Carmen Brandt

Why Dots and Dashes Matter

Writing Bengali in Roman Script

A friend recently sent me a photo of an Indian restaurant called "Vagina Tandoori" and asked for my 'official statement' as a scholar of South Asian studies. After we had had a good laugh – as many other Internet users had before us – I was not shy of providing an answer, as I explained the potential reason behind this awkward name: the restaurant owners could be of Bengali origin and might have transcribed the Bengali term ভাগিনা/*bhāginā*[1] "nephew"[2], the way they thought it should be written in Roman script. I have come across this transcription several times in Bangladesh. Beyond that, a male colleague of mine regularly gets emails from his "vagina", i.e. from a friend in Bangladesh who is much younger to him and refers to himself as his "nephew". Even though the awkward restaurant name turned out to be a hoax,[3] the writing of South Asian languages – especially Bengali – in Roman script is, nonetheless, often a challenge. Hence, this essay reflects on these challenges, discusses in general why South Asian languages are often written in Roman script, and which options might be best to do so.

The Roman Script – a Symbol of Colonial Conquest

As I have already discussed in detail (Brandt 2020), in South Asia, the Roman script is often perceived as a symbol of foreign domination in past and present. Its

[1] All Bengali words are transliterated according to the rules in Table 4 below.

[2] That means the son of one's sister or the son of one's husband's sister. The colloquial form of ভাগিনেয়/*bhāgineẏa* cited here is the one most often used in Bangladesh.

[3] Seemingly, someone either manipulated a digital photo of the restaurant, or, the first letter of the actual restaurant name – "Nagina Tandoori" – was broken, and without the first vertical line the N looked like a V. Even the British newspaper *The Independent* published the erroneous name of the restaurant in a list of the ten worst named restaurants (The Independent 2010), and over the years several photos of the restaurant have been posted on various social media platforms, e.g. Twitter, and other Internet websites. Unsurprisingly, the restaurant owners changed the name. It is now named "Diwali".

Veröffentlicht in: U. Niklas, H. W. Wessler, P. Wyzlic, S. Zimmer (Hg.): *»Das alles hier«: Festschrift für Konrad Klaus zum 65. Geburtstag*. Heidelberg, Berlin: CrossAsia-eBooks, 2021. DOI: <https://doi.org/10.11588/xabooks.815>

use for South Asian languages by Christian missionaries, colonial administrators and European linguists, in the wake of European colonialism, even today causes negative feelings among some people in South Asia, even though a host of people write their first language, first and foremost, in the Roman script, e.g. Bodo, Garo, Konkani and Santali, and, in the case of Mizo, even officially. There are several reasons for the widespread use of the Roman script for South Asian languages.[4]

The most important reason goes back to the nexus between the conquest of South Asian regions by European powers, and the spread of printing technology during that time. With the invention of the modern movable-type printing system, and the rise of printed publications – the so-called "printing revolution" (Eisenstein 2012) – the Roman script was often also applied to languages written predominantly in other scripts, or none, i.e. having no written form yet. While the production of fonts for other scripts was at times difficult and costly – thus, in some cases, the main reason for applying the Roman script – in other cases European languages and the Roman script spread among the colonised population as a means of upward social mobility. Moreover, the Roman script was at times also an instrument of control; an example is the case of the Portuguese language in Goa, and the writing of Konkani in Roman script.

In contrast to Konkani (which, before the Portuguese conquest, was written, among others, in Goykanadi script; and today officially in Modern Nagari, but by many Christian Konkani speakers still in Roman script), other languages – so-called 'tribal' languages – were only written down for the first time by European or American Christian missionaries in the 19th and 20th century. Those missionaries overwhelmingly preferred the Roman script – the writing system they were most familiar with and for which printing technology was cheapest at that time – instead of applying an indigenous South Asian script, or inventing a new one. For this very reason, for instance, Mizo is written exclusively in Roman script until today, while in many other cases indigenous South Asian scripts – 'borrowed' from languages of the majority population (e.g. Bengali, Hindi, Oriya, Tamil, etc.) or more recently invented ones (cf. Brandt 2014: 88–91) – are applied additionally (e.g. Garo, Ho, Santali, etc.), making these languages bi- or multiscriptal.

Missionaries contributed to the standardisation of so-called 'tribal' languages and their writing into Roman script, from the 19th century onwards. But even before then, Orientalists, who were contributing to establish the power of the British Empire by studying the colonised populations – their cultures, religions and languages – preferred, at times, the Roman script, even for languages which had written literary traditions in their own scripts – as in the reproduction, also in non-

[4] Since I have discussed these reasons in detail before (cf. Brandt 2020), I will provide only a summary here.

colonial contexts, of non-Roman scripts generally, except for Greek. William Jones (1746–1794) points out one reason for this circumstance in the beginning of his publication "A Dissertation on the Orthography of Asiatick Words in Roman Letters" (Jones 1787: 1):

> Every man, who has occasion to compose tracts on *Asiatick* literature, or to translate from the *Asiatick* languages, must always find it convenient, and sometimes necessary, to express *Arabian*, *Indian* and *Persian* words, or sentences, in the characters generally used among *Europeans* […].

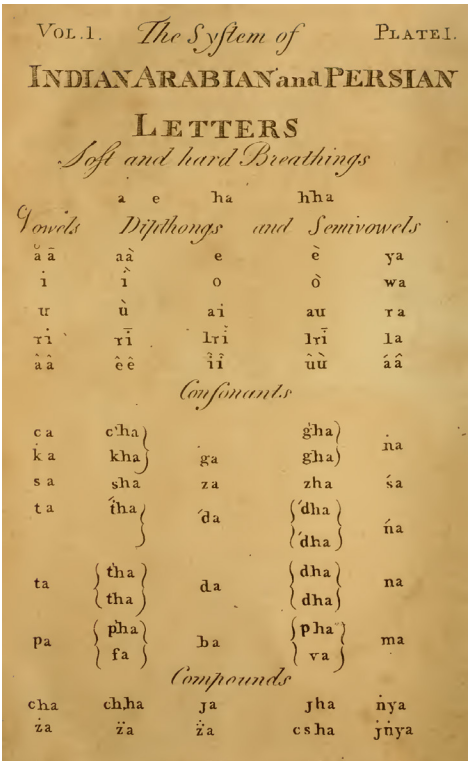
Besides this convenience for European Orientalists, another important factor was also the comparison of languages by Jones and his colleagues, which laid the foundation stone for the establishment of the modern disciplines of historical and comparative linguistics, and, more specifically, Indo-European studies. The latter became popular after Jones postulated, during a lecture at the *Asiatick Society of Bengal* in Calcutta in 1786, that most European languages are related to Persian and Sanskrit and hence form one language family (Jones 1799), even though he was not the first to make this observation.[5] However, the discovery of the relationship of Indo-European languages might have been accelerated by writing all these languages in only one script, i.e. the Roman script. Be that as it may, William Jones suggested, in the essay mentioned above, a standardised transliteration system for several Asian languages, or rather, for a host of scripts usually applied to these languages, e.g. the Perso-Arabic and Devanagari scripts and the script known from the Bengal region, today called "Bengali script"[6] (see Figure 1).

Although linguists even today prefer (modified versions of) the Roman script for comparing languages with each other, the fact that the Roman script has a history engraved in colonialism is a heavy burden. Thus, the Roman script is often perceived as having been applied to South Asian languages solely for reasons of control over the colonised population and in this context highly criticised by some scholars (cf., e.g., Majeed 2019: 110f.). However, people who resent use of the Roman script often tend to ignore the fact that the standardisation of South Asian

[5] For an overview on scholars observing the connections among Indo-European languages before William Jones, and erroneous conclusions and observations made by Jones, see, for instance, Campbell 2006.

[6] Today, this script is designated, according to the International Organization of Standardization, as "Bengali (Bangla)". However, the exclusive naming of this script as "Bengali script", even though it has been used for centuries for other languages too, such as Assamese and Bishnupriya, causes resentment – for instance, among Assamese script activists (cf. Brandt & Sohoni 2018: 7). Therefore, I shall write Bengali in this context with inverted commas in the present article, i.e. 'Bengali' script, to signal to the reader that this appellation is problematic. A more appropriate name for this script is "Eastern Nagari".

languages *and* scripts themselves goes overwhelmingly back to European or American Orientalists, Christian missionaries, and administrators employed by the British colonial rulers (cf. Ross 2018).


VOL. I. *The System of* PLATE I.

INDIAN ARABIAN and PERSIAN

LETTERS

Soft and hard Breathings

a e ha hha

Vowels Diphthongs and Semivowels

a ā	aà	e	è	ya
i	î	o	ò	wa
u	ù	ai	au	ra
ṛi	ṛī	ḷṛi	ḷṛī	la
â â	ê ê	î î	û ù	á â

Consonants

c a	c'ha		g'ha	
k a	kha	ga	gha	ṅa
s a	sha	za	zha	śa
t a	ṫha	ḋa	ḋha / ḋha	ṅa
ta	ṫha / tha	da	dha / dha	na
pa	ṗha / fa	ba	pha / va	ma

Compounds

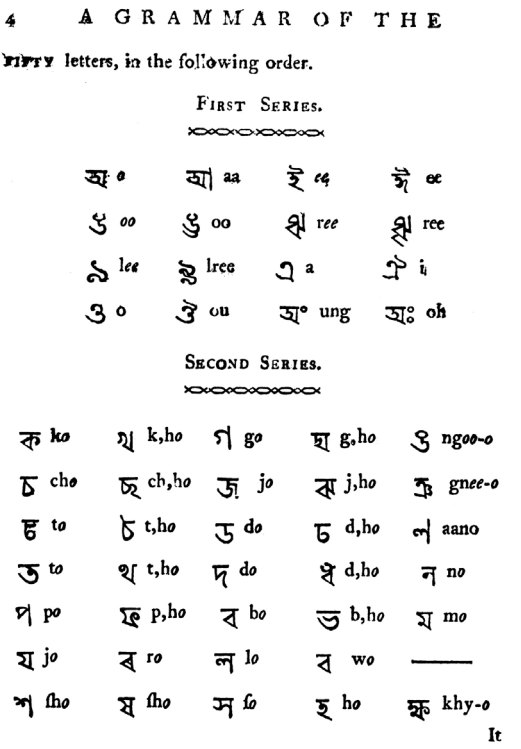
cha	ch.ha	ja	jha	ṅya
ża	z̈a	z̈a	csha	j.ṅya

Figure 1: Plate 1 of William Jones's essay displaying various options for transliterating different scripts (Jones 1787: [0]).

This refers, for instance, to Bengali and Nathaniel Brassey Halhed (1751–1830), an English Orientalist working for several years for the East India Company in Bengal.[7] The first Bengali grammar was published by the Portuguese missionary Manuel da Assumpção who used the Roman script for Bengali in his book (1743), whereas Halhed's Bengali grammar (Halhed 1778) was the first book displaying printed Bengali in its script (see Figure 2). The wooden typeface for this book was produced by Charles Wilkins (1749–1836), with the help of the Bengali blacksmith Pañcānan Karmakār (Ross 2018: 449). This typeface and the grammar have

[7] For an excellent study on the development of printing in Bengali and the evolution of the 'Bengali' character see Ross 1999.

doubtlessly contributed to the standardisation of the Bengali language and its script, even though they were not meant for the local population (ibid.: 447). Nonetheless, they were only the vanguards for the following linguistic and philological studies of the Bengali language and the necessary development for a suitable typeface – by Europeans and Bengalis.



4 A GRAMMAR OF THE

fifty letters, in the following order.

FIRST SERIES.

অ o	আ aa	ই ee	ঈ ee
উ oo	ঊ oo	ঋ ree	ৠ ree
ঌ lee	ৡ lree	এ a	ঐ i
ও o	ঔ ou	অং ung	অঃ oh

SECOND SERIES.

ক ko	খ k,ho	গ go	ঘ g,ho	ঙ ngoo-o
চ cho	ছ ch,ho	জ jo	ঝ j,ho	ঞ gnee-o
ট to	ঠ t,ho	ড do	ঢ d,ho	ণ aano
ত to	থ t,ho	দ do	ধ d,ho	ন no
প po	ফ p,ho	ব bo	ভ b,ho	ম mo
য jo	র ro	ল lo	ব wo	———
শ sho	ষ sho	স so	হ ho	ক্ষ khy-o

It

Figure 2: Bengali vowels and consonants in Halhed's grammar and their transcription (Halhed 1778: 4).

Ultimately, the standardisation processes of the Bengali language and its script, impelled by this new technology, were, in the long run, also crucial for the emergence of Bengali ethnic consciousness during the so-called Bengal Renaissance (Ahmed 1965: 79–85). Retrospectively, the "printing revolution" in Bengali, i.e. in the vernacular language of the colonised population, became possible only after the establishment of Calcutta as the centre/capital of British dominated India (1773–1911). This new metropolitan space enabled the ensuing cultural and political interaction between Europeans and Bengali intellectuals, and the emergence

of the so-called *bhadralok* ("gentlemen"). Only later on, the British had to realise that the printed Bengali word became a weapon of the growing resistance against colonial rule, whereas Halhed had still justified his grammar as serving the British Empire (Halhed 1778: i f.):

> […,] one of its most important desiderata is the cultivation of a right understanding and of a general medium of intercourse between the Government and its Subjects; between the Natives of Europe who are to rule, and the Inhabitants of India who are to obey.

The nexus between British colonial rule and the spread of printing technology for vernacular languages contributed not only to the emergence of Bengali and Indian nationalism in general (Bayly 1996: 241f.), but also to the importance of modern South Asian languages and scripts for group identity politics along ethnolinguistic and religious lines (e.g. in the case of Hindi/Urdu) in contemporary South Asia (cf. Brandt & Sohoni 2018) and the resentment against the Roman script for South Asian languages (cf. Brandt 2020).

Keeping these complex, and at times clearly contradictory, historical developments in mind, hence it is not astonishing, too, that today many people in India, Pakistan and Bangladesh are, on one hand, very proud of their first language, but on the other, still often prefer writing it in Roman script – in a similar way many condemn British colonialism, but at the same time learn English for higher education and upward social mobility. The symbolic value of one's own 'authentic' language and script obviously stands in sharp contrast to the communicative and socioeconomic value of English and the Roman script. The latter is doubtlessly the *scripta franca*[8] of modern South Asia.

The Roman Script – the *Scripta Franca* of South Asia

The status of the Roman alphabet as the *scripta franca* of South Asia is a phenomenon still little studied, even though the important role of this script in South Asia is obvious (cf. Brandt 2020). The reasons for this can be found in the developments described above, but also in the importance of English today on a global level, and of new technologies, such as the Internet and mobile phones, and their respective applications which, in many cases, often demand basic command over the Roman script. Similar to the invention of the modern movable-type printing press (suitable for mass print production) by Johannes Gutenberg (1400–1468), with its spread in

[8] Even though the term *scripta franca* is a neologism, and neither grammatically correct in Latin nor in English (or any other language), it reflects ideally its function and importance comparable to a *lingua franca*.

Europe from the middle of the 15th century, new communication technologies are also today, at their inception, often only available for the language/script of their region of origin. And even though new technologies can nowadays spread much faster to all corners of the world, due to modern globalisation and global capitalism, further developments of new technologies, and their modification for other languages/scripts, depend highly on the demands of the various markets. Moreover, one has to consider the fact that the development of computer technology for the masses, and its respective applications – still predominantly in the hands of US-American companies – assures English and the Roman script their important roles in this field.

Since English is the *lingua franca* among the socioeconomically higher strata in South Asia, and additionally, every formally educated person in South Asia can at least read the Roman script because English is a mandatory subject in most South Asian schools, the sale of new communication products hardly suffers any setback in this region – at least none that can be blamed on a missing 'cultural' modification but rather on the low purchasing power of the masses. In general, compared to regions in which English and the Roman script are less popular (and the purchasing power of the masses is on average much higher), e.g. China, Japan, Korea and Russia, the development of new technologies for modern South Asian languages and their respective scripts, such as text messaging with mobile phones, took much longer. Interestingly enough, even though all major South Asian scripts are by now included in the Unicode Standard (Unicode 2020) and their usage for mobile phones and computer technology is consequently no problem anymore, many people still write their language digitally in Roman script.

However, there are several reasons why the Roman script is still consciously used for South Asian languages: the overwhelming variety of scripts in South Asia, especially in India; the multiscriptality of some languages, e.g. Hindi/Urdu; and the fact that the Roman alphabet is the only script which can be read across state and ethnolinguistic borders. But while the Roman script is used, for instance, for Hindi film screenplays, Bollywood posters and other kinds of advertisement in order to reach more people, in other cases, it might be applied more out of convenience.

Bengali in Roman Script

Among others, this seems to be the case for Bengali, a language which otherwise can be deemed to be monoscriptal. It is far less widespread as a second language than Hindi/Urdu, and digital communication technologies are readily at hand for Bengali and its script. Especially in Bangladesh, the Bengali language has today

an incomparably high status due to its role in the self-assertion of Bengalis when this region (today Bangladesh) was part of Pakistan (before 1971). In contrast to India and Pakistan, where English is an official language (besides other languages), Bengali is the sole national and official language of Bangladesh, and English and the Roman script are less visible than in India, such as in Calcutta, the cradle of Bengali nationalism. Nonetheless, on the Internet and with their mobile phones, even Bengalis living in Bangladesh often write their language in Roman script.

While many Bengalis might have become accustomed to writing their language in Roman script while using digital technologies during the time when the script for their language was not yet available, other reasons might play a role as well. Unfortunately, without an in-depth study on this phenomenon, we can for the time-being only speculate about those: one reason might be related to the fact that code-switching is widely spread in South Asia and therefore writing, for instance Bengali and English, in one common script – in this case the Roman script – ensures an easier writing flow. The other way round – writing English words and whole phrases in 'Bengali' script – is also well known, for instance on signboards (see Figure 3) and from Bengali newspapers, novels, etc., while today only a few publications use both scripts. Another reason might be owed to the quite complicated orthography of Bengali (which will become obvious from the discussion below): an *ad hoc* Romanisation can disguise the deficits one has in this regard; I observed, especially among Bengalis who have attended exclusively English-medium schools, that many of them face problems writing Bengali in its respective script without making any mistakes, even if they have a fluent spoken command over their first language. Furthermore, for writing Bengali, it is still often necessary to install a specific software and/or font and/or to change between English, i.e. Roman script, and Bengali and its script. Thus, nowadays Bengali in Roman script in mobile text messages, or on the Internet, might be at times the result of sheer convenience, and not technological limitations.

Hence, it is also not surprising that seemingly no effort is put into a standardised transcription system for Bengali in Roman script, for instance by government institutions in India or Bangladesh. Of course, with the International Phonetic Alphabet (IPA), there is already a transcription system which is based on the Roman script, but this is far from being suitable for daily use, since it consists of several graphemes only used for this alphabet. Another option would obviously be systematic transliteration.

Figure 3: English in 'Bengali' script in Dhaka, Bangladesh (Photo by Carmen Brandt).[9]

Transcription and Transliteration

As already mentioned above, William Jones proposed a systematic writing of Asian languages in Roman script as early as in 1787. Even though he did not mention the term "transliteration", this is exactly what he suggested: a systematic way of transforming a text from one script into another, i.e. a method to reproduce the orthography of the original as accurately as possible, in Jones's case, into Roman script. Among others, he laments in his essay that while studying as a young student an Arabic couplet transcribed into Roman script, seemingly by the French Orientalist Barthélemy d'Herbelot de Molainville (1625–1695), he asked several Arabic experts to write this couplet in Arabic script, but "they all wrote it differently, and all, in my present opinion, erroneously" (Jones 1787: 3).

[9] The transliteration of the writing in 'Bengali' script on this banner is *omen·s plejār paẏent̤*.

The fact that especially the Arabic vowels in this couplet were either reproduced equivocally in Roman script, or not at all, made it difficult for Jones, and the Arabic experts consulted by him, to ascertain the original orthography in Arabic script, which made it, hence, difficult to translate the couplet correctly. One main problem, of which Jones was aware, is that the letters of the Roman alphabet are pronounced differently in the various European languages (which are commonly written in Roman script), at times even differently in the same languages, and additionally, distinctly in their regional variants.[10] This leads to the problem that an *ad hoc* transcription – i.e. an unsystematic reproduction of a language in Roman script usually based on its pronunciation – can vary enormously, depending on the first language of the transcriber and even the recording time, as the pronunciation of words in every language can change over time. Matters become even more complicated when the pronunciation of words changes, but their orthography does not. This circumstance – a language with a, hence, imprecise orthography with regard to pronunciation – applies also to English. In this context, William Jones writes the following (Jones 1787: 13):

> Our *English* alphabet and orthography are disgracefully, and almost ridiculously, imperfect; and it would be impossible to express either *Indian*, *Persian* or *Arabian* words in *Roman* characters, as we are absurdly taught to pronounce them: but a mixture of new characters would be inconvenient; and, by the help of the diacritical marks used by the *French*, with a few of those adopted in our own treatises on *fluxions*, we may apply our present alphabet so happily to the notations of all *Asiatick* languages, as to equal the *Dévanágarì* itself in precision and clearness; and so regularly, that any one, who knew the original letters, might rapidly and unerringly transpose into them all the proper names, appellatives, or cited passages, occurring in tracts of *Asiatick* literature.

Although Jones proposed a precise reproduction of several Asian languages in Roman script, his criticism of the English orthography indicates that he did not suggest this because he believed in the superiority of the English language or the Roman script. On the contrary, Jones was one of those European Orientalists who was fascinated by Sanskrit and the "precision and clearness" of the "Dévanágarì" (i.e. *Devanāgarī* or *Devˈnāgˈrī*). After all, especially his English translation of Kālidāsa's Sanskrit drama *Śakuntalā* provoked a growing romantic interest in India and its rich literatures among intellectuals in Europe, especially among German literati and scholars who had no geopolitical interest in this region (Brandt & Hackenbroch 2017: 40f.). The latter might have been the case for many British

[10] One of the best-known examples in German is the term *Chemie* which is predominantly pronounced as [çeˈmiː] in Northern and Central Germany and as [keˈmiː] in the south of Germany.

Orientalists, too, but however that may be, the nexus between their research while serving the East India Company and/or the British Empire, and hence their subsequent contribution to the consolidation of British colonial rule in South Asia, is undeniable in retrospect. Halhed's quote above, from 1778, illustrates that also during that time British Orientalists were very much aware of this nexus. Thus, it is understandable that the Roman script is today often perceived as a symbol of colonial conquest.

Following on from Jones's proposal to transliterate Asian languages systematically into Roman script with diacritical marks, this is now the standard practise in most philological disciplines dedicated to the study of non-European languages[11] and their literature. Jones's system, and those used by his contemporaries, were developed further, and even though scholars could not agree on a single one until today,[12] the systems are very similar, predominantly intelligible among each other, and, most importantly, reproduce languages written in diverse scripts precisely in Roman script.

However, while such transliteration is common in philological contexts, in many other disciplines scholars still reproduce words, and even whole sentences – especially in South Asian languages such as Bengali – in Roman script without transliterating, or even transcribing, them systematically;[13] similar to the reproduction of those languages by many of their native speakers in mobile phone text messages or on the Internet, or to many colonial English sources.[14] This refers predominantly to social studies, for instance social anthropology, sociology and human geography,[15] and to their respective academic publications, whose empirical data are often based on oral and not written sources, such as qualitative interviews. There are several reasons why, for instance, non-Bengali, mostly European or North American, scholars from these disciplines prefer transcription over

[11] This does not refer to all languages. Languages, for example, written with logographic scripts, such as Chinese and Japanese, are transcribed.

[12] There are, for instance, several transliteration systems to write Sanskrit and other languages written in scripts of South Asian origin, such as the International Alphabet of Sanskrit Transliteration (IAST), the Hunterian System (officially adopted by the Indian state) or the system standardised by the International Organization for Standardization (ISO 15919). These three systems use diacritics, while the Harvard-Kyoto Convention System does not.

[13] A rather rare exception is the Dutch anthropologist Lotte Hoek who uses a self-defined transcription system without any diacritics for the Bengali terms, in her book on cut-pieces in Bangladeshi cinema films (Hoek 2014: ix), even though this is not without problems. The same refers to Sudeep Chakravarti and his book *The Bengalis* in which he also uses a self-defined transcription system, but with diacritics (Chakravarti 2017: xi ff.).

[14] For instance, in the various administrative and ethnographic reports in English produced during British colonial time, the Bengali term *bādiẏā* was transcribed as 'Budeea', 'Bhudiya', 'Badya', 'Badia' or 'Badiya' (Brandt 2018: 201f.).

[15] The list could be much longer; the phenomenon is also known from film studies, musicology, political science, etc.

transliteration: they hardly include written sources in Bengali; they might have 'only' command over spoken Bengali and/or conduct their interviews with the help of an interpreter. In those cases, Bengali terms are often spread throughout the publication, otherwise written in another language, for instance German or English. At times, this pertains to Bengali terms, which are difficult to translate into another language and carry a complex meaning which cannot be translated by a single word. But in many other cases, when terms which can be easily translated are also included in Bengali, one gets the impression that the dispersion of these unsystematically transcribed Bengali terms also serves to convince the reader that the author is highly familiar with the Bengal region and its people, or, so to say, "speaks their language", even if this is not the case.

However, I could observe that in English publications by Bengali scholars, who know their first language very well – spoken and written – and who include Bengali literature in their research work, Bengali words and sentences are also often transcribed in an unsystematic way. The same refers to many other Indian and Bangladeshi publications (newspapers, magazines, reports by state or non-state institutions, Internet websites, etc.) written in English. Seemingly, in such cases, the authors take it for granted that the readers know Bengali well enough to understand what they mean, often disregarding the fact that the orthography of their transcriptions has sometimes hardly anything to do with the original in 'Bengali' script. Fortunately, some Bengali scholars, especially from philological disciplines, are aware of the problems caused by transcribing their language and thus follow one or the other transliteration system for Bengali.[16] But before discussing the advantages and disadvantages of a transliteration or transcription of South Asian languages – especially Bengali – in the conclusion, I shall illustrate the challenges pertaining to Bengali in Roman script.

The Challenging Bengali Orthography

One of the main reasons for the challenging transcription of Bengali is linked to its orthography in its script, the so-called 'Bengali' script,[17] because orthography and pronunciation do not necessarily correspond to each other systematically. Therefore, Bengali in Roman script holds more potential for misunderstanding than Hindi in Roman script. This can be demonstrated by some examples of

[16] Most authors employ one or the other system containing diacritics, such as the ones mentioned in note 12, even though some scholars transliterate according to a self-defined transliteration system (cf., e.g., Ahmed 2001: xv).

[17] See note 6 for an explanation of why the term "Bengali" in the appellation "Bengali script" is problematic.

lexemes shared by both languages; for instance, of Perso-Arabic and Sanskrit origin, which are often pronounced differently, especially in the latter case. Once transliterated into Roman script, these shared words of Sanskrit origin are clearly seen to be overwhelmingly identically spelled as in the source language, Sanskrit. However, since the phonology of both languages is divergent, the pronunciation of shared words is, more often than not, different, even though their spelling might be similar or the same. This is also well known from various European languages which often share the same words, even spelled identically in Roman script, but which are pronounced very differently. With regard to words of Sanskrit origin, Hindi seeks basically to adhere to the Old Indo-Aryan pronunciation of Sanskrit – though the two pronunciations are not identical, particularly, though not only, in the treatment of the short vowel transliterated as *a*. Bengali, however, while conservative in the orthography of such words, basically adheres to a Middle Indo-Aryan pronunciation of them, leading often to pronunciations far removed from what the orthography would suggest. I will return to this below.

One striking difference between Bengali and Hindi lies in the inherent short vowel for consonant graphemes that nearly all autochthonous South Asian scripts possess. While in Hindi this short vowel represents the phoneme [ə], in Bengali it is either an [ɔ] or an [o],[18] depending on various pronunciation rules. According to the rules of most transliterating systems, this short vowel is reproduced as an *a* in the case of both Hindi and Bengali, following the usual transliteration rules of South Asian languages developed on the basis of Sanskrit.[19] Interestingly, when Bengali is transcribed often this inherent vowel is also reproduced as an 'a', even though this stands in stark contrast to its actual pronunciation. Well-known examples are anglicised names of famous Bengali people and places, listed in Table 1.

From some of the examples in Table 1 another difference between Hindi and Bengali becomes obvious, which can complicate matters in a similar way: spoken Modern Standard Bengali[20] does not generally differentiate between the three sibilants of its script – শ (*śa*), ষ (*ṣa*) and স (*sa*) – whereas Hindi, whose script also has three sibilants – श (*śa*), ष (*ṣa*) and स (*sa*) – in pronunciation distinguishes consistently between [ʃə] (for *śa* and *ṣa*) and [sə] (for *sa*). In Bengali, on the other hand, the reproduction of all three sibilants in pronunciation is generally [ʃɔ], though in

[18] For an overview on the International Phonetic Alphabet (IPA) for the Bengali script see Table 4 below.

[19] This refers also to the transliteration systems listed in note 12.

[20] All examples in this article refer to Modern Standard Bengali – the spoken and written Bengali – standardised by the end of the 19th century and in the beginning of the 20th century, based on the local variants in the region forming today the districts Hooghly, Howrah, Nadia, and the North and South 24 Paraganas in the Indian state of West Bengal and its capital, Calcutta.

certain consonantal combinations and positions their pronunciation is [sɔ]. Thus, the sibilant *ś* of the honorific *śrī* is pronounced [ʃ] in Hindi, but [s] in Bengali. However, *ṣ* is pronounced [s] only in consonantal combinations in a few certain non-Sanskritic loanwords, e.g. *māṣṭār* [mas̪tar] "master, teacher" (today written mostly *māsṭār*), whereas *sa* is pronounced [sɔ] in many non-Sanskritic loanwords. These circumstances make it difficult even for many Bengalis to write their language without mistakes in the 'Bengali' script. As regards transcription, the main problem, especially for people who have no spoken or written command over Bengali, is that স (*sa*) very often (see example 1, 2, 8, 9 and 10 in Table 1) – just like in the case of Hindi – and at times even শ (*śa*) (see example 7 in Table 1) and ষ (*ṣa*) are transcribed as an 's', and non-Bengalis may, in consequence, mispronounce words affected by this phenomenon.

Table 1: The inherent vowel in Bengali

	Original	**Transliteration**	**IPA**	**Common anglicised form**
1.	অমর্ত্য সেন	*Amartya Sen*	[ɔmɔrt̪o ʃen̪]	Amartya Sen
2.	আশিস নন্দী	*Āśis Nandī*	[aʃiʃ n̪on̪di]	Ashis Nandy
3.	কলকাতা	*Kal'kātā*	[kolkat̪a]	Calcutta
4.	পার্থ চট্টোপাধ্যায়	*Pārtha Caṭṭopādhyāẏ*	[par̪t̪ʰo cɔʈʈopad̪d̪ʰaj]	Partha Chatterjee
5.	বর্ধমান	*Bardhamān*	[bɔrd̪ʰoman̪]	Bardhaman/ Burdwan
6.	রবীন্দ্রনাথ ঠাকুর	*Rabīndranāth Ṭhākur*	[ɾobin̪d̪ɾon̪at̪ʰ ʈʰakuɾ]	Rabindranath Tagore
7.	শরৎচন্দ্র চট্টোপাধ্যায়	*Śarat·candra Caṭṭopādhyāẏ*	[ʃɔɾot̪cɔn̪d̪ɾo cɔʈʈopad̪d̪ʰaj]	Saratchandra Chattopadhyay
8.	সত্যজিৎ রায়	*Satyajit Rāẏ*	[ʃɔt̪t̪oɟit̪ ɾaj]	Satyajit Ray
9.	সুন্দরবন	*Sundar'ban*	[ʃun̪d̪ɔrbon̪]/[ʃun̪d̪ɔrbɔn̪]	Sundarbans
10.	সুভাষচন্দ্র বসু	*Subhāṣcandra Basu*	[ʃubʰaʃcɔn̪d̪ɾo boʃu]	Subhash Chandra Bose

A similar phenomenon which can affect the pronunciation of Bengali in Roman script, but not that extensively, refers to graphemes representing vowels. There are, for instance, respectively a short and a long 'u' and 'i' in written Bengali: ই/*i* [i], ঈ/*ī* [i], উ/*u* [u] and ঊ/*ū* [u]. The IPA in brackets behind these characters illustrates that also here the orthography distinguishes between them, but they

are pronounced in the same way.[21] On the other hand, the vowel graphemes এ/*e* ([e] or [æ]) and অ/*a* ([ɔ] or [o]), and the inherent vowel ([ɔ] or [o]) can carry two different phonetic values, depending on various pronunciation rules, for instance এটা/*eṭā* [eṭa] "this" but একটা/*ekʹṭā* [ækṭa] "one" and করি/*kari* [kori] "(I/we) do" but কর/*kara* [kɔro] "(you) do". However, the discrepancy between vowel graphemes and their phonetic value is well-known from many other languages too, particularly English.

More challenging seem to be several ligatures, i.e. special combinatory letters representing consonant clusters, and their pronunciation in Bengali. Written Bengali is rich in ligatures which can be found in words of various etymological origins. Although the ones found in terms of Perso-Arabic or English origin can be challenging, too, ligatures in words of Sanskrit origin cause more difficulties regarding pronunciation and especially their transcription into Roman script, since, as already pointed out above, the original Sanskrit orthography, which is retained, does not match the pronunciation which clearly indicates the sound shift known from Middle Indo-Aryan languages. Since it is beyond the scope of this essay to deal with the historic processes which have led to this phenomenon, a brief overview (with significant examples in Table 2) about these discrepancies is hopefully sufficient for illustrating the further challenges of transcribing Bengali into Roman script.

All these terms are also known in Hindi but, as a rule, pronounced differently from their Bengali versions. For instance, examples 1, 2 and 7 in Table 2 are spelled the same in Hindi: क्षेत्र/*kṣetra* [kʃeːt̪rə], शिक्षक/*śikṣak* [ʃikʃək] and स्मरण/*smaraṇ* [smərəɳ]/[smərəɳ̪]. Clearly, in contrast to Bengali, their pronunciation reflects their orthography much better. This is owed to the fact that in Modern Standard Hindi graphemes in ligatures carry predominantly the same phonetic value as when they are written singly, while in Bengali they are often either pronounced very differently in consonant clusters, or not at all. This refers overwhelmingly also to ব/*ba* as the last member of a ligature in words of Sanskrit origin.[22] In such cases, this grapheme, if it is the second member of a ligature,[23] is not pronounced at the beginning of the word (see example 9 in Table 2); internally, however, it is assimilated to the

[21] *All* Bengali vowels are short irrespective of orthography. However, when pronounced in a monosyllabic word standing isolated (i.e. not in a compound, with a suffix, or in a sentence), they are long irrespective of orthography.

[22] An exception is a compound in which a ligature arises due to an initial ব/*ba*, coalescing with the end consonant of the preceding compound member; if the compound is generally analysable as such, then the previously initial ব/*ba* retains its pronunciation. This is also mostly the case after the prefix উদ্/*ud*.

[23] If ব/*ba* is the third member of a ligature, it is not pronounced (see example 9 in Table 2).

preceding consonant, thus producing a geminated consonant (see example 10 in Table 2).[24] Moreover, if these words of Sanskrit origin are also part of the contemporary Hindi lexicon, then in the position of ব/*ba* in Bengali ligatures, we can find a ব/*va* in Hindi which represents the phoneme [ʊə].

Table 2: Discrepancies between spelling and pronouncing Bengali words of Sanskrit origin

	Ligature	**Example**	**TL**	**IPA**	**Meaning**
1.	ক্ষ = ক্ + ষ	ক্ষেত্র	*kṣetra*	[kʰeʈ̪ro]	field
2.	ক্ষ = ক্ + ষ	শিক্ষক	*śikṣak*	[ʃikkʰɔk]	teacher
3.	জ্ঞা = জ্ + ঞ + আ	বিজ্ঞান	*bijñān*	[biggæn̪]/ [biggan̪]	science
4.	দ্যা = দ্ + য্ + আ	উদ্যান	*udyān*	[ud̪d̪an̪]	garden
5.	ত্যু = ত্ + য্ + উ	মৃত্যু	*mṛtyu*	[mɾiʈ̪ʈ̪u]	death
6.	দ্ম = দ্ + ম	পদ্ম	*padma*	[pɔd̪d̪o]/[pɔd̪d̪õ]	lotus
7.	স্ম = স্ + ম	স্মরণ	*smaraṇ*	[ʃɔɾon̪]/[ʃɔ̃ɾon̪]	remembering
8.	ত্ত্ব = ত্ + ত্ + ব	নৃতত্ত্ব	*nṛtattba*	[n̪ɾiʈ̪ɔʈ̪ʈ̪o]	anthropology
9.	দ্ব = দ্ + ব	দ্বন্দ্ব	*dbandba*	[d̪ɔn̪d̪o]	pair, conflict
10.	শ্ব = শ্ + ব	বিশ্ববিদ্যালয়	*biśbabidyālay*	[biʃʃobid̪d̪alɔj]	university

In fact, Modern Standard Bengali does not possess the phoneme [ʊɔ] which is also reflected in its written form. Not only in ligatures in words of Sanskrit origin, where we can find a ব/*va* in the case of Hindi, is a ব/*ba* written in Bengali, but also in many other cases when these graphemes are not part of any consonant cluster. But in contrast to ব/*ba* as a last member in a consonant cluster, it is then pronounced as [bɔ]. This refers for instance to the first letter of example 10 in Table 2 which is written in Hindi विश्वविद्यालय/*viśvavidyālay* and pronounced [ʊiʃʊaʊid̪yaːlaj]. Other examples are the god "Shiva" (Hindi: शिव/*śiv* [ʃiʊ] / Bengali: শিব/*śib* [ʃib])[25], "colour, letter, caste" (Hindi: वर्ण/*varṇ* [ʊəɾɳ]/[ʊəɾn̪] / Bengali: বর্ণ/*barṇa* [bɔɾn̪o]) and "day" (Hindi: दिवस/*divas* [d̪iʊəs] / Bengali: দিবস/*dibas* [d̪ibɔʃ]).

Unsurprisingly, the fact that the Bengali ব/*ba* represents orthographically what is known from Sanskrit and also Hindi as ব/*va*, but at the same time stands as well

[24] This is realised like in Italian or Spanish, i.e. the tongue retains its position a bit longer than for a simple consonant. In this essay, geminated consonants are written in IPA as double consonants, as also reflected in Bengali orthography, e.g. চট্টোপাধ্যায় *Caṭṭopādhyāy* [cɔʈ̪ʈ̪opad̪d̪ʰaj] and not [cɔtːopad̪ʰːaj], since the use of the IPA symbol [ːˈ] to show gemination obliterates the aspiration and is thus unsuitable.

[25] Here [ʃiːb]; cf. note 21.

for the phoneme [bɔ] in general, also in words of English, Perso-Arabic or any other origin, leads sometimes not only to confusing transcriptions but also transliterations. In the context of transliterations, the phenomenon of the so-called ব-ফলা/*ba-phalā* is one example in this regard known from many academic publications: instead of simply transliterating the ব when it is the last member of a ligature as *ba* some scholars write a *va*, such as in the term for "world" (Hindi: বিশ্ব/*viśva* / Bengali: বিশ্ব/*biśba*, i.e. *biśva*). In this way, scholars either show that they are aware that the word is of Sanskrit origin or want to mark that ব/*ba* is in those cases not pronounced as [bɔ], or both. At times, the ব/*ba* written singly in words of Sanskrit origin is also transliterated as *va*, but then often solely in the case of terms which are also well known in their anglicised versions based on their Sanskrit form, such as Shiva, i.e. শিব as *śiva* (and not *śib*). In those cases, the impreciseness of transliteration is obvious.

Complicating Matters More: The English 'v' in 'Bengali' Script

Besides the discrepancies between Bengali phonology and orthography, especially in words of Sanskrit origin, and the challenges arising from those for transliterating and transcribing Bengali into Roman script, matters seem to become even more complicated since more and more English words have been entering the Bengali word-stock. In particular, the whole ৱ/*va*-ব/*ba* complex, combined with certain English words written in 'Bengali' script, has led to a new orthographic phenomenon of Bengali in Roman script which leads us to the example from the beginning of this essay: why do some Bengalis transcribe ভাগিনা/*bhāginā* [bʰagiɳa] ("nephew") as 'vagina' in Roman script?

I have had many lively discussions with Bengali friends about the missing grapheme in their language for the phoneme [ʊɔ]. And not once was this discussion finished with something we all agreed upon. *Inter alia* I observed that, on the one hand, people in general tend to mix up phonemes and graphemes, and on the other hand, many Bengalis I could talk to, especially from Bangladesh, are convinced that the script usually denoted as the "Bengali script" is perfect and can display any phoneme unambiguously. This assertion is then proven by giving examples of foreign, particularly English words containing a 'v' by writing them in 'Bengali' script: ড্রাইভার/*ḍrāibhār* [ɖraibʰaɾ] "driver", নার্ভ/*nārbh* [naɾbʰ] "nerve", ভিডিও/*bhiḍio* [bʰiɖio] "video", etc.

Seemingly, the ever growing number of English words in the Bengali lexicon has caused the presumption among some Bengalis that the Bengali grapheme ভ/*bha* [bʰɔ], commonly used for representing the English phoneme [v], should

always be transcribed as 'v' in Roman script, also in the case of words of South Asian origin. Besides ভাগিনা/*bhāginā* [bʰagiɳa] 'vagina' ("nephew"), this is also the case for, among others, ভালো/*bhālo* [bʰalo] 'valo' ("good"), ভাই/*bhāi* [bʰai] 'vai' ("brother") (see Figure 4), ভাইয়া/*bhāiẏā* [bʰaija] 'vaiya' ("elder bother") and ভাষা/*bhāṣā* [bʰaʃa] 'vasha', but also for first names of Sanskrit origin such as অমিতাভ/*amitābha* [ɔmiṯabʰo] 'Amitav' (in this case further complicated by contamination with the Hindi pronunciation [ɔmiṯabʰ] *amitābh*) and সৌরভ/*saurabh* [ʃourɔbʰ] 'Sourav'. However, while in the case of proper names such transcription makes it difficult to pronounce them properly only for people who do not know Bengali, the imprecise and inconsistent transcription of many other words makes it sometimes difficult even for Bengalis to understand each other.

Figure 5 illustrates how things can get even more complicated when a third script plays a role in the public sphere, a circumstance common, for instance, in the Indian union state of West Bengal. The name of the shop on the signboard on the very top (with yellow background colour) is written correctly in 'Bengali' script: সিদ্ধি বিনায়ক ট্রেডার্স/*siddhi bināẏak ṭreḍārs*. The term বিনায়ক/*bināẏak* is of Sanskrit origin and a synonym for the Hindu god Ganesha (*Gaṇeśa*). In Nagari, it is written with a व/*va* in the beginning, as is apparent from the Hindi/Modern Nagari version of the shop name on the left-hand side of the shop, where vertically is written: सिद्धी [sic!] विनायक ट्रेडार्स/*siddhī* [sic!] *vināyak ṭreḍārs*. The shop name in Roman script on top (with green background colour) and vertically in the middle is spelled in the same way as on the very top of the shop: Siddhi Vinayak Traders. But then the Bengali versions written on top and on the right side are in both cases: সিদ্ধি ভিনায়ক ট্রেডার্স/*siddhi bhināẏak ṭreḍārs*. In this case, বিনায়ক/*bināẏak* is spelled wrongly as ভিনায়ক/*bhināẏak* representing the 'v' of the Modern Nagari and Roman script versions as a ভ/*bha*, as it is normally done only for English words transcribed into 'Bengali' script.

Transcribing from Complicated to Complicated

While a transcription, as a rule, should serve the written reproduction of the pronunciation of a specific language into a particular script, especially the transcription of a language with a marked discrepancy between orthography and pronunciation, such as Bengali, into a familiar script used for another language, such as English, which also displays such a discrepancy, makes matters even more complicated. Other factors need to be considered, too: the most problematic being that the person transcribing Bengali into Roman script might know neither Bengali nor English in their written forms well enough. This applies to quite a few people in Bangladesh and India, where poor education is still widespread, but (luckily)

Figure 4 (left) and 5 (right) illustrating the ব/va-ৰ/ba complex in the public sphere in Cooch Behar (কোচবিহার/Koc'bihār), West Bengal, India (Photos by Carmen Brandt).

hardly stops people from participating in written communication. Moreover, the person also might not know spoken English well, and/or the language which is transcribed might not be Modern Standard Bengali but a regional spoken variant – which can cause even more irregularities, apparent only for people who not only know that the person transcribes this specific variant of Bengali but are also familiar with the latter. Above all, there may also be typing errors and autocorrections, especially in digital communication with mobile phones, which can distort written text too.

But even if these unpropitious circumstances do not apply, the discrepancies between spoken and written Bengali illustrated above, and the ones between spoken and written English, are more than enough to make the transcription from Bengali into Roman script, based on its English pronunciation, often an adventure. Not without reason did William Jones write already in 1787, as mentioned above, that "[o]ur *English* alphabet and orthography are disgracefully, and almost ridiculously, imperfect; and it would be impossible to express either *Indian*, *Persian* or *Arabian* words in *Roman* characters, as we are absurdly taught to pronounce them" (Jones 1787: 13). The example below illustrates a situation of the sort Jones might have had in mind when he wrote those harsh words.

The Spelling Adventures of Murad the Baldhead

The name "Murad Takla", Murad the Baldhead,[26] actually does not refer to a particular person. Rather, it has become a synonym for people who transcribe Bengali into Roman script in such an erroneous way that either the meaning can hardly be understood by others, or a funny twist is added to the writing due to the ambiguity of one or several words. According to one of the co-founders of the Facebook page called মুরাদ টাকলা (TL: *murād ṭāk'lā*) (Murād Ṭāk'lā 2020a), "Murad Takla" was born in 2012 when he saw an angry comment on Facebook which started with "murad takla" (Dhaka Tribune 2018). He and his friends were "scandalized by the horrible spelling" (ibid.) of this comment. Shortly after, they decided to start this Facebook page which, according to its "About" section, is dedicated to the fight against language distortion (Murād Ṭāk'lā 2020b)[27] and has today more than 360,000 followers. The whole sentence which led to the creation of this Facebook page is reproduced in Table 3, together with other examples taken from the article "The Legend of 'Murad Takla': Origin, Examples and Other Issues" published in the Bangladeshi English newspaper *The Daily Star* (Hasan 2017). However, only the name of the page shall receive our attention in detail.

Table 3: Examples by 'Murad Taklas'[28]

1.	source: abasar.net	Murad takla jukti dia bal, falti pic dicos kan! Lakapara koira kata bal,
	probable Bengali original:	মুরদ থাকলে যুক্তি দিয়ে বল্, ফালতু পিক দিচ্ছিস্ কেন, লেখাপড়া করে কথা বল্।
	transliteration:	*murad thāk'le yukti diẏe bal·, phāl'tu pik dicchis· kena, lekhāpaṙā kare kathā bal·.*
	translation:	If you have the strength, say with arguments; why are you blowing [your] whistle uselessly; talk [only] after receiving education.
2.	source: Hasan 2017	2016 amira bondary sabai cox'bazar ra onak moga korlam????

[26] I am grateful to Shabnam Surita for introducing me to Murad Takla.

[27] The complete "About" entry is: ভাষা বিকৃতির বিরুদ্ধে যুদ্ধ চলুক অবিরাম (TL: *bhāṣā bikṛtir biruddhe yuddha caluk abirām*) "may the war against language distortion go on uninterruptedly".

[28] Even though some of these examples might refer to local variants of Bengali, the probable Bengali originals in this table will, as a rule, refer to Modern Standard Bengali. Anything else would be speculation, since none of these examples seemingly follow a transcription system.

Table 3 (continued)

	transcription by Hasan:	2016 amra bondhura sobai Cox's Bazar e onek moja korlam!!!!
	probable Bengali original:	2016 আমরা বন্ধুরা সবাই কক্সবাজারে অনেক মজা করলাম!!!!
	transliteration:	*2016 ām'rā bandhurā sabāi kaks'bājāre anek majā kar'lām!!!!*
	translation:	In 2016, all we friends had a lot of fun in Cox's Bazar.
3.	source: Hasan 2017	Jakan rattri nijom nai coke gom, Akla sonna gare… Tomay mane pade Allah.., tomay mane pade[29]…
	transcription by Hasan:	Jokhon ratri nijhum nai chokhe ghum, Ekla sunno ghore Tomay mone pore Allah, Tomay mone pore
	probable Bengali original:	যখন রাত্রি নিঝুম নাই চোখে ঘুম, একলা শুন্য ঘরে তোমায় মনে পড়ে আল্লাহ, তোমায় মনে পড়ে।
	transliteration:	*yakhan rātri nijhum nāi cokhe ghum, ek'lā śunya ghare,* *tomāẏ mane paŕe āllāh, tomāẏ mane paŕe.*
	translation:	When the night is silent, no sleep is in the eyes; alone in [my] empty room, I remember you, Allah, I remember you.
4.	source: Hasan 2017	mayara shob kichu para ekta chalayar Mon vengtya para abar gortay para.
	probable Bengali original:	মেয়েরা সব কিছু পারে একটা ছেলের মন ভাঙতে পারে আবার গড়তে পারে।
	transliteration:	*meẏerā sab kichu pāre ek'ṭā cheler man bhāṅ'te pāre ābār gaŕ'te pāre.*
	translation:	Girls can do everything. They can break the heart of a boy [and] can shape [it] again.
5.	source: Hasan 2017	maja maja tobo dekha pai, cirodin kano pai na. kano magh asa hridoy o akasha, tumara dekhita pai na.
	probable Bengali original:	মাঝে মাঝে তব দেখা পাই, চিরদিন কেন পাই না। কেন মেঘ আসে হৃদয় ও আকাশে, তোমারে দেখিতে পাই না।
	transliteration:	*mājhe mājhe taba dekhā pāi, ciradin kena pāi nā. kena megh āse hṛdaẏ o ākāśe, tomare dekhite pāi na.*
	translation:	Sometimes I can see you, why can I not forever. Why do clouds come in [my] heart and in the sky; I cannot see you.
6.	source: Hasan 2017	Happiness is Baby Vagina ke shower korano and dustami kora.

[29] This may be read as *pāde* "farts".

Table 3 (continued)

	probable Bengali original:	Happiness is baby ভাগিনাকে shower করানো and দুষ্টামি করা।
	transliteration:	Happiness is baby *bhāginā̃ke* shower *karāno* and *duṣṭāmi karā*.
	translation:	Happiness is bathing [my] baby nephew and teasing [him].

The angry commentator wrote "murad takla", which seemingly should be মুরদ থাকলে/*murad thāk'le* [muro̪d̪ t̪hakle]. While the transcription of the first word as "murad" is quite common and hence less problematic (see Table 1 and the discussions on the inherent vowel above), the second word illustrates several other problems. For instance, spoken Standard Bengali distinguishes strictly between aspirated and unaspirated consonants, for instance ত/*ta* [t̪ɔ] and থ/*tha* [t̪ʰɔ]; Germanic languages such as English and German also have aspirated consonants in some positions, but these are not reproduced in the script, and, moreover, the standard forms of these languages have not the kind of dentals known from Bengali and other South Asian languages, and as such no equivalent of Bengali থ/*tha*. Furthermore, 'th' in English stands for two fricative sounds, as in "this" [ðɪs] and "thin" [θɪn], and not for an aspirate; writing "takla" instead of "thakla" might thus have been even a conscious choice to avoid this very different pronunciation peculiar to English.

But then, one might easily take the Roman letter 't' in "takla" to refer to the Bengali grapheme ট/*ṭa* [ʈɔ]. Bengali not only differentiates strictly between aspirated and unaspirated consonants, but also between the dentals ত/*ta* [t̪ɔ], থ/*tha* [t̪ʰɔ], দ/*da* [d̪ɔ] and ধ/*dha* [d̪ʰɔ] and the retroflexes ট/*ṭa* [ʈɔ], ঠ/*ṭha* [ʈʰɔ], ড/*ḍa* [ɖɔ] and ঢ/*ḍha* [ɖʰɔ]. The latter four phonemes are not known at all from English phonology, but interestingly the letters 'd' and 't' in English words are predominantly reproduced with ট/*ṭa* [ʈɔ] and ড/*ḍa* [ɖɔ] in Bengali,[30] such as ট্যাক্সি/*ṭyāksi* [ʈæksi] "taxi" and ডাক্তার/*ḍāktār* [ɖakʈaɾ] "doctor" (see also Figures 3 and 5). It is hence not surprising that "takla" could easily be understood as টাকলা/*ṭāk'lā*, a colloquial term for someone with a bald head (Bengali: টাক/*ṭāk* "bald"), also similarly known from Hindi/Urdu, Marathi and so-called Bollywood films (e.g. Marathi: टक्ला/*ṭak'lā* "bald").

However, the main problem when transcribing Bengali into Roman script is related to the vowels, due to the fact that written English does not have any obvious rules regarding the representation of vowel phonemes. Some phonemes are

[30] The reason is that the English alveolars, which these letters represent, have no equivalents in Bengali; to Bengali speakers, their closest approximations seem to be retroflexes. This refers also to other modern South Asian languages, such as Hindi.

represented by several graphemes, while one grapheme can represent different phonemes. For instance, the phoneme [ʌ] can be written as a 'u' (e.g. "f**u**n") or 'o' (e.g. "m**o**ney"), while the letter 'a' can be pronounced as [æ] (e.g. "c**a**t"), [ɒ] (e.g. "w**a**tch") or [ə] (e.g. "**a**live"). It is thus no surprise at all that the commentator transcribed the two different phonemes in the word থাকলে/*thāk'le* with the letter 'a', as "takla", instead of distinguishing between these two phonemes also in written form, particularly if we have here a non-standard pronunciation of *e* in those words as [æ]. Another Bengali word which is often transcribed in a similar manner is একটা/*ek'ṭā* [ækʈa] "one", i.e. as 'akta'; many similar examples may be found in Hasan 2017.

Thus, it should not be at all surprising when someone transcribes থাকলে/*thāk'le* as 'takla', especially if we consider that there is no effort from the side of government institutions in Bangladesh and India to introduce a standardised transcription system which children could already learn in school. The examples in Table 3 illustrate how adventurous some transcriptions can get. Except for the sentence which has led to the Facebook page of/for মুরাদ টাকলা (TL: *murād ṭāk'lā*), Murad the Baldhead, all other examples are taken from *The Daily Star* article (Hasan 2017), already mentioned above, for the very reason that the author, Mahdi Hasan, of this article makes fun of the transcriptions of other people and offers, in some cases, a transcription of his own which is also not unproblematic.

Even though these examples are to some extent extreme cases of unsystematic transcription, they are evidence that there is indeed a necessity to raise awareness for writing Bengali systematically in Roman script. It is particularly important to develop a system which is easy to handle for everyone, and on different digital devices – such as mobile phones. The remarks in the first half of this article, about the important role of the Roman script for South Asian languages – about its status as the *scripta franca* of South Asia – were hopefully convincing enough to conclude that ignoring this obvious problem can be no solution at all.[31] For instance, writing Bengali exclusively in 'Bengali' script and making fun of people who write Bengali in Roman script – as it is done by the administrators of the Facebook page মুরাদ টাকলা (TL: *murād ṭāk'lā*) – is not really helpful. By tying Bengali exclusively to only one script, the problem is only ignored and not approached pragmatically, and more importantly, systematically.

However, this problem is not one peculiar to Bengali and its script alone, but can be found, in varying degrees, with regard to other South Asian languages and their common scripts too. Bengali was chosen as it seems to be an extreme example, and thus can highlight the various problems associated with the use of the Roman script in a very graphic manner. Being aware of these problems and the

[31] For further elaborations on the Roman script being the *scripta franca* of South Asia see Brandt 2020.

fact that the Roman script was, is, and will continue to be used for Bengali and other South Asian languages – for various reasons and in various contexts and ways – I shall outline which way might be the best, in which context, in the conclusion below.

Conclusion

As illustrated in this essay, there is a huge difference between transliterating and transcribing a language into a script usually not employed for it. This refers to the methods and also the objectives. Doubtlessly, both transliteration and transcription are eligible, depending on the context in which one or the other is required or preferred. As a philologist, I hope that there is no doubt any more that, particularly in academic contexts, transliterating should be *the* priority when the author quotes from written sources. In this case, the accurate representation of a written book title and of the author's name in Roman script will help the reader to ascertain the original in 'Bengali' script and to detect the original source. Additionally, adding the original quote transliterated precisely, besides its translation, allows the reader to draw their own conclusions, since every translation is already an interpretation. Table 4 below refers to one of these transliteration systems for doing so and demonstrates that this endeavour is nowadays mostly made by using dots and dashes, i.e. diacritics.[32] However, Table 4 is only one example, and the one I prefer. There are several other systems, which might be as good as this one – as long as the reader is also aware which system the author employed, and the orthography of the Bengali original is represented accurately in Roman script.

One could argue that writing Bengali in Roman script should be obsolete – at least in academia. After all, technically it is no problem at all any more to write texts in different scripts, as was demonstrated in this essay. However, writing every language in its very own script, especially in the case of South Asia considering the variety of scripts, can hamper the comprehension of the written text. The very fact that many South Asian languages share the same or similar lexemes allows, for instance, a scholar of Hindi, Nepali, Oriya, Sanskrit or Urdu literature to understand at least some words or even more of a quote from Bengali literature, without knowing the 'Bengali' script, if the quote is reproduced in Roman script.[33] Since in this case the accurate orthographic reproduction of the written word is a precondition to enable the reader to make the comparison between different languages, the correct pronunciation does not play any role. This refers particularly to words of Sanskrit origin, such as দীপ/*dīp* ("light"),

[32] An exception is the Harvard-Kyoto System.

[33] This applies even more to Hindi and Urdu, which are basically the same language but written in different scripts.

দ্বিপ/*dbip* ("elephant") and দ্বীপ/*dbīp* ("island") which in Bengali are all pronounced as [d̪ip] in combination with suffixes or other words, and [d̪i:p] when isolated. Transcribing these three Bengali words in academic texts as 'dip' or 'deep' in Roman script, respectively, or writing them in IPA – [d̪ip] and [d̪i:p] – would be pretty confusing not only for scholars not familiar with Bengali, but even for Bengalis and scholars of Bengali, particularly if the context is missing. The confusion caused by transcribing Bengali in an unsystematic way was hopefully sufficiently demonstrated in this essay. But when should Bengali be transcribed and not transliterated in academia?

This case could apply to oral sources in general, such as folk songs, dialogues in films, speeches by politicians, but also interviews with people who do not speak Modern Standard Bengali. As a matter of fact, spoken Bengali, as most languages, can be so divergent that the reproduction of quotes by people who speak, for instance, Sylheti, Chittagonian, or any other local speech – which is predominantly categorised as a 'dialect'[34] of Modern Standard Bengali – can be quite a challenge. If one tries to transliterate those, one might face difficulties to do so since most local variants of Bengali lack a standardised orthography. One could opt for trying to represent this oral document by transliterating it into Roman script on the basis of an *ad hoc* orthography in 'Bengali' script. However, the phonology of some local variants is so different from Modern Standard Bengali that the present set of 'Bengali' characters and their transliterated variants in Roman script is not sufficient to represent those unambiguously.[35] Furthermore, this *ad hoc* orthography might already be an interpretation by the author, which can distort the original content. Nevertheless, this is still a better option than transliterating the oral quote according to the orthography of Modern Standard Bengali, which would mean relegating the original variant of Bengali under Modern Standard Bengali and, moreover, distorting it extensively. Especially in a region where illiteracy is still widely spread, the written standardised language would therefore be given more authority than the spoken, neglecting the actual diversity of spoken Bengali, depending on region, socioeconomic stratum, age, gender, etc. Therefore, in order to represent an oral source accurately in text, the scholar could employ the IPA, which is the only easily accessible system that can represent the spoken word in written form precisely. A less precise option could be a transcription system which is less

[34] "A language is a dialect with an army and navy", a quote attributed to a publication in Yiddish from 1945 by the linguist Max Weinreich (Splunder 2019: 36).

[35] This refers, for instance, to the phoneme [z], which is not represented by any distinguished 'Bengali' character, or a common character with an additional diacritic. As a rule, it is represented solely by জ/*ja* [ɟɔ], while in Modern Nagari for Hindi it is represented by the character ज/*ja* [ɟə] with a dot below, i.e. by ज़/*za* [zə]. Today, some publications (like the Calcutta journal *Deś*) have opted to reproduce sounds not provided for by the 'Bengali' script by copying the manner used for Hindi, namely by using subscribed dots, in the case above therefore জ়.

complicated than the IPA and might, in consequence, also be an option for writing Bengali in Roman script with mobile phones and on the Internet.

So, while William Jones postulated a system for transliterating languages systematically, there is now doubtlessly also a need to transcribe Bengali – as well as other South Asian languages – systematically to avoid misunderstandings, as illustrated above. Of course, the IPA is far too complicated to be employed for this task in everyday life and also by scholars who are not familiar with this system. However, I want to handle this matter in a similar way to William Jones, as in the case of his own transliteration rules – written more than 200 years ago – and "close this paper with specimens [for transcribing Bengali into Roman script systematically]; not as fixed standards of orthography, which no individual has a right to settle, but as examples of the method which I recommend" (Jones 1787: 33). Hence, the transcription rules below (Table 4) are only suggestions, and I neither claim them to be complete nor perfect.

Table 4: The transliteration (TL), International Phonetic Alphabet (IPA)[36], and systematic transcription (TC) suggested by me for the script used for Modern Standard Bengali

	অ	আ	ই	ঈ	উ
TL	*a*	*ā*	*i*	*ī*	*u*
IPA	ɔ/o	a	i	i	u
TC	ó/o	a	i	i	u
	ঋ	এ	ঐ	ও	ঔ
TL	*ṛ*	*e*	*ai*	*o*	*au*
IPA	ri	e/æ	o̯i	o	o̯u
TC	ri	e/ê	oi	o	ou
	ক	খ	গ	ঘ	ঙ
TL	*ka*	*kha*	*ga*	*gha*	*ṅa*
IPA	kɔ	kʰɔ	gɔ	gʰɔ	ŋɔ
TC	kó	khó	gó	ghó	nó
	চ	ছ	জ	ঝ	ঞ
TL	*ca*	*cha*	*ja*	*jha*	*ña*

[36] A complete chart of the IPA as PDF can be downloaded via the website of the International Phonetic Association: <https://www.internationalphoneticassociation.org/content/full-ipa-chart>. I abstain from marking the stress in Bengali terms since it is predominantly on the first syllable and hence much easier than in other languages.

Table 4 (continued)

IPA	cɔ	cʰɔ	ɟɔ	ɟʰɔ	nɔ/j̃ɔ̃
TC	có	chó	jó	jhó	nó/yó
	ট	ঠ	ড	ঢ	ণ
TL	*ṭa*	*ṭha*	*ḍa*	*ḍha*	*ṇa*
IPA	ʈɔ	ʈʰɔ	ɖɔ	ɖʰɔ	ɳɔ/ɳ̆ɔ
TC	Tó	Thó	Dó	Dhó	nó
	ত	থ	দ	ধ	ন
TL	*ta*	*tha*	*da*	*dha*	*na*
IPA	t̪ɔ	t̪ʰɔ	d̪ɔ	d̪ʰɔ	n̪ɔ
TC	tó	thó	dó	dhó	nó
	প	ফ	ব	ভ	ম
TL	*pa*	*pha*	*ba*	*bha*	*ma*
IPA	pɔ	pʰɔ/fɔ	bɔ	bʰɔ	mɔ
TC	pó	phó/fó	bó	bhó	mó
	য	র	ল	শ	ষ	স	হ
TL	*ya*	*ra*	*la*	*śa*	*ṣa*	*sa*	*ha*
IPA	ɟɔ	ɾɔ	lɔ	ʃɔ/sɔ	ʃɔ	sɔ/ʃɔ	hɔ
TC	jó	ró	ló	shó/só	shó/só	shó/só	hó
	ড়	ঢ়	য়	◌ং	◌ঃ	◌ঁ
TL	*ṙa*	*ṙha*	*ẏa*	*ṃ*	*ḥ*	~
IPA	ɽɔ	ɽʰɔ	jɔ	ŋ	ʿ	~
TC	Ró	Rhó	yó	ng	h/-	n/m[37]

As Table 4 also illustrates, a systematic transcription is not meant to replicate the orthography of Bengali words unambiguously in Roman script but, instead, serves to represent its pronunciation. Therefore, on the one hand, different characters of the 'Bengali' script can be represented by the same Roman letter or letter combination since they are pronounced in the same way, such as 'sh' for শ, ষ and স if they are

[37] Instead of adding a diacritic, as in transliterating the *candrabindu* showing vowel nasalisation, 'n' or 'm' is to be added behind the respective vowel since it is not possible to combine more than one diacritic with a vowel on mobile phones; cf., e.g., চাঁদ [cãn̪d] "moon", transliterated as *cãd*, but in my proposed transcription 'cand'.

pronounced as [ʃɔ]. On the other hand, one written 'Bengali' character can be represented by diverging Roman letters, as is the case for the vowels অ/*a* and এ/e. Moreover, the diacritics used for transcribing অ/*a* when it is pronounced as [ɔ], i.e. 'ó', and এ/e when it is pronounced as [æ], i.e. 'ê', might appear random, and that is what they indeed are: the main criteria relevant for me was that the diacritics can easily be added when typing on a computer and with mobile phones; they are found on every standard keyboard for languages using the Roman script and can be employed with mobile phones by pressing the key for the letter 'o' or 'e' for more than a second. Unfortunately, it is not that easy to add a diacritic to consonants, for which reason I suggest to write the retroflex phonemes in capital letters.[38] After all, the 'Bengali' script does not differentiate between lower-case and upper-case characters, and if everything – including words at the beginning of sentences and proper nouns – is accordingly written in small letters in Roman script, too, the capital letter can easily be recognised as representing a retroflex.

According to the transliteration system preferred by me[39] and the transcription rules suggested in Table 4, the sentence which led to the creation of the Facebook page মুরাদ টাকলা (TL: *murād ṭāk'lā* / TC: murad Takla) would, thus, be transliterated and transcribed in the following way:

Source: abasar.net	Murad takla jukti dia bal, falti pic dicos kan! Lakapara koira kata bal,
probable Bengali original:	মুরদ থাকলে যুক্তি দিয়ে বল্, ফালতু পিক দিচ্ছিস্ কেন, লেখাপড়া করে কথা বল্।

[38] According to the Harvard-Kyoto System, the retroflexes are also transliterated with capital letters. Such capitalisation is also found in various transliteration systems for Tamil.

[39] This transliteration system (see Table 4) is based on the International Alphabet of Sanskrit Transliteration (IAST) but includes some important additions and modifications, which make it more useful for transliterating New Indo-Aryan, exemplified here by Modern Standard Bengali. The modifications to this transliteration system by Rahul Peter Das (Das 2020: 6f.) refer mainly to vowels, such as marking the omission of the inherent vowel with a vertical dash if this is not marked in the original, e.g. আলবত/*āl'bat* "certainly", while vowel omission at the end of the word, as a rule, is not marked, e.g. মত/*mat* "opinion" (in contrast to মত/*mata* "similar to") (ibid.: 10). However, if the omission of the inherent vowel is marked, then this is marked in the transliteration with a middle dot, e.g. আলবৎ/*āl'bat·* "certainly" (ৎ = ত্) (ibid.). Ultimately, this transliteration system reflects the pronunciation at least to some extent, which can be essential, especially if two different words are written in the same way but pronounced differently, e.g. মত/*mat* or *mata*. Furthermore, to distinguish between the specific 'Bengali' characters representing the diphthongs – ঐ/*ai* and ঔ/*au* – and combinations of two other 'Bengali' vowel characters forming these sounds, in the latter case the second vowel is written with diaeresis, e.g. ঐ/*ai* but অই/*aï* and বৌ/*bau* but বউ/*baü* (ibid.: 14). Another exception is the character য/*ya* being part of a vowel cluster in the beginning of a word of foreign origin, such as in অ্যাডভোকেট/*äḍ'bhokeṭ* and এ্যাডভোকেট/*æḍ'bhokeṭ* "advocate" (ibid.: 67f.).

transliteration (TL):	*murad thāk'le yukti diẏe bal·, phāl'tu pik dicchis· kena, lekhāpaŕā kare kathā bal·.*
transcription (TC):	murod thakle jukti diye ból, faltu pik dicchish kêno, lêkhapóRa kore ból.

The main challenges regarding a simplified but systematic way of transcribing Bengali, so that its meaning can be widely understood, are particularly the unambiguous representation of its vowels, such as the inherent vowel, which is often transcribed erroneously with an 'a'. Other important domains are the differentiation between aspirated and non-aspirated consonants, between the dentals and retroflexes, and the complicated orthography of words of Sanskrit origin. The latter refers, for instance, to the examples in Table 2 which can be transcribed according to my transcription system as ক্ষেত্র 'khetro', শিক্ষক 'shikkhók', বিজ্ঞান 'biggên', উদ্যান 'uddan', মৃত্যু 'mrittu', পদ্ম 'póddo', স্মরণ 'shó(n)ron', নৃতত্ত্ব 'nritótto', দ্বন্দ্ব 'dóndo', and বিশ্ববিদ্যালয় 'bishshobiddalóy'.

Other challenges refer to words of foreign – for instance, Perso-Arabic, Portuguese or English origin – and to phonemes known from local variants of Bengali, but not Modern Standard Bengali. For example, this is the case for the phoneme [z], in written Bengali mostly represented with জ/*ja*, for instance in the Portuguese word কামিজ/*kāmij* "shirt", mostly pronounced as [kamiʤ] but, by some Bengalis, also as [kamiz]; or the word সোজা/*sojā* "straight", in Modern Standard Bengali [ʃoʤa] but in some local variants [ʃoza]. Hence, these words can be transcribed as 'kamij' and 'shoja', or 'kamiz' and 'shoza' if one wants to represent his/her actual pronunciation correctly.

Nonetheless, this simplified transcription system suggested by me still requires a sufficient command over spoken Bengali to recognise and pronounce everything correctly. Its usefulness for academia might hence be limited, especially if correct translations and a proper contextualisation are missing. Most importantly, if applied in academia, then this should be only the case for oral sources deviating from Modern Standard Bengali, and, by no means, for written ones. However, the main reason for this exercise was to address the problematic writing of Bengali in Roman script on the Internet and with mobile phones. And for avoiding misunderstandings *à la* Murad Takla, this transcription system is sufficient and easy to implement. Even though it seems unnecessary to mention the following, in these days, one cannot be careful enough to avoid accusations of a political nature: this essay does not suggest replacing the 'Bengali' script by the Roman alphabet for good. But in case Bengali *is* written in Roman script, this should be done in a way to ensure that misunderstandings are avoided.

References

Ahmed, A. F. Salahuddin 1965. *Social Ideas and Social Change in Bengal : 1818–1835*. Leiden: E. J. Brill.

Ahmed, Syed Jamil 2001. *In Praise of Nirañjan : Islam, Theatre and Bangladesh.* Dhaka: Pathak Shamabesh.

Bal, Ellen 2007. *They Ask If We Eat Frogs : Garo Ethnicity in Bangladesh*. Leiden: International Institute of Southeast Asian Studies; Singapore: ISEAS Publishing.

Bayly, C[hristopher]. A[lan]. 1996. *Empire and Information : Intelligence Gathering and Social Communication in India, 1780–1870*. Cambridge: Cambridge University Press.

Brandt, Carmen 2014. "Script as a Potential Demarcator and Stabilizer of Languages in South Asia", in: Hugo C. Cardoso (ed.): *Language Endangerment and Preservation in South Asia*. (Language Documentation & Conservation Special Publication 7.) Honolulu: University of Hawai'i Press, 78–99 (<http://hdl.handle.net/10125/4602>, accessed: November 5, 2020).

Brandt, Carmen & Kirsten Hackenbroch 2017. "Die deutsche Südasienforschung im Wandel der Zeit", in: *ASIEN* 144: 36–57.

Brandt, Carmen 2018. *The 'Bedes' of Bengal : Establishing an Ethnic Group through Portrayals*. (Transdisciplinary South Asian Studies 1.) Zürich: LIT Verlag.

Brandt, Carmen & Pushkar Sohoni 2018. "Script and Identity – The Politics of Writing in South Asia: An Introduction", in: *South Asian History and Culture* 9 (1): 1–15 (<https://doi.org/10.1080/19472498.2017.1411048>, accessed: November 5, 2020).

Brandt, Carmen 2020. "From a Symbol of Colonial Conquest to the Scripta Franca: The Roman Script for South Asian Languages", in: Brandt, Carmen & Hans Harder (eds): *Wege durchs Labyrinth : Festschrift zu Ehren von Rahul Peter Das*. Heidelberg; Berlin: CrossAsia-eBooks, 1–36 (<https://doi.org/10.11588/xabooks.642>, accessed: October 25, 2020).

Campbell, Lyle 2006. "Why Sir William Jones Got it All Wrong, or Jones' Role in How to Establish Language Families", in: *Anuario del Seminario de Filología Vasca Julio de Urquijo* 40: 245–264.

Chakravarti, Sudeep 2017. *The Bengalis : Portrait of a Community*. New Delhi: Aleph.

Das, Rahul Peter 2020. *Lehrbuch der bengalischen Hochsprachen.* 3. korr. Aufl. (Südasienwissenschaftliche Arbeitsblätter 14). Halle (Saale): Seminar für Südasienkunde und Indologie des Orientalischen Instituts der Martin-Luther-Universität Halle-Wittenberg (<https://doi.org/10.11588/xarep.00004387>, accessed October 25, 2020).

Dhaka Tribune 2018. "Murad Takla Turns 6, Still Fighting Strong", in: *Dhaka Tribune* (November 19, 2018) (<https://www.dhakatribune.com/feature/2018/11/19/murad-takla-turns-6-still-fighting-strong>, accessed September 21, 2020).

Eisenstein, Elizabeth L. 2012. *The Printing Revolution in Early Modern Europe*. 2nd ed. (Canto Classics Edition.) Cambridge: Cambridge University Press.

Halhed, Nathaniel Brassey 1778. *A Grammar of the Bengal Language*. Hoogly.

Hasan, Mahdi 2017. "The Legend of 'Murad Takla': Origin, Examples and Other Issues", in: *The Daily Star* (February 22, 2017) (<https://www.thedailystar.net/online/the-legend-murad-takla-origin-examples-and-other-issues-1365298>, accessed September 21, 2020).

Hoek, Lotte 2014. *Cut-pieces : Celluloid Obscenity and Popular Cinema in Bangladesh.* New York: Columbia University Press.

Jones, William 1806 [1787]. "A Dissertation on the Orthography of Asiatick Words in Roman Letters", in: *Asiatick Researches* 1 (5th ed.): 1–56.

Jones, William 1799 [1786]. "The Third Anniversary Discourse, Delivered 2 February, 1786: On the Hindus", in: *The Works of Sir William Jones*. In Six Volumes; Vol 1. London: Robinson, 19–34.

Majeed, Javed 2019. *Colonialism and Knowledge in Grierson's Linguistic Survey of India*. London; New York: Routledge.

Murād Ṭāk'lā 2020a. "মুরাদ টাকলা [Murād Ṭāk'lā]", in: *Facebook* (<https://www.facebook.com/MuradTakla>, accessed: September 21, 2020).

Murād Ṭāk'lā 2020b. "About", in: *Facebook* (<https://www.facebook.com/MuradTakla/about/?ref=page_internal>, accessed September 21, 2020).

Ross, Fiona G. E. 1999. *The Printed Bengali Character and its Evolution*. Richmond: Curzon.

Ross, Fiona 2018. "Historical Technological Impacts on the Visual Representation of Language with Reference to South-Asian Typeforms", in: *Philological Encounters* 3 (4): 441–468 (<https://doi.org/10.1163/24519197-12340054>, accessed October 25, 2020).

Splunder, Frank van 2020. *Language is Politics : Exploring the Ecological Approach to Language*. London; New York: Routledge.

The Independent 2010. "'You Ate Where?' Ten 'Worst-Named' Restaurants", in: *The Independent* (May 2, 2010) (<https://www.independent.co.uk/lifestyle/food-and-drink/you-ate-where-ten-worst-named-restaurants-5538385.html>, accessed August 2, 2020).

Unicode 2020. "Supported Scripts", in: *Unicode Standard* (<https://unicode.org/standard/supported.html>, accessed September 21, 2020).