaithi ᱠ /a/ not the Kaithi ᱠᱤ /ā/. This is clear from the older Syloti Nagri manuscript form ᱠ . But the dependent vowel sign ᱠ /ā/ is that of the Kaithi /ā/.

The inherent vowel in Sylheti is /o/, which is the phoneme otherwise explicitly represented in Syloti Nagri script as ᱠ . That is, any consonant not marked with a vowel symbol is assumed to have this vowel, so that the sign ᱠ generally represents /ko/ not /k/.

Where a Bengali word has an /o/ (other than at the end of a word), the corresponding Sylheti word is generally spelled and pronounced with a /u/. The Bengali short /a/ at the end of a word generally becomes Sylheti /o/, but is usually not written. In recent decades, an innovation has occurred in Bengali spelling such that the word-final /a/ (pronounced /o/) of certain words is now usually written ᱠ /ō/ (still pronounced /o/). Recently, some Sylheti authors have expressed interest in exactly the same innovation for Syloti Nagri. (This came about quite independently of the changes that have occurred in Bengali.) Thus, these Sylheti authors wish to spell word-final /a/ (pronounced /o/) using a newly-innovated vowel sign ᱠ /ō/ (still pronounced /o/).

In Sylheti orthography, ᱠ corresponds to both the Bengali vowels ᱠ /a/ and ᱠ /o/, cognates of which in Sylheti are pronounced the same; the Syloti Nagri ᱠ is also used to represent the Sylheti semivowel /w/.

It appears likely that the Syloti Nagri vowel letter **৳** /o/ is derived from the Kaithi semivowel **𑂏** which corresponds to the Gujarati character **વ** and the semivowel use of the Bengali character **ব** /v, w/. This would explain its otherwise anomalous placement as the last Syloti Nagri vowel.

Syloti Nagri does not have letters corresponding to Bengali characters **য** /ý/ or **য়** /y/. In cognate words, one will find Syloti Nagri **পৈ** /e/ used where the Bengali words are spelled with **য়ে**, pronounced /y/, and Syloti Nagri **৳** /o/ used where the Bengali words are spelled with **ওয়ে**, pronounced /w/.

Diphthongs ending in /i/ are common in Sylheti. It is acceptable to spell these either in full or by placing the diphthong (dvisvara) symbol **ঁ** over the syllable. Thus, **দাঃ** = **দৌ** /koi/; **দ্রঃ** = **দ্রৌ** /kui/; **দাঃ** = **দৌ** /kai/; **৳ঃ** = **৳ৌ** /oi/.

Vowel	Bengali		Gujarati		Kaithi		Syloti Nagri	
	Indep.	Dep.	Indep.	Dep.	Indep.	Dep.	Indep.	Dep.
a	অ	ক	અ	ક	𑂏	𑂐		
ā	আ	কা	આ	કા	𑂏𑂐	𑂐𑂐	𑂐	দা
i	ই	কি	ઈ	કિ	𑂏	𑂐	𑂏	দী
ī	ঈ	কী	ઈ	કી				
u	উ	কু	ઉ	કુ	𑂏	𑂐	𑂏	দু
ū	ঊ	কূ	ઊ	કૂ				
r	ঋ	কৃ	ઋ	કૃ				
e	এ	কে	એ	કે	𑂏	𑂐	পৈ	দো
ai	ঐ	কৈ	ઐ	કૈ	𑂏	𑂐	৳ৈ	দৌ
o	ও	কো	ઓ	કો	𑂏	𑂐	৳	দা (দো)
au	ঔ	কৌ	ઔ	કૌ	𑂏	𑂐		

Table A1.1: Comparison of vowel signs

A1.2 Consonants

Table A1.2 shows a comparison of consonants. The first column shows a generic Roman transliteration. (Note that the actual pronunciations may vary from one language to another.)

The following notes are provided as annotation to the table. Roman transliterations are shown between slash characters.

The Bengali alphabet enumerates the character ব twice as distinct graphemes, once as the consonant /b/ and once as the semivowel /v, w/. These were written differently in medieval Bengali, but in modern Bengali have the same form though still retaining two positions in the alphabet.

Syloti Nagri does not have letters corresponding to Bengali characters য় /ý/ or য় /y/. In cognate words, one will find Syloti Nagri ऐ /e/ used where the Bengali words are spelled with য়, pronounced /y/, and Syloti Nagri ঔ /o/ used where the Bengali words are spelled with ঔয়, pronounced /w/. Where Bengali words use য় /ý/, Sylheti words will be spelled using Syloti Nagri ञ /j/.

Syloti Nagri has only one character श corresponding to the Bengali characters শ, ষ and ঞ, which are all pronounced /sh/.

The Syloti Nagri *anusvara* ँ corresponds to the Bengali *candrabindu* ঁ and the Gujarati *anusvara* ઁ, and is similarly written above the syllable. Whereas the Bengali and Gujarati characters are used to indicate nasalization, however, the Syloti Nagri character is used to indicate the Sylheti phoneme /ng/, corresponding to the Bengali *anusvara* ঙ /m/ and the Kaithi • .

Cons.	Bengali	Gujarati	Kaithi	Syloti Nagri
k	ক	ક	क	दा
kh	খ	ખ	ख	दा
g	গ	ગ	ग	ग
gh	ঘ	ઘ	घ	ब
ṅ	ঙ	ંગ		
c	চ	ચ	च	ब
ch	ছ	છ	छ	द्र
j	জ	જ	ज	ज
jh	ঝ	ઝ	झ	दा
ñ	ঞ	ઞ		
t̪	ট	ટ	ट	ठ
th̪	ঠ	ઠ		ठ
d̪	ড	ડ	ड	ड
r̪	ড়	ઢ	ढ	ढ
dh̪	ঢ	ढ	ढ	ढ
rh̪	ঢ়			
ṇ	ণ	ણ		
t	ত, ৎ	ત	त	त
th	থ	थ	थ	त
d	দ	દ	द	द

Table A1.2: Comparison of consonant symbols

Cons.	Bengali	Gujarati	Kaithi	Syloti Nagri
dh	ধ	ધ	ध	दा
n	ন	ન	न	न
p	প	પ	प	दा
ph	ফ	ફ		दा
b	ব	બ	ब	ब
bh	ভ	ભ	भ	न
m	ম	મ	म	म
y	য	ય		
y	য়			
r	র	ર	र	न
l	ল	લ	ल	न
v	ব	વ	व	(र)
ś	শ	શ	श	श
ṣ	ষ	ષ		
s	স	સ		
h	হ	હ	ह	दा
m̐	ং			॰
ḥ	ঃ	:		
~	্	.	.	

References

- Bhuiya, M.A. 2000. *Jalalavadi Nagri: a unique script & literature of Sylheti Bangla*. Badarpur, Assam, India: National Publishers.
- Chalmers, R. 1996. *Learning Sylheti*. London: Centre for Bangladeshi Studies.
- Grierson, G.A. 1899. *A Handbook to the Kaithi Character, 2nd. edition*. Calcutta.
- Grierson, G.A. 1903. *Linguistic Survey of India, Vol. 5, Part I*. Calcutta.
- Grierson, G.A. 1928. *Linguistic Survey of India, Vol. 1 Part II*. Calcutta.
- Qadir, Dr. S.M. Ghulam. 1999. *Sileti Nagri Lipi - Bhasha O Sahitya* (The Sylheti Nagri script - language and literature). PhD thesis, Bangla Academy, Dhaka.
- Rizvi, S.N.H. (ed.) 1975. *Bangladesh District Gazetteers: Sylhet*. Dhaka: Government of Bangladesh.
- Sarkar, J.N. 1948. *History of Bengal, Vol. II*. Dhaka.
- Spratt, David, and Nancy Spratt. 1987. *The Phonology of Sylheti Bangla*. SIL.