

# Language Documentation and Description

ISSN 2756-1224

---

This article appears in: *Language Documentation and Description*,  
vol 18. Editors: Candide Simard, Sarah M. Dopierala & E. Marie Thaut

## Does Camden Sylheti have consonant clusters?

S. ELIZABETH EDEN

Cite this article: Eden, S. Elizabeth. 2020. Does Camden Sylheti have  
consonant clusters? In Candide Simard, Sarah M. Dopierala & E. Marie  
Thaut (eds.) *Language Documentation and Description* 18, 23-41.  
London: EL Publishing.

Link to this article: <http://www.elpublishing.org/PID/197>

This electronic version first published: August 2020



This article is published under a Creative Commons  
License CC-BY-NC (Attribution-NonCommercial). The  
licence permits users to use, reproduce, disseminate  
or display the article provided that the author is attributed as the  
original creator and that the reuse is restricted to non-commercial  
purposes i.e. research or educational use. See  
<http://creativecommons.org/licenses/by-nc/4.0/>

---

## EL Publishing

For more EL Publishing articles and services:

Website: <http://www.elpublishing.org>  
Submissions: <http://www.elpublishing.org/submissions>

# Does Camden Sylheti have consonant clusters?

S. Elizabeth Eden  
*Independent scholar*

## Abstract

By examining the distribution of consonantal segments word-medially in Camden Sylheti, as documented in the SOAS Sylheti Dictionary, I will show that the native lexicon has only (C)V(C) syllables, contrary to previously observed transfer effects seen in Sylheti-English bilinguals. I will briefly discuss repair strategies observed in the Sylheti lexicon for historic loan words, and question which of these are still extant. I will also demonstrate the computational tools which I have developed to help with such analyses.

A simple view of syllable structure is that any sequence of consonants that occurs word-initially is a valid syllable onset, and any sequence that occurs word-finally is a valid coda (Zec 2007). However, a closer cross-linguistic examination shows that consonant distributions may vary between word-edge and word-medial positions (Kaye 1990).

Word-edge consonant sequences in Sylheti are relatively rare. Word-initial clusters are predominantly found in loan words e.g. English [profesər], [skol] and Sanskrit [ɡrɑm], [prɑfnɔ]. Likewise, most of the final-consonant sequences in the Dictionary are either loan items e.g. [ɡɪf], or morpheme edges which do not surface without a following vowel e.g. [ɑfn-]. However, word-medial consonant sequences are relatively common in the native lexicon.

Word-medial sequences can be divided into valid word-final sequences and remaining consonants. In English, for example, [ŋɡɪ] in angry splits into [ŋ]/[ɡɪ], implying that [ɡɪ] is a valid word-medial complex onset. For Sylheti, I demonstrate that the word initial ‘complex onsets’ are not in fact found word-medially; they never follow a valid word-final sequence. I similarly demonstrate that only two of the 27 word-final sequences are found as internal codas. The status of these two sequences (as allophonic variants or loan phonology) is debatable. By contrast, all word-initial singletons are also found as word-medial onsets, and all word-final singletons are also found as word-medial codas.

I conclude that all word-medial consonant sequences in the native lexicon contain syllable boundaries.

## 1. Background

In this paper, I present a systematic overview of consonant sequences in Camden Sylheti.<sup>1</sup> This account brings together information on inventory and syllable structure with repair strategies and cross-linguistic comparisons, comparing Sylheti to other Eastern Indo-Aryan languages. It also offers a case study for the use of Nidaba, my lexical analysis and comparison tool, in assisting phonological analyses.

I start with introductions to: the role of syllable structure in phonology; the Camden Sylheti lexicon; the functions of Nidaba; and the evidence for marginal contrasts found in this data. I examine the consonantal inventory of Sylheti in Section 2, syllable finals in Section 3, syllable initials in Section 4, and the origins of marginal contrasts in Section 5.

### 1.1 Syllable structure

Syllables are a way of organising consonants (C) and vowels (V). The basic structure of a syllable consists of an onset and a rime. The rime consists of a nucleus (generally a vowel) and a coda. The universal syllable, found in every language, has the shape CV (cf. Breen & Pensalfini 1999), i.e. a single consonant onset, and a rime containing just a nucleus. Therefore, initial vowels are cross-linguistically less common, as are initial consonant clusters. Likewise, codas are dispreferred, so final consonants and internal consonant sequences are not universal (e.g. Blevins 1995); final consonant sequences are particularly rare.

A simple view of syllable structure is that any sequence of consonants that occurs word-initially is a valid syllable onset, and any sequence that occurs word-finally is a valid coda (Zec 2007). However, a closer cross-linguistic examination shows that consonant distributions may vary between word-edge and word-medial positions.

It is worth investigating the syllable structure of a language beyond superficial phonotactic observations because it can have concrete effects on production and perception. For example, the order of acquisition in children can be determined by syllabic role (Pan & Snyder 2004), as can the perception of non-native consonant sequences (Kabak & Idsardi 2007). In the case of Sylheti, studies of acquisition currently rely on data from Bangla, due to a paucity of alternative sources (McCarthy & Skoruppa 2015).

---

<sup>1</sup> I thank Vijay D'Souza, Farhana Ferdous, Kathleen McCarthy, Candide Simard, and Marie Thaut for insight and discussion, and two anonymous reviewers for their comments.

Furthermore, it is always worth expanding our typological knowledge so that theoretical work on the syllable and alternative organisational structures is based on data, not assumptions.

## 1.2 Camden Sylheti data

My main source of data is the Sylheti lexicon compiled by the *SOAS Sylheti Project* (see Simard, Dopierala & Thaut, this volume), as of November 2016, consisting of 4,525 lexical items. This is supplemented by consultation with members of the project, and in particular with native Sylheti speaker Farhana Ferdous.

Given the multilingual nature of the Sylheti community, particularly those consultants whose work provides the foundation of the SOAS Sylheti Project, the lexicon may contain examples of Bangla words, drawn from sources where code-switching has occurred unnoticed. Various English-language items have also entered the lexicon, likely because the community has been in regular contact with the English-speaking population of London for several generations now. It is not known how the frequency and phonology of these Camden Sylheti loan items compare to Bangladeshi or Indian Sylheti.

Examples throughout are glossed in English. Where possible I have included Bangla translations, and Sanskrit items which I infer to be cognate, with IAST transliteration. All Sylheti examples and their glosses are taken from the SOAS Sylheti corpus, via Nidaba (see below). Sanskrit items were sourced from Glashoff et al. (2017).

## 1.3 Nidaba

Nidaba is an online database and lexical analysis tool which I have developed in the course of my doctoral research.<sup>2</sup> Its core functions are search and comparison of segmental patterns in transcribed lexicons.

To analyse a language with Nidaba, two sets of input data are required:

1. a list of lexical items in some transcription system, together with any data with which the researcher would like to tag items (e.g. English gloss, part of speech, origin of loan items); and
2. a conversion from that transcription system to IPA transcription.

Initially, conversion will be a simple phonetic mapping. This stage allows the researcher to confirm the phonetic inventory of their initial transcription,

---

<sup>2</sup> <http://nidaba.co.uk> (accessed 2020-06-19)

identifying any typographical errors (e.g. [c] in place of [k]). The mapping system can handle combinations of characters, using a longest-match-first approach. This allows for lexicons derived from semi-regular orthographic systems.

Once a lexicon has been uploaded, the researcher can compare the occurrence of different segments in different positions (word-initial, word-medial and word-final), which can assist in identifying allophones. Once the researcher has completed a phonemic analysis, the list of lexical items can be retranscribed with a new, phonemic, mapping, for use in further analysis.

The principal analysis tool in Nidaba is a search for segmental patterns. For any given transcription of a lexicon, the researcher can view word-initial, word-medial or word-final sequences of vowels or consonants. These sequences are displayed with the number of lexical items in which they are found, and a link to all known examples. This latter feature can help in discovering commonalities, such as all examples of a given sequence deriving from the same morpheme. If corpus frequency data is available, this tool can also give the total frequency of a sequence summed over all items, and similar statistics.

From this basic overview, more detailed searches can be conducted. The researcher can specify properties of sequences such as:

- length, number of items, frequency, or sonority profile;
- place, manner and/or voicing features; and
- part of speech or other lexical tags.

A second analysis tool in Nidaba is set comparison. The results of the detailed searches can be automatically compared, making it easy to see e.g. which sequences occur word-initially but not word-finally, in nouns but not in verbs, or in high frequency items but not in numerous ones.

A third analysis tool in Nidaba locates sub-sequences, which can then be fed into the set comparison tool (see Section 3.2).

The final analysis tool used in this paper is a minimal set generator. This tool produces lists of all minimal pairs, triplets, etc. in a lexicon. Thus, it also produces a list of all phonemic contrasts with documented examples.

## 1.4 Locating contrasts

In this section, I discuss the criteria available for deciding whether two sounds A and B are *contrastive* or *allophonic* (predictably distributed variants) using the Camden Sylheti lexicon. In some situations, such a decision is difficult or even undesirable, giving rise to marginal contrasts. However, the existence of marginal contrasts does not negate the usefulness of performing a phonemic

analysis. Indeed, locating ambiguous situations is a key step in describing a language accurately. Hall (2013) gives the following criteria: predictability of distribution, phonological behaviour, phonetic similarity, commutation, native speaker judgement, orthography, and alternations.

Nidaba’s core tools are designed for locating patterns in lexical datasets. The primary types of evidence used in this paper are therefore predictability of distribution and phonological behaviour. In addition, in interpreting the results, I have relied on the admittedly imprecise notion of phonetic similarity, assuming that sounds which share neither place nor manner nor voicing are unlikely to be allophones. Nidaba also contains a tool for identifying commutation, but the majority of the resulting pairs have not been verified with a native speaker.

Regarding native speaker judgements, speakers of Camden Sylheti are generally also speakers of Bangla and British English to some extent. The principal consultants for the lexicon speak all these languages, among others. Bilingualism influences both production and perception (e.g. Pavlenko 2000), but if code-switching is a feature of the community, there may not be a separate monolingual ‘standard’.

Regarding orthography, Sylheti is for the most part unwritten, with speakers writing in Bangla, the medium of education (see Simard, Dopierala & Thaut, this volume). Historically, Sylheti was written with the Sylheti-Nagri script, but this cannot be relied upon as a guide to modern Sylheti. For example, Sylheti-Nagri contains an aspiration contrast in stops which modern Sylheti lacks, causing difficulties for Sylheti speakers who are trying to revive the script (SOAS Sylheti Language Society, personal communication, 2015-01-27).

This paper lacks evidence from morphological alternations. However, Sylheti is highly inflected, so that would be a fruitful area for future research. I return to the origins of marginal contrasts in Section 7.

## 2. Consonant phonemes

In this section, I establish a probable phonemic inventory for Camden Sylheti, using Nidaba to examine the consonants found in the raw phonetic transcriptions of the SOAS Sylheti Project (see Table 1).

*Table 1: Full set of consonants used in phonetic transcription*

p	b	ṭ	ḍ	t̥	d̥	t̪	d̪	k	g	
f		s	z	ʃ		ʒ		x		h
	m		n		ɳ				ŋ	
		l	r		ɽ					

The subset of these consonants found as *singletons*, not neighbouring any other consonants, is found in Table 2. Those consonants not found in all positions (initial, medial and final) are in parentheses; consonants not found as singletons in any position are replaced with a dash.

Table 2: *Singleton consonants*

(p)	b	<u>t</u>	<u>ɖ</u>	t	ɖ	<u>tʃ</u>	( <u>ɖʒ</u> )	k	g	
f		s	z	–		ʃ		x		(h)
	m		n		–				(ŋ)	
		l	r		(ɾ)					

## 2.1 Nasals

The retroflex nasal [ŋ] is only found preceding retroflex stops. Given the relative incidence of homorganic nasal-stop sequences compared to heterorganic sequences for other nasals (examined in further detail in Section 3.1), and the complete absence of any alveolar nasal-retroflex stop sequences, I conclude that [ŋ] is an allophone of /n/.

The velar nasal [ŋ] is not found word-initially, and like the other nasals, is most commonly found in homorganic sequences. Whilst found in many fewer items than the labial or alveolar nasal (comparing only instances in word-medial or word-final position), I do not conclude that it is an allophone of /n/. A large proportion of word-medial sequences containing [ŋ] are heterorganic, and the majority of word-final occurrences are in isolation. This latter could be an artefact of the transcription, since transcribers were often native speakers of English, which forbids [ŋg] final sequences. However, this seems unlikely given the word-medial data. This is a separate phoneme of other eastern Indo-Aryan languages, including Bangla (Khan 2010), and Assamese (Mahanta 2012). It is found contrasting with both /m/ and /n/: [gam] ‘sweat’; [gan] ‘song’ and [gan] ‘river’.

The status of final nasal-stop clusters in Sylheti is discussed in Section 3.1.

## 2.2 Retroflex flap and stop

Like the velar nasal, the retroflex flap [ɾ] is also not found word-initially. By contrast, the voiced retroflex stop [ɖ] is only found word-finally in two items, [blɛɪɖ] ‘blade (of grass)’ and [bɛrɛɖ] ‘bread’. These are probably borrowed: both items have synonyms with Bangla cognates, and English alveolar stops are borrowed as retroflexes in most Indo-Aryan languages.

These two consonants are not quite in complementary distribution in word-medial position. [d̪] is found word-medially between two vowels in 11 items, whereas [ɽ] is found in 112. [d̪] is also found following [ŋ] and as a geminate; and in [maɽd̪al] ‘to strain’ and [d̪ald̪a] ‘Dalda’, a brand name. [ɽ] is found preceding [b], [d], [n], [t], [d̪], [k], and [ʃ]; following [m]; and in [fifɽa] ‘ant’, [laxɽi] ‘wood’, [zɔgɽa] ‘argument’ and [lɛŋgɽa] ‘lame’.

The distribution of these two sounds in Sylheti appears to be similar to that in other Indo-Aryan languages, such as Bangla and Hindi, including the apparent contrast found in loan words (Dasgupta 2003; Masica 1991: 91, 97; Śa’ 2001).

Both of these sounds are found contrasting with the voiceless retroflex stop [ʈ]. For example, [aʈ] ‘eight’ versus [aɽ] ‘month (3)’ and [d̪ali] ‘solider’ versus [ʈali] ‘pan’.

The aspiration contrast in voiced retroflexes was lost in Assamese and Bangla during the development of New Indo-Aryan (Cardona 2017); Sylheti preserves the contrast between retroflexes and the dental rhotic that was subsequently lost in Assamese.

### 2.3 Affricates

The postalveolar affricate [d̪ʒ] is not found individually in Sylheti; the Sylheti cognates of Bangla words containing [d̪ʒ] are realised with [z] (Ferdous p.c.). This is the same development found in Assamese and neighbouring Bengali dialects (Masica 1991: 94-95). With the development of fricative [z] from the voiced stop [d̪] (via [d̪ʒ]), Sylheti now has a voicing opposition in its fricatives, unlike most Indo-Aryan languages. For example, [sal] ‘ash’ versus [zal] ‘net’.

Using Nidaba’s transcription search, I find that [d̪ʒ] is only present in the contexts [nd̪ʒ] and [d̪ʒd̪ʒ]. Appearances in other contexts are as a variant of [z], possibly Bangla forms: [xɔɪld̪ʒa] (a variant of [xɔɪlza] ‘liver’); [rad̪ʒniti] (a variant of [razniti], ‘politics’); [ɽɔrd̪ʒɔni] (a variant of [ɽɔrɔni], ‘ring finger’); and as an English loan [sad̪ʒɛʃt-xɔɽ] ‘to suggest’.

[tʃ] is found individually predominantly in loan items: [tʃɛri] ‘cherry’, [tʃɔkɔɛt] ‘chocolate’, [bitʃ] ‘shore (beach)’, and [pɔtʃɪr] ‘enough’. Like [d̪ʒ], [tʃ] is found in the contexts [ntʃ] and [tʃtʃ]. Otherwise, it occurs only in [lalʃɛ] ‘reason’, [tʃɪp] ‘quiet’ and [tʃɔk] ‘bright’. The vast majority of [tʃ]-initial Bangla glosses in the lexicon correspond to [s]-initial Sylheti items.

The majority of nasal-affricate sequences correspond to Bangla nasal (vowel)-affricate sequences. It appears that post-nasal position is enough to protect the affricate from lenition, which accords with the cross-linguistic phenomenon of post-nasal fortition.

Based solely on the distribution of these two affricates in native Sylheti words, I would conclude that they behave, and should be treated, identically.



However, native speakers produce loan items differently in the two cases: [d̥ʒ] is pronounced as [z], but [t̥ʃ] is retained. It may be that Camden Sylheti is transitioning, or has already transitioned, to treating [t̥ʃ] as a phoneme in its own right.

## 2.4 Other fricatives

The retroflex fricative [ʂ] is only found before the retroflex stop [ʈ]; it is an allophone of either /s/ or /ʃ/, both of which occur independently.

The glottal fricative [h] is not found in consonant sequences (with the sole exception [brahmi] ‘type of plant’). It is found word-initially but not finally, and contrasts with the other fricative phonemes, e.g. [xasi] ‘knife’, [xaʃi] ‘cough’, and [xahi] ‘bowl’. [h] predominantly corresponds to Bangla [ʃ], with 61 [ʃ]-initial and 20 [h]-initial Bangla translations of Sylheti [h]-initial words. These are reflexes of Old Indo-Aryan /s/ (Pattanayak 1966; Masica 1991: 186).

Unlike in Assamese, [h] is not an allophone of [x]: [hɔɾ] ‘to move’ contrasts with [xɔɾ] ‘to do’. Instead, [x] and [k] are allophones. [k] is found preceding or following a high vowel, as a geminate, and in a few loan items, with [x] found elsewhere. Given the existence of a number of loan items with [k] where [x] would usually be expected (e.g. [nɛkɫɛs], [kampuɾɔɾ]), it is possible that the allophony rule has become fossilised. For example, the borrowed word [ɾɪʃka] ‘rickshaw’ has had metathesis applied, but [k] is retained as though still in the environment of a high vowel. We may see a split into two separate phonemes over the next few decades, particularly if there is an influx of English loanwords into Camden Sylheti.

## 2.5 Labials

The voiceless labial stop [p] is found only infrequently, and predominantly in two environments: following a labial nasal, and word-initially in the sequence [pɾ]. Items which are cognates with, or loans of, English items that contain [p] usually have [f] instead. Several items in the lexicon are recorded with both pronunciations (e.g. [ɪʃtɛmp] / [ɪʃtɛmf], [sappanno] / [saffanno]). I therefore conclude that [p] is an allophone of /f/. In terms of the development of this allophony, the fricative /f/ may be pronounced as [ɸ] or [f]; it may be that exposure to English labiodental [f] in Camden Sylheti is having an effect. Most Indo-Aryan languages have aspirated versions of both voiceless and voiced stops; in Bangla, the aspirated labial stops tend to be pronounced as bilabial fricatives (Masica 1991: 103).

## 2.6 Phonemic inventory

Table 3: Consonantal phonemes of Sylheti

b	<u>t</u>	<u>d</u>	t	<u>tʃ</u>	k	g	
f	s	z		ʃ			h
m		n				ŋ	
	l	r	ɾ				

The most striking feature of Sylheti as an Indo-Aryan language is the absence of contrastive aspiration. Bangla has lost the aspiration contrast in final and pre-consonantal position (Pattanayak 1966; Masica 1991: 204). Sylheti, like other East Bengali dialects (Masica 1991: 103), has extended this merger to all positions.

Sylheti appears to maintain the basic Indo-Aryan system of five places for stops: bilabial, dental, retroflex, palatal, and velar (see Table 3). However, the voiceless stop /p/ has mostly been lost through spirantization, as in Chittagong Bengali (Masica 1991: 103). Cognates of Sanskrit /c/ are now predominantly found with /s/, and [tʃ] is found in non-cognate items.

In terms of innovations from Middle Indo-Aryan to New Indo-Aryan, Sylheti parallels Bangla in its voiced retroflexes and loss of (final) aspiration, Assamese in its lack of palatal stops or affricates, and both Bangla and Assamese in use of the velar nasal.

## 2.7 Comparison to other proposed inventories

This phonemic inventory is broadly similar to the transliteration scheme developed by Baker et al. (2000), which, however, lacks the voiceless retroflex stop, whose existence has been confirmed here by native speaker judgements. I have used the minimal pair [at] ‘eight’ and [at] ‘hand’ to confirm the correct transcription of multiple lexical items. Baker et al.’s scheme treats [tʃ] and [s] as allophonic, parallel to their voiced counterparts, and contains no alveolar tap or trill. It also lacks the voiced retroflex stop as an allophone of the continuant, but is not intended to give a narrow phonetic transcription.

By contrast, the inventory given for Sylheti in Singha & Ahmed (2016) bears a closer resemblance to Bangla, with aspirate counterparts for all stops, and no labiodental fricative.

Baker et al. (2000) developed their scheme for Sylheti as it is spoken in the UK, whereas the spoken corpus data used in Singha & Ahmed (2016) was presumably collected in Assam and Sylhet. This may account for some differences, but more research is required to isolate the different factors which influence Sylheti pronunciation.

## 2.8 Natural classes

Sylheti has a deficient voicing opposition in its stops, with the voiceless bilabial stop missing, as is typologically common (Gamkrelidze 1975). The Sylheti speaking community has a long history of contact with Arabic, which also follows this pattern.

Sylheti has a marginal voicing contrast in its fricatives, with only the alveolar fricative having a voiced counterpart. Again, this is typologically common, with the coronal place of articulation being considered unmarked (e.g. Rice 2007), and so more capable of supporting contrast.

Stops, fricatives and nasals all display contrast in place between bilabial, dental/alveolar and velar/glottal. By contrast, there are manners of articulation missing from both the retroflex and palatal series. This is a feature of Assamese-Bengali (Pattanayak 1966) and many other Indo-Aryan languages (Masica 1991: 94-99).

## 3. Syllable finals

In this section, I discuss deviations from the canonical CV syllable structure relating to codas.

Firstly, let us examine sequences found in word-final position in the lexicon. The lexicon contains both stems and bound morphemes. It therefore contains final sequences such as [fn] belonging to bound morphemes [afn-] which do not appear as free morphemes, but only with a following vowel. These can be filtered out by Nidaba if they are consistently tagged in the source data. If not, a combination of other tags such as part of speech data can be used to eliminate verbal roots and other bound morphemes.

### 3.1 Word-final codas

Sylheti has word-final codas, with nearly 45% of items ending in a consonant. The following consonants and clusters were found finally, in order of decreasing frequency: [ɾ], [l], [n], [ʃ], [t], [m], [tʃ], [x], [s], [tʰ], [k], [z], [d], [f] (>1% of items); [b], [g], [ŋ], [nd], [ndʒ], [ŋt] (>0.1% of items).

Voiced obstruents were not permitted in Sanskrit codas (Kessler 1994); this may account for the low frequency of [b], [g] and [d] relative to their voiceless counterparts.

Setting aside sequences found in only one item, and those mostly occurring in loan items, e.g. [ɛbarɪst] ‘Everest’, we find the following multi-segment sequences: [nd], [ndʒ], [ŋt], [rt] and [rd].

[nd] is found in verbal stems, and in nouns (see Table 4). These are mostly cognate with Bangla nouns which have a nasal vowel, instead of a nasal+stop cluster. NC (nasal-consonant) clusters were present in the protolanguage of

Assamese-Bengali (see Table 5), though many were subsequently lost through a variety of processes (Pattanayak 1966). Final clusters are not allowed in modern Bangla, but are in Assamese (Masica 1991: 126). More investigation is needed to determine whether Sylheti retained the NC clusters like Oriya, or redeveloped them more recently from a nasalised vowel system like Bangla's.

Table 4: Examples of word-final [nd] in nouns

Sylheti	English	Bangla		Sanskrit	
[tɔbɔnd]	knot	বাঁধন	[bãd <sup>h</sup> ana]	बन्ध	<bandha>
[sand]	moon	চাঁদ	[tʃãd]	चन्द्र	<candra>
[xand]	shoulder	কাঁধ	[kãd <sup>h</sup> a]	स्कन्ध	<skandha>
[damand]	son-in-law	জামাতা	[dʒamata]	जमात्	<jamātr>
[fand]	trap	ফাঁদ	[p <sup>h</sup> ãda]		
[ɪŋɡland]	England	ইংল্যান্ড			

Table 5: Correspondences involving NC clusters

	Bengali	Assamese	English	Reconstructed form
ɔŋ	ɔŋ	ɔŋ	colour	*ɔŋg
ɔŋga	-	ɔŋga	red	*ɔŋg
aʃ	hãs	pati hãh	duck	-
sand	tʃãd	sɔndrɔ	moon	*tʃãnd

[ndʒ] is found in a single morphological item, [gɔndʒ] <গঞ্জ> 'district', and in place names derived from it: [hɔbigɔndʒ] 'Habiganj', [xɔrimgɔndʒ] 'Karimganj', [sunamgɔndʒ] 'Sunamganj'.

[rd] is found in the nouns [mɔrd] 'man' and [dɔrd] 'pain', which appear to be loan items from Persian (Cormac Anderson, p.c.): <مرد *mrd*> [mærd] and <درد *drd*> [dærd].

[ŋt] is found in four items which appear to be loans from English: [kɔrɔŋt] 'electricity (current)', [rɛʃtʃurɔŋt] 'restaurant', [fɛŋt] 'trousers, pants', and [happɛŋt], 'shorts, half pants'. Likewise, [rt] is found only in [ɛrfɔrt] 'airport' and [ʃart], 'shirt'.

Sylheti is more tolerant of syllable structure violations than segment quality violations; there are no cases where [p] is retained but a complex onset or coda is repaired. By contrast, in [happɛŋt], not only is [p] retained, but [f] is adapted to match it. We have seen that [tʃ] and [dʒ] are protected from spirantization in geminates. Sylheti does not allow differing allophones within a sequence, and has a preference for stops over fricatives in geminates, resulting in these 'non-native' geminates in all three cases.

Regarding the other segment quality adaptations, we have seen that English alveolar stops are borrowed as retroflexes. Nasals and fricatives are

normally borrowed as dental/alveolar (e.g. [brɪʈan] ‘Britain’, [profɛsar] ‘Professor’), but undergo place assimilation to retroflex, as in native items.

[ɛrfɔrt] and [ʃart] have been borrowed from a rhotic variety of English (cf. Masica 1991: 75-76). In both onset and coda position, [ɾ] is borrowed as dental/alveolar, and does not undergo place assimilation. [ɾ] is an allophone of /d/, and the sequence \*/dɾ/ would be ungrammatical; Sylheti does not have any homorganic stop sequences which differ in voicing. This results in the unusual sequence [ɾt], otherwise found only in the loan item [xarʈɔn] ‘curtain’ and the pronouns [arʈa] ‘next’ and [amarʈa] ‘mine’.

### 3.2 Word-internal codas

In this section, I examine evidence for word-internal codas in Sylheti, derived from the application of Nidaba’s subsequence and set comparison tools.

Consider [ʃɔnda] as an example: it contains the word-internal consonant sequence [nd]. [d] appears word-initially in the lexicon, but [nd] does not. The longest possible internal ‘onset’ sequence in [ʃɔnda] is therefore [d], leaving [n] as the preceding coda.

The set of word-final consonants can be compared to the set of internal codas calculated this way. Such a comparison shows that all word-final consonants can also occur as word-internal codas.

There are only a few items transcribed with CCC word-internal sequences. Firstly, I examine two bimorphemic items, [dɔkknɔr] ‘southern’ and [ɔttɾɛ] ‘northern’. They appear to be formed by suffixation plus deletion from [dɔkkin] ‘south’ and [ɔttɔr] ‘north’. In Bangla, there is a preference for disyllabic trochees, which Nagarajan (2014) proposes has been the case since at least the 17<sup>th</sup> century. This may account for the deletion. However, there is limited other evidence of this preference in the Sylheti lexicon, since the creation of disyllables through epenthesis (see Section 4.5) is more easily explained as a side-effect of syllable structure repair. Furthermore, [kn] is not otherwise valid as either an onset or a coda sequence in Sylheti. More detailed studies are required into geminate behaviour under adjective and adverb formation.

Other than [zɔlfrɔfat], which I will return to in the next section, the remaining items are of the form [ŋgC], and mostly [ŋgL] (see Table 6). [ŋg] is not found as a word-final coda, and there are no minimal pairs contrasting [ŋg] and [ŋ]. The loan item ‘English’ is pronounced variously with and without the [g], and the Bangla cognates also lack it. A more detailed phonetic study of these items is required.

Table 6:  $\eta(g)C$  sequences

Sylheti	English	Bangla	
[hʊŋgi]	type of eel		
[tɛŋgra]	type of catfish	টেংরা	
[xɑŋla]	type of fish	ফলি	
[hamokbaŋgra]	'snail shell' stork	শামুকখোল	
[baŋgladeʃ]	Bangladesh		
[baŋgla]	Bangla	বাংলা	[baŋla]
[ɪŋgland]	England		
[leŋgra]	lame	লেংড়া	[leŋra]
[baŋglaɔr]	room		
[ɪŋgrezi]/[ɪŋgrez]	English		
[ɪŋliʃ]	English		
[fiŋla]	pink		
[siŋla]	bamboo switch		

## 4. Syllable initials

### 4.1 Vowel initial syllables

Like all contemporary Indo-Aryan languages (Masica 1991: 128), Sylheti has syllable-initial vowels (e.g. [afne] 'you'), and allows morpheme-internal vowel hiatus (e.g. [gaɪɔx] 'male singer').

A fuller analysis of vowel phonotactics in Sylheti requires a detailed examination of the status of diphthongs and their potential interactions with tone. As a first approximation, Sylheti allows diphthongs and short open syllables both word-internally and finally. The maximum number of morpheme-internal vowel qualities in a sequence is three, probably structured as one diphthong and one monophthong. By combining vowel-final verb stems with vowel-initial suffixes, Sylheti can have sequences of up to five vowels (three syllables), like Assamese. For example, [xava] 'to cough' may be inflected [amɪ xavaɪar] 'I'm coughing'.

### 4.2 Word-initial consonantal sequences

The onset consonants of Sylheti, in decreasing order of frequency, are the singletons [b], [f], [x], [s], [m], [ʃ], [h], [g], [d], [t], [z], [k], [n], [l], [r], [t], [d] (found in >1% of items) and the sequences [br], [pr], [fr], [kl], [st] and [gr]. There are other sequences, but each is found in only one lexical item, such as Hindi and Arabic greetings.

The infrequent sequences appear to represent borrowings or re-borrowings from English and Sanskrit (see Table 7). Almost all are nouns, the most frequently borrowed class of lexical items (Campbell 1993).

Table 7: A selection of onset clusters

Sylheti	English	Bangla		Sanskrit	
[brʃʌn]	Britain				
[bru]	brow	ভুরু	[b <sup>h</sup> uru]	भ्रू	<bhrū>
[brʃʈi]	rain	বৃষ্টি	[brʃʈi]	वृष्टि	<vr̥ṣṭi>
[klas]	class				
[klantə]	tired	ক্লান্ত	[klantə]	कलान्त	<klānta>
[gram]	village	গ্রাম	[gram]	ग्राम	<grāma>
[griʃo]	‘hot season’	গ্রীষ্ম	[griʃmə]	ग्रीष्म	<grīṣma>
[praʃno]	question	প্রশ্ন	[praʃno]	प्रश्न	<prazna>
[protizogita]	competition	প্রতিদ্বন্দ্বিতা	[prətɪdbəndbita]	प्रतियोगिता	<pratiyogitā>
[profesar]	professor				
[stiri]	wife	স্ত্রী	[stri]	स्त्री	<strī>
[stən]	breast	স্তন	[ʃtən]	स्तन	<stana>
[zɔʃrɔfat]	waterfall	জলপ্রপাত	[jalaprapāta]	प्रपात	<prapāta>

### 4.3 Word-medial onsets

Examining word-medial sequences using Nidaba, I find that all word-initial singletons also appear in word-internal onset position, as well as the previously mentioned retroflex allophones and geminates. Aside from [dɔkknɔʃ] and [brahmi], the only candidate for a word-internal complex onset in Sylheti is the bimorphemic item [zɔʃ-rɔfat], ‘waterfall’.

### 4.4 Syllable contact

Of the sequences of two word-medial consonants in the Sylheti lexicon, nearly 50% have falling sonority, 20% are identical consonants, 5% are non-geminates with level sonority, and 25% have rising sonority. Some of the rising sequences are loan items from languages with complex onsets (e.g. [madrasa] ‘religious school’), whereas others have been retained from Sanskrit. Whilst the Syllable Contact principle holds that sonority should drop across syllable boundaries, it is ‘often [overridden by] the prohibition of complex syllable onsets’ (Clements 2009). The incidence of word-medial rising sonority sequences in Sylheti therefore does not rule out a prohibition on complex onsets both initially and medially.

These potentially heterosyllabic sequences appear to assist sequential bilinguals in acquiring their tautosyllabic English counterparts, even though the English structures are not licit in Sylheti (McCarthy & Skoruppa 2015).

## 4.5 Repair strategies

### 4.5.1 Metathesis

A repair strategy which maximises retention of the original sounds is metathesis. Syllable structure requirements are met by transposing vowels and consonants, in this case to convert CCV.CV sequences to CVC.CV sequences.

I have not located any examples of this strategy being applied to English borrowings; metathesis may no longer be an active repair strategy in modern Sylheti.

*Table 8: Metathesis between Sanskrit (Old Indo-Aryan) and modern Sylheti*

क्रमि	<krimi>	→	[kiɾa]	worm
प्रति	<prati>	→	[forti]	every
प्रोष	<proṣa>	→	[fɔrsa]	light

### 4.5.2 Anaptyxis

Syllables with a pre-existing coda cannot be repaired by metathesis, given Sylheti's ban on complex codas. Instead, they are repaired with anaptyxis, the insertion of a vowel.

*Table 9: Anaptyxis*

[bɛɾɛd]	'bread'
[fɛlɛɪt]	'plate'
[dʒɛɾɛɪm]	'drain'
[tɛɾɛɪm]	'train'
[gɔllas]	'glass'

Singha & Ahmed (2016) record three different vowels used in epenthesis: [i], [e] and [o]. Given limited examples in both corpora, there is not yet conclusive evidence for whether vowel choice is determined by vowel harmony (a feature of Bangla and Assamese, e.g. Mahanta 2008) or by consonant quality. If the former, high vowels require [i], front vowels require [e], and back vowels require [o]. If the latter, [i] is used with [k], preventing its adaptation to [x], [e] is used following labials and retroflexes (front consonants), and [o] is used following velars (back consonants).



### 4.5.3 Prothesis

Loan words with an initial sT (s+stop) cluster are repaired through prothesis, the insertion of a vowel preceding the sequence. This holds for both sCVC(C) words, which cannot undergo metathesis, and for sCVCV(C) words, which could. This result is consistent with Goswami's (2013) findings for North Tripura Sylheti.

*Table 10: Prothesis*

[ɪspid]	'speed'
[ɪstɔf]	'stop'
[ɪstɛmf]	'stamp'
[ɪstɪʃɔn]	'station'

All examples of prothesis use [ɪ], regardless of the vowel quality of the following syllable, so again epenthetic vowel quality could be determined by the (here empty) onset. Alternatively, the intervention of a coda between the epenthetic vowel and the following one might also play a role in blocking harmony, as in Assamese (Mahanta 2008).

The treatment of sC clusters as coda+onset, with repair being through prothesis instead of anaptyxis, is cross-linguistically common (Goad 2012). The location of the boundary between the two strategies varies. For example, Hindi treats sT- and sm- clusters with prothesis, and sn-, sl- and s+sonorant sequences with anaptyxis. The single example of this in Sylheti is the repair via anaptyxis of [selet] 'slate'. It is not clear from this limited data if sn-sequences would be adapted with anaptyxis or with prothesis.

## 5. Types of intermediate relationship

Hall (2013) lists several types of intermediate relationship which can give rise to marginal contrasts. Sylheti displays several of them: varying predictability; distinct strata; production effects; frequency effects; and partial contrasts per natural class.

Firstly, Sylheti shows variation in predictability, and much of this article is dedicated to the exceptions to mostly predictable relationships. I have not dealt in as much detail with cases of neutralisation.

Secondly, Sylheti has distinct strata. In both syllable structure and consonantal inventory, there are broad generalisations which are violated only or almost only by words of foreign origin. I have not attempted a theoretical account of this variability, but rather tried to describe the native Sylheti stratum, approximately corresponding to the Japanese Yamato or Greek Demotic strata. The equivalent stratum in Bangla is known as *tadbhava* (Kar 2009). I have noted throughout where the gloss indicates Bangla, English or Sanskrit influence on Sylheti. The existence of re-borrowings from cognate languages complicates the situation, since assimilated cognates may be

indistinguishable from native lexical items, leaving a skewed picture of nativisation processes. Nagarajan (2014) demonstrates the effect of an evolving grammar on loanword adaptation in Bangla; Sylheti is in a very similar situation. We do not yet have enough data to examine other strata such as technical or fossilised vocabulary.

Variability in production would be a good area for future investigation, particularly those areas where a ‘foreign’ and a ‘native’ pronunciation have both been observed for the same word. The only variability noted is where differences are stark enough to have been transcribed with different phonemes, in a lexicon where the transcription is not maximally narrow. Given the aforementioned diglossic situation, it is difficult to conclude whether these observed variations should be considered part of the native Sylheti stratum or not at this stage. No detailed phonetic analysis is included here, a major drawback of relying primarily on a transcribed lexicon.

Given the lack of a written corpus of Sylheti, we currently have very little data on lexical frequency. However, I have noted the number of lexical items in which each pattern is found by Nidaba.

Finally, Sylheti displays a partial voicing contrast, with only some places and manners participating, and has a deficient retroflex series. A summary of the natural classes of Sylheti was included in Section 2.6.

## **6. Conclusion**

Nidaba has enabled me to perform a relatively rapid analysis of the phonological inventory and syllable structure of Camden Sylheti, starting from the SOAS Sylheti Project lexicon. It has helped me to locate allophonic patterns and minimal sets, and compare loanword adaption from English, Sanskrit and Bangla.

In terms of deviation from a basic CV syllable, I find that Sylheti permits both word-medial and word-final codas, but not complex codas or complex onsets, except in a minority of loan items. It permits word-initial vowels and vowel hiatus. Sylheti has rising sonority across some heterosyllabic CC sequences. Consonant clusters are found only in certain (foreign) strata of the language, and are subject to (variable) repair strategies. Given this analysis, studies of, e.g., bilingual acquisition or disruption of speech fluency, should not rely on Bangla data as necessarily being representative of Sylheti.

For a more complete and accurate analysis, three different types of additional data are required. Firstly, an in-depth acoustic study, looking at variation in different phonological and sociolinguistic contexts. Secondly, an articulatory study could shed light on, e.g., the syllabification of sT clusters and word-medial sequences (as in Hermes, Mücke & Grice 2013) and vowel harmony (as in Archangeli & Yip 2016). Thirdly, a comparative study is needed of Camden Sylheti with Bengali Sylheti, with Bangla, and with Indo-

Aryan in general, as Pattanayak (1966) did for Oriya, Assamese, Bangla and Hindi. Such a study would provide insight not just into the history and the synchronic state of the Sylheti language, but also into its complex sociolinguistic situation.

## References

- Archangeli, Diana & Jonathan Yip. 2016. The effect of context on the articulation of harmonic vowels in Assamese: An ultrasound study. *The Journal of the Acoustical Society of America* 140(4), 3105.
- Baker, Paul, Mabel Lie, Tony McEnery & Mark Sebba. 2000. The construction of a corpus of spoken Sylheti. *Literary and Linguistic Computing* 15(4), 421-432.
- Blevins, Juliette. 1995. The syllable in phonological theory. In John Goldsmith (ed.) *The Handbook of Phonological Theory*, 206-244. Oxford: Blackwell.
- Breen, Gavan & Rob Pensalfini. 1999. Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30(1), 1-25.
- Campbell, Lyle. 1993. On proposed universals of grammatical borrowing. In Henk Aertsen & Robert J. Jeffers (eds.) *Historical linguistics 1989*, 91-109. Amsterdam: John Benjamins.
- Cardona, George. 2017. Indo-Aryan languages (accessed 2017-05-08) <https://www.britannica.com/topic/Indo-Aryan-languages/Characteristics-of-the-modern-Indo-Aryan-languages>
- Clements, George N. 2009. Does sonority have a phonetic basis? In Eric Raimy & Charles E. Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 165-176. Cambridge, MA: MIT Press.
- Dasgupta, Probal. 2003. Bangla. In George Cardona & Dhanesh Jain (eds.) *The Indo-Aryan languages*, 351-390. London: Routledge.
- Goad, Heather. 2012. sC clusters are (almost always) coda-initial. *The Linguistic Review* 29(3), 335-373.
- Goswami, Arpita. 2013. Simplification of CC sequence of loan words in Sylheti Bangla. *Language in India* 13(6), 182-191.
- Gamkrelidze, Tamaz V. 1975. On the correlation of stops and fricatives in a phonological system. *Lingua* 35(3-4), 231-261.
- Glashoff, Klaus et al. 2017. Sanskrit dictionary for spoken Sanskrit. <http://spokensanskrit.org> (accessed 2016-06-01)
- Hall, Kathleen C. 2013. A typology of intermediate phonological relationships. *The Linguistic Review* 30(2), 215-275.
- Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of Italian word-initial clusters: The case of ‘impure s’. *Phonology* 30(1), 1-25.

- Kabak, Barış & William Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints. *Language and Speech* 50(1), 23-52.
- Kar, Somdev. 2009. The syllable structure of Bangla in optimality theory and its application to the analysis of verbal inflectional paradigms in distributed morphology. Ph.D. dissertation, Universität Tübingen.
- Khan, Sameer ud Dowla. 2010. Bengali (Bangladeshi standard). *Journal of the International Phonetic Association* 40(2), 221-225.
- Kessler, Brett. 1994. Sandhi and syllables in classical Sanskrit. In Erin Duncan, Donka Farkas & Philip Spaelti (eds.) *The proceedings of the twelfth west coast conference on formal linguistics [WCCFL]*, 35-50. Somerville, MA: Cascadilla Press.
- Mahanta, Shakuntala. 2008. *Directionality and locality in vowel harmony: With special reference to vowel harmony in Assamese*. Utrecht: LOT.
- Mahanta, Shakuntala. 2012. Assamese. *Journal of the International Phonetic Association* 42(2), 217-224.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- McCarthy, Kathleen M. & Katrin Skoruppa. 2015. The interplay between L1 and L2 phonotactics in sequential bilingual children. Poster session presented at: The 40th Annual Boston University Conference on Language Development. November 13-15. Boston: Boston University.
- Nagarajan, Hemalatha. 2014. Constraints through the ages: Loanwords in Bangla. *The EFL Journal* 5(1), 41-63.
- Pan, Ning & William Snyder. 2004. Acquisition of /s/-initial clusters: A parametric approach. In Alejna Brugos, Linnea Micciulla & Christine E. Smith (eds.) *Proceedings of the 28th annual Boston University conference on language development [BUCLD]*, 436-446. Somerville, MA: Cascadilla Press.
- Pattanayak, Debi P. 1966. *A controlled historical reconstruction of Oriya, Assamese, Bengali, and Hindi*. The Hague: Mouton.
- Pavlenko, Aneta. 2000. L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics* 11(2), 75-205.
- Rice, Keren. 2007. Markedness in phonology. In Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*, 79-98. Cambridge: Cambridge University Press.
- Śa', Rāmeśvara. 2001. *Synchronic comparative phonology of Bengali and German*. Kolkata: Pustak Bipani.
- Singha, Kh. Dhiren & Md. Ishaque Ahmed. 2016. Phonological adaptations of some English loanwords in Sylheti. *Language in India* 16(7), 269-286.