Non-allographed Consonant Graphemes and Diacritics in Bengali Words: Defining their Patterns of Usage and Nature of Pronunciation

Niladri Sekhar Dash[*]

Abstract

In this paper I have made an attempt to define the patterns of usage and pronunciation of Bengali consonants and diacritics when these characters are found to be used in formation of non-allographed words in the language. The basic goal of this study is to find out how consonants and diacritics are used in formation of non-allographed words and how these characters are pronounced within words in standard Bengali speech. Another goal of this study is to investigate how the original pronunciation of consonants is affected at certain contexts of their usages within the words. For this study I have utilized a lexical database of nearly three hundred thousand words tokenized from a text corpus of nearly five million words of modern Bengali written prose. This is, perhaps, the first study of its kind that can lead the Bengali speakers to understand how the pronunciation of non-allographed words can reflect on the standard Bengali speech habits. The application relevance of this study may be attested in several domains of mainstream and applied linguistics, such as, language description, language teaching, dictionary compilation, usage-based grammar writing, language processing, language cognition, speech recognition and synthesis, and text-to-speech conversion.

Keywords: Pronunciation, Allograph, Consonants, Graphemes, Diacritics, Bengali

1. Introduction

It is generally known that the patterns of usage and pronunciation of some of the Bengali consonant graphemes and diacritics are not always identical with their usage and pronunciation observed in those situations when they are treated as isolated characters in the language. Due to several linguistic and extra-linguistic factors (Hai 1985, Sarkar 1992, Ansari 1999, Pal 2001, Dash 2006), pronunciation of some consonant graphemes and diacritics is heavily affected. And this is reflected in pronunciation of those words where these consonant graphemes and diacritics are used.

[*] Linguistic Research Unit, Indian Statistical Institute, Kolkata, India

In this paper I have made an attempt to investigate the usage patterns as well as to identify those factors and contexts, which are solely responsible for causing variations in pronunciation of consonant graphemes and diacritics when these characters are used in formation of words where no vowel allograph (e.g., *ā-kār, i-kār, u-kār, e-kār, o-kār,* etc.) is attached. In a step-by-step process, I have studied the usage variations of consonant graphemes within Bengali words (Section 2); the unique pronunciation variations of some consonants when used in formation of words (Section 3); the pronunciation patterns of the consonant graphemes in case of non-allographed words made with two, three or more characters (Section 4); and the usage and pronunciation variations of the diacritic signs (Section 5). The applicational relevance of this work is highlighted in the concluding section of this paper.

2. Usage Variations of Consonant Graphemes

Bengali has small set of vowels, vowel allographs and consonant graphemes along with a large set of consonant grapheme clusters (or conjuncts) and a few diacritics to form orthographic words. When the consonant graphemes are used in formation of orthographic words in the language, we observe the following two important phenomena with regard to their form and occurrence:

(a)   Due to the phenomenon of *sequence redundancy* some consonants are restricted in use at some positions within words. That means, while many of the consonants are permitted to occur at word-initial, medial, and final positions, some are restricted in their occurrence in all these positions.

(b)   Some of the consonants undergo structural change (i.e.. change in their original forms) when these are used in formation of consonant clusters.

With regard to restrictions in positional usage and *sequence redundancy* (Schane 1972: 40). it is noted that all consonants are not permitted to occur at all the positions of words. For instance, the consonants like *ṅ, ñ, ṇ, ṛ, ṛh, ṭ (khaṇḍa-ta)* and *y* are not allowed to take place at initial position of a word, as these characters are not possible to pronounce at that position. On the contrary, the consonants like *ḍ* and *ḍh* are not allowed to occur at final position of a word, because of the same reason mentioned above. The most important thing is that the phenomenon of restricted usage does not apply to all the consonants used in Bengali. It is applicable to only a few consonants and these consonants are taken into consideration in this section. The information of restriction in use of consonants in word formation play crucial role at the time of teaching rules and strategies of word formation (Aronoff 1981: 21) in Bengali to the language learners. Similar observation stands pertinent in case of defining pronunciation of some consonants and this issue is addressed in the Section 3.

The number of consonants used in the Bengali script is 36 (thirty-six) if we include the *khanda-ta* (i.e., *haft-t*) in the list. However, most often this particular character is considered as a consonant graphic variant of the unvoiced dental stop consonant *t* rather than a unique consonant. Although this particular character (i.e., *khanda-ta*) shows some differences in occurrence and distribution in word formation as well as reveals notable difference in pronunciation with comparison to unvoiced dental stop consonant *t*, I have considered it as a graphic variant of *t* and treated it accordingly in the present frame of investigation, analysis and description. I have also adopted similar strategy for other consonant graphic variants and diacritics used in the script of the language.

In the following sub-sections I have tried to discuss about those consonants that have exhibited variations and have exercised restrictions in usage. Their patterns of usage are highlighted in the following sub-sections with examples taken from the Bengali written corpus. I have also tried to identify if there is any factor that controls their usages.

2.1 Velar nasal consonant grapheme ṅ

The velar nasal consonant ṅ never occurs at the initial position of a Bengali word, because it is impossible to pronounce a word with this character at the very first position. Therefore, it is found to occur either at word- medial or at word-final, as shown below.

(a)   At word-medial position

kāṅgāl	"beggar"	noṅgar	"anchor"
lāṅgal	"plough"	hāṅgar	"shark", etc.

(b)   At word-final position

gāṅ	"river"	coṅ	"flannel"
jaṅ	"war"	byāṅ	"frog"
bhāṅ	"bhang"	raṅ	"colour"
saṅ	"buffoon"	hiṅ	"asafoetida", etc.

This consonant, however, has registered its highest percentage of usage as the 1st member of a consonant clusters (e.g., *ṅk, ṅg, ṅgh, ṅm,* etc.), as the following examples show.

(c)   As first member of a consonant cluster

aṅka	"lap"	cuṅgi	"octroi"
jaṅgal	"forest"	paṅkil	"muddy"
phiṅge	"drongo"	baṅga	"Bengal"
bhaṅgi	"design"	maṅgal	"good"
raṅgin	"colourful"	saṅgit	"song"
saṅghāt	"clash"	jaṅghā	"thigh"

bāṅmay   "eloquent"   diṅmaṇḍal   "horizon". etc.

However, it should be stated here that these clusters where ṅ is used as the first member, never occur at the word-initial position due to pronunciation limitation. Therefore, these clusters usually occur at word-medial or word-final positions in the language.

2.2 Palatal nasal consonant grapheme ñ

The palatal nasal consonant ñ occurs neither at the initial nor at the medial position of a Bengali word. Also, it does not have much use at the final position of a word as this consonant is very much restricted in use in the language. I have found only a few words in the corpus database where it has occurred at the final position of a word, as in, miñā "mister", yācñā "want". etc.

This consonant grapheme, however, is most often used either as the first or the last member in formation of clusters (e.g., ñc, ñch, jñ, etc.) as the following examples show.

(a)   As first member of a consonant cluster

añcal	"area"	bāñchā	"desire"
mañjari	"bud"	jhañjhā	"storm"
pañcam	"five"	cañcal	"frolic". etc.,

(b)   As last member of a consonant cluster

jñān	"knowledge"	bijñān	"science"
ajñān	"senseless"	abhijñān	"proof", etc.

2.3 Voiced retroflex consonant grapheme ḍ

The voiced retroflex consonant ḍ is quite frequently used at the initial position of a word, but hardly found to take place at the medial and final position of a word, as the following examples show.

(a)   Word-initial position

ḍak	"dock"	ḍan	"don"
ḍar	"fear"	ḍabal	"double"
ḍāl	"branch"	ḍāb	"green coconut"
ḍāk	"call"	ḍim	"egg"
ḍāini	"witch"	ḍākāt	"robber"
ḍāgar	"wide"	ḍānā	"wing"
ḍali	"tray"	ḍiṅgi	"boat"
ḍiś	"dish"	ḍubanta	"drowning"
ḍumur	"fig"	ḍek	"deck"
ḍerā	"shelter"	ḍobā	"pond"
ḍor	"string"	ḍorā	"stripe"

ḍaul   "shape" ḍela   "lump", etc.

At the medial position of a word it is found only in case of a few compound words and in case of some English words like the followings.

(b)   Word Medial position

suḍaul   "well formed"   ābḍāl   "off-side"
magḍāl "high branch"   raṇaḍaṅkā   "sound of war"
maḍel   "model"   pyāḍel   "paddle", etc.

At word-final position this consonant grapheme is found to occur in case of some English words borrowed into Bengali such as the followings.

(c)   Word-final position
maḍ   "mod"   huḍ   "hood"
heḍ   "head"   muḍ   "mood"
raḍ   "rod"   roḍ   "road", etc.

A pertinent observation may be raised in this context with regard to usage of this consonant at the word-final position. It is noted that it does not occur in this position in case of Sanskrit and native Bengali words but it is frequently used in case of English words. It is surprising that there is not a single Sanskrit or Bengali word where this consonant character is used at the final position. One may argue that it is not found at that position (i.e., word-final) in Sanskrit and Bengali because it is difficult to pronounce the consonant properly at this position of a word. But this argument cannot be true, because many English words, where the last character is a voiced retroflex consonant, are borrowed into Bengali and these words are rightly pronounced by the native Bengali people quite easily without any problem. In fact, the borrowing of such English words into Bengali has expanded the range of use of this particular consonant sound to the final position of words for native Bengali speakers. Therefore, it is still an unsolved question why this character is not found to be used at word-initial position in case of Sanskrit and Bengali words.

In most cases this consonant is used in formation of consonant clusters (e.g., *ḍḍ, ṇḍ, nḍ, ḍy,* etc.) – either as a first or as a second member. And these clusters are mostly used either at the medial or final position of words, but never at the initial position, as the following examples show.

(d)   As a member of a consonant cluster

āḍḍā   "gossip"   gāḍḍā   "ditch"
hāḍḍi   "bone"   maṇḍa   "lump"
gaṇḍa   "cheek"   jāḍya   "pace"
pāṣaṇḍa   "devil"   ṣaṇḍa   "bull"

pyāṇḍel "pandal"		bāṇḍil	"bundle"
kanḍom "condom"		anḍa	"egg", etc.

2.4 Voiced aspirated retroflex consonant grapheme ḍh

The voiced aspirated retroflex consonant *ḍh* has a very limited usage in the language. It mostly occurs at the initial position of a word. At word-medial position, it has two or three occurrences, as in, *aḍhel* "many", *āḍhkā* "not covered", *beḍhap* "disproportionate", etc. At word final position, it hardly occurs.

In case for cluster formation, this consonant is found only in two or three clusters, such as, *meḍhra* "ram", *barṇāḍhya* "colourful", *dhanāḍhya* "rich and affluent", etc.

2.5 Retroflex nasal consonant grapheme ṇ

The retroflex nasal consonant *ṇ* is highly restricted in its use in the language. It is rarely used at initial position of words, although it may occur quite frequently at medial and final position of words. There are only two or three words in Bengali where this particular consonant is found to take place at word-initial position, e.g., *ṇijanta* "a pattern of lexical derivation", *ṇatvabidhān* "rules for using ṇ", etc.

2.6 Voiced flap consonant grapheme ṛ

Due to pronunciation limitations and inabilities, the voiced flap consonant *ṛ* registers no use at the initial position of Bengali words. Therefore, it usually occurs at word-medial or at word-final position, as the following examples exhibit.

taṛbaṛ	"hurried"	kharkhaṛi	"window pane"
narbaṛe	"loose"	jaṛatā	"sluggishness"
āṛ	"side"	gaṛ	"fort"
paṛ	"(you) read"	caṛ	"slap"
jaṛa	"lifeless"	jhaṛ	"storm"
hāṛ	"bone"	muṛo	"head"
pāṛ	"shore"	sāṛā	"response"
sāṛi	"saree",	gāṛi	"car", etc.

2.7 Voiced aspirated flap consonant grapheme ṛh

The voiced aspirated flap consonant *ṛh* does not take place at word-initial position for the same reason applicable to *ṛ* stated above. Therefore, this consonant is normally used either at word-medial or at word-final position in Bengali, such as the followings.

dṛṛha	"strong"	dṛṛhatā	"strength"
mūṛha	"idiot"	mūṛhatā	"foolishness"
gāṛha	"thick"	prauṛha	"aged"
pragāṛha	"profound"	āsāṛh	"a rainy month", etc.

2.8 Consonant graphic variant grapheme t̪

The usage of t̪ (i.e., *khaṇḍa-ta*) is also highly restricted in words in Bengali. It never occurs at the word-initial position. Usually, it occurs at word-medial or at word-final position, such as the followings.

utpāt	"nuisance"	utsab	"festival"
satkār	"death ritual"	satpātra	"suitable groom"
hrtpiṇda	"heart"	mutsuddi	"merchant"
butpatti	"origin"	madat	"help"
parṣat	"society"	hathāt	"suddenly"
nacet	"else"	mahat	"great"
bhabiṣyat	"future", etc.		

This character has some uniqueness with regard to the use of vowel allograph with it. It is never allowed to carry any vowel allograph with it. Therefore, whenever a situation arises for using a vowel allograph with this character, particularly at the time of using case markers and suffixes, it immediately changes into its source consonant (i.e., unvoiced dental stop consonant *t*) as the examples show.

mahat	"great"	but	mahater	"of great"
taphāt	"distance"	but	taphāte	"at distance"
bhabiṣyat	"future"	but	bhabiṣyate	"in future"

However, there are some exceptions to the rule reported above. Sometimes the case markers and suffixes are attached with this character by using a hyphen mark immediately after it keeping the character unchanged in its original shape, as shown below.

mahat-er	"of great"	bhabiṣyat-er	"of future"
parṣat-ke	"to the society"	hathāt-i	"by chance",
nehāt-i	"only just", etc.		

In my observation hyphenation is a useful method of preserving the original shape of the character in Bengali words. However, this leads towards increase of number of spelling variation of some words in the language (Dash 2006: 28). For instance, the word *bhabiṣyat-er* "of future" can have now at least two different spellings: *bhabiṣyater* (with regular *t* and no hyphen) and *bhabiṣyat-er* (with *khaṇḍa-ta* and a hyphen). Both the spellings are quite frequently used in the language, as the corpus under my study shows.

2.9 Semivowel consonant grapheme y

The semivowel consonant *y* does not usually occur at the initial position of the Bengali words. However, there are some words, which are written with this consonant at this position, such as, *yunivārsiṭi* "university", *yurop* "Europe",

yunian "union" etc. This was, however, an old practice of writing such words in Bengali and it is almost obsolete now in the language.

The semivowel consonant has, however, recorded its highest frequency of use both at word-medial and word-final positions, as the following examples show.

āynā	"mirror"	āyā	"ayah"
nāyak	"hero"	bayas	"age"
maydā	"flour"	baniyād	"base"
gāyak	"singer"	cayan	"selection"
kāyā	"body"	jāyā	"wife"
samay	"time"	hāy	"alas"
nay	"not"	jay	"victory"
balay	"circle"	nayan	"eye"
hay	"becomes"	gharoyā	"homely", etc.

The remaining consonants used in the Bengali script have no restriction in their use with regard to position in the words. That means these can occur at all three positions (i.e., initial, medial, and final) of words without any restriction although their frequency of usage at different positions may vary based on their nature of occurrence and their role in formation of words in the language (Hossain 1989, Sarkar 1992, Bhattacharya 2005).

3. Pronunciation Variations of Consonant Graphemes

Since the total number of consonants (36) used in the Bengali script is more than the number of consonant sounds (30) they are able to denote, it is natural that some consonants will be similar in pronunciation (Sarkar 1992). A simple analysis shows that some of the consonants are almost similar in pronunciation, although these are different in their orthographic forms and representation. For example, the voiced palatal affricate *j* and semivowel *y* are pronounced as [ɟ], although they are two different consonant characters with different orthographic forms. Similarly, the palatal sibilant consonant *ś* and the retroflex sibilant consonant *ṣ* are pronounced as [ʃ] in spite of their different shape and design. The retroflex nasal consonant *ṇ* and the dental nasal consonant *n* are pronounced as [n]. Finally, the unvoiced dental stop consonant *t* and its graphic variant *ṭ* *(khaṇḍa-ta)* are pronounced as [t]. Other consonants of the language are more or less unique in their orthographic forms and in denotation of discrete sounds.

All Bengali consonants, by default, are vocalic (i.e., ɔ-ending) in isolation but non-vocalic (i.e., consonant-ending) when used with vowel allographs. Due to this reason, all Bengali consonants are *diaphonic* in nature. That means they have double phonetic identities. In isolated situation, each consonant carries an inherent [ɔ] sound, which is often manifested in pronunciation. In other situation, this inherent [ɔ] sound of the consonant becomes silent. For example, the unvoiced velar stop consonant *k*, in isolation, is pronounced as [kɔ] ([k+ɔ]). When used within words it can be pronounced either as [kɔ] or as simple [k].

Within words it can occur either with or without a vowel allograph. When it is without a vowel allograph, it can be either pronounced with inherent [ɔ], as in, *kam* [kɔm] "less" or without it, as in, *bak* [bɔk] "crane", etc.

On the other hand, when this consonant is attached with as a vowel allograph, it invariably drops its inherent [ɔ] sound to become *isophonic* (i.e., [kɔ] > [k] / [ɔ] > ø). Thus, the process of change of a diaphonic consonant into an isophonic one in Bengali can be explained in the following four stages:

[1]   Phonemic form of the consonant : k=[kɔ] (diaphonic)
[2]   The vowel allograph *i-kār* [i] is tagged to it: [kɔ +i]
[3]   Rule of transformation: [ɔ] > ø and [i] inserted
[4]   Final form of the consonant: ki = [ki] (output)

These conversion rules stand valid for almost all the consonants (and clusters) used in the Bengali script. However, there are some exceptions, and these are discussed in the following sub-sections.

3.1 Velar nasal consonant grapheme ṅ

The velar nasal consonant *ṅ* has no pronunciation information at the word initial position because it does not occur here. At the word-medial position, it is always open or non-vocalic in pronunciation. That means at this position it is pronounced without the support of the inherent [ɔ] sound of the grapheme, as shown below.

aṅka	[ɔŋko] "lap"	kāṅgāl	[kaŋgal] "beggar"
baṅga	[bɔŋgo] "Bengal"	cuṅgi [cuŋgi] "octroi"	
jaṅgal	[ɉɔŋgol] "forest"	phiṅge [phiŋge] "drongo"	
bhaṅgi	[bɦoŋgi] "design"	maṅgal [mɔŋgol] "good"	
saṅgit	[ʃoŋgit] "song" rangin [roŋgin] "colourful"		
noṅgar [noŋgor] "anchor"		lāṅgal	[laŋgol] "plough"
hāṅgar [ɦaŋgor] "shark"		hāṅgāmā [ɦaŋgama] "chaos".	

At the word-final position also this consonant (ṅ) is pronounced in the same manner (i.e., as a non-vocalic consonant character without the assistance of the inherent [ɔ] sound of the grapheme) as the following examples exhibit.

gāṅ [gaŋ] "river"	jaṅ [ɉɔŋ] "war"
ṭaṅ [ṭɔŋ] "top"	bhāṅ [bɦaŋ] "bhang"
byāṅ [bæŋ] "frog"	raṅ [rɔŋ] "colour"
saṅ [ʃɔŋ] "buffoon"	hiṅ [ɦiŋ] "asafoetida", etc.

3.2 Voiced flap consonant grapheme ṛ

Since the voiced flap consonant *ṛ* does not occur at word-initial position, it has no record of pronunciation at this position. In word-medial and word-final positions, this consonant – when not attached with any vowel allograph – is often non-vocalic in pronunciation as the following examples show.

ār [aɽ] "side"	gar [gɔɽ] "fort"
gur [guɽ] "molasses"	car [cɔɽ] "slap"
chār [chaɽ] "rebate"	par [pɔɽ] "read"
pār [paɽ] "shore"	jhar [ɟɦɔɽ] "storm"
hār [ɦaɽ] "bone"	khar [khɔɽ] "straw"
dhar [dhɔɽ] "body"	mīr [miɽ] "scale"
nīr [niɽ] "nest"	sār [ʃaɽ] "response", etc.

However, there are some exceptions to the pattern mentioned above. It is noted that in some words this character is vowel-ended in standard pronunciation, as in, *jara* [ɟɔɽo] "lifeless", *dara* [dɔɽo] "strong", *saragara* [ʃɔɽogɔɽo] "accustomed", *bara* [bɔɽo] "large", etc. The number of such examples, however, is quite low in the language.

3.3 Voiced aspirated flap consonant grapheme ṛh

The voiced aspirated flap consonant *ṛh* has no record of pronunciation at word-initial position because it does not occur here. At word-medial and word-final position, it is always vocalic or vowel-ended in pronunciation in standard Bengali speech, as shown below.

dṛṛha [driɽɦo] "strong"	dṛrhatā [driɽɦota] "strength"
mūrha [muɽɦo] "idiot"	mūrhatā [muɽɦota] "stupidity"
gārha [gaɽɦo] "thick"	pragārha [progaɽɦo] "profound"
praurha [prouɽɦo] "aged", etc.	

Only one exception is found to this rule where this consonant is non-vocalic in pronunciation at word-final position, e.g., āṣāṛh [aʃaɽ] "name of rainy month in Bengali" etc.

3.4 Consonant graphic variant grapheme ṯ

The graphic variant *ṯ* (*khaṇḍa-ta*), irrespective to its use at any position within words, is always non-vocalic in pronunciation. And without exception, the pronunciation of the character is exactly the same to that of unvoiced dental consonant **t**, as the following examples show.

ālbāṯ [albat] "surely"	utpāṯ [utpat] "nuisance"
uṯsab [utsɔb] "festival"	kupokāṯ [kupokat] "pulled flat"
jagaṯ [ɟɔgɔt] "world"	taphāṯ [tɔphat] "difference"
daibāṯ [doibat] "by chance"	naceṯ [nɔcet] "else"
nehāṯ [neɦat] "only"	parṣaṯ [pɔrʃɔt] "society"
bajjāṯ [bɔʲɟat] "rascal"	bidyuṯ [biᵈdut]"thunder"
bhabiṣyaṯ [bɦobiʲʃɔt] "future"	buṯpatti[butpoˈti]"origin"
madaṯ [mɔdɔt] "help"	mahaṯ [mɔhot] "great"
muṯsuddi [mutsuᵈdi "seller"	saṯ[ʃɔt] "honest"
haṯhāṯ [ɦɔthat] "suddenly"	hṛṯ [ɦrit "heart"

saṭkār [ʃɔtkar] "performing death rituals"　　　etc.

3.5 Aspirated glottal fricative grapheme h

Among the total consonants available in the language, the aspirated glottal fricative *h* is always vocalic in pronunciation irrespective to its occurrence at any position of a word. That means it is always pronounced with [ɔ] sound wherever it is used within a word. However, in some situations, due to vowel harmony or similar other phonological process, the inherent [ɔ] sound is changed into [o] sound, as the following examples show.

āhata [afiɔto] "wounded"　　ābaha [abɔfio] "wounded"
nihata [nifiɔto] "dead"　　jahar [ɟɔfiɔr] "poison"
bahatā [bɔfiɔta] "flowing"　　gahanā [gɔfiɔna] "ornament"
deha [defio] "body"　　mahat [mɔfiɔt] "great"
dehalī [defioli] "house"　　prahar [profiɔr] "time"
bibāha [bibafio] "marriage"　　pratyaha [proˈtofio] "daily"
bigraha [biᵍgrofio] "idol"　　āgraha [aᵍgrofio] "eagerness"
saha [ʃɔfio] "with"　　grahaṇ [grofiɔn] "accept", etc.

Some exceptions to this rule are observed in case of some Arabic and Persian words, where this consonant is partially pronounced with [ɦ] sound, e.g., *āllāh* [aˈlaʰ] "god", *bādśāh* [badʃaʰ] "king", *dargāh* [dɔrgaʰ] "mosque", *darmāh* [dɔrmaʰ] "daily wage", etc.

4. Consonant Graphemes in Non-allographed Words

Because of the diaphonic nature of Bengali consonants it becomes necessary to look into the pronunciation patterns of the characters within those words, which are formed with consonants (and clusters) only without the use of any vowel allograph. Since this particular issue has never been addressed before with proper empirical evidences, I have tried to probe details into the nature of variation of pronunciation of the consonants when these graphemes take part in formation of words.

The study becomes quite intriguing in the sense that when the words are made with consonants only, each consonant is potentially eligible to exhibit one of the three following pronunciation variations based on form, meaning, and etymology of the words:

[1]　A consonant can be pronounced with [ɔ] sound,
[2]　A consonant can be pronounced with [o] sound, and
[3]　A consonant can be non-vocalic [ø] in pronunciation

Therefore, defining the patterns of pronunciation of Bengali words made with non-allographed consonants becomes a challenging task. However, this study

becomes a valuable contribution for people involved in the works relating to language teaching, pronunciation dictionary compilation, text-to-speech conversion, and language cognition as understanding the patterns of pronunciation of the words provide necessary insights to design syllabus for teaching pronunciation of words to learners; provide proper information of pronunciation of words in dictionary; formulate necessary rules to be used in machine learning for capturing accurate pronunciation of consonants; and train a system with necessary inputs for proper recognition of speech units within the larger domain of machine recognition of a language and its speech units.

From a lexical database of three hundred thousand Bengali words I have found nearly 5,000 words, which are formed with consonants (and clusters) only without any vowel allograph. These words are actually formed by way of using consonants (and clusters) which are mostly diaphonic in nature, e.g., *ghånå* "thick", *såråbå* "vocal", *måtå* "like", *kåmålå* "lotus", *påråbå* "festival", *kåtå* "how much", *dålå* "party", *jånåpådå* "locality",[3] etc. This kind of word formation is quite possible in Bengali because all non-allographed consonants (and clusters), due to their diaphonic entity, are often vocalic in pronunciation with [ɔ] or [o].

Taking this factor into consideration I formulate the following observations about the general patterns of pronunciation of non-allographed Bengali words:

(a) Each consonant used in a non-allographed word can have three pronunciation variations: [ɔ] or [o] or [ø].
(b) The first consonant of the word, by default, is pronounced with either [ɔ] or [o] sound. It can never be non-vocalic.
(c) The pronunciation of the subsequent consonants varies among three alternatives: [ɔ], [o] or [ø].

For illustration consider a list of Bengali words, which are formed with two consonants only (without any vowel allograph): *kårå, chålå, bårå, jårå, målå, khålå, nåbå, såbå, råtå, rånå, bånå, bålå, cålå, gåmå, pånå, hålå, tåbå, mågå, dålå, pålå, dåśå, sårå, håtå,* etc. For these words, the first consonant is mostly pronounced with [ɔ] or with [o] sound, while the second consonant is pronounced with [ɔ], [o] or [ø] sound. Thus, in a simple estimation, we can have 6 pronunciation variations for a single word made with 2 non-allographed consonants:

(a)   1st consonant with [ɔ] and 2nd consonant with [ɔ]
(b)   1st consonant with [ɔ] and 2nd consonant with [o]
(c)   1st consonant with [ɔ] and 2nd consonant with [ø]
(d)   1st consonant with [o] and 2nd consonant with [ɔ]
(e)   1st consonant with [o] and 2nd consonant with [o]
(f)   1st consonant with [o] and 2nd consonant with [ø]

It is, therefore, very difficult for the language users as well as for the language learners to determine the actual pronunciation of these words, since it is not known, at the time of pronunciation, which consonant is non-vocalic and which consonant is vocalic in which manner in words. We require necessary etymological and semantic information of words, beside other types of information, to determine their actual pronunciation.

Similarly, there are many Bengali words, which are formed with 3 (three) non-allographed consonants, e.g., *kåtåkå, kådåmå, kåbåcå, gåthånå, gåtårå, gålådå, câmåkå, cåpålå, cåråmå, jåthårå, jånåmå, jåmåjå, någådå, någårå, pålåkå, påsåmå, bådånå, bådålå, bålådå, måtånå, mårådå, såtåkå, såråmå, såråņå, såcålå, såbålå, såmåyå, såtåtå, såmårå, håjåmå, hålåphå, håråņå,* etc. More complications arise in pronunciation of these words as we find several options open with regard to pronunciation of the individual consonants. While the first consonant is pronounced with [ɔ] or [o] sound, second, and third consonant is pronounced either with [ɔ], [o] or [ø] sound. Thus we can have 18 possible pronunciation variations for a single word made with 3 non-allographed consonants[4]. Since it is very difficult to determine the pronunciation of such words, we require to have detailed information of various types from the words as well as from the language to determine their actual pronunciation.

It is understood that the pronunciation of a word made with two or three non-allographed consonants may vary depending on its meaning, part-of-speech, or etymology besides other factors embedded with the text or the language. To justify this observation I present below a sample list of words made with two and three non-allographed consonant to show how pronunciation of words varies due to factors stated above (Table 1).

Word	POS	Meaning	Pronunciation
måtå	Noun	opinion	[mɔt]
måtå	Adverb	like	[mɔto]
hålå	Noun	hall	[ɦɔl]
hålå	Finite verb	became	[ɦolo]
kårå	Noun	hand	[kɔr]
kårå	Finite verb	you do	[kɔro]
pårå	Pronoun	other	[pɔr]
pårå	Finite verb	you wear	[pɔro]
kåmålå	Noun	lotus	[kɔmol]
kåmålå	Finite Verb	decreased	[komlo]
påråbå	Noun	festival	[pɔrob]
påråbå	Finite Verb	will wear	[porbo]
påråtå	Noun	layer	[pɔrot]
påråtå	Finite verb	used to wear	[porto]
såråbå	Adjective	loud	[ʃɔrob]
såråbå	Finite verb	will move	[ʃorbo]

Table 1: Pronunciation variation of words due to part-of-speech and meaning

The data given in the list above (Table 1) shows that determining the pronunciation of non-allographed words is a tough problem in Bengali as variations are quite high in case of such words[5]. However, based on the analysis of the words I have formulated the following two rules which are more or less true to all the words made with two, three or more non-allographed consonants (and clusters).

Rule 1: At word-initial position. a non-allographed consonant (or a cluster) is vocalic with [ɔ] or [o] sound.

Rule 2: At word-final position, a non-allographed consonant is usually non-vocalic, while a cluster is usually vocalic in pronunciation.

Even if we do not consider the information regarding part-of-speech and meaning of words in determining their pronunciation, we can postulate some general patterns of pronunciation, which appear to be true to a large number of words made with non-allographed consonants in the language. These general patterns are illustrated below with examples taken from the Bengali corpus

Pattern 1: The pronunciation patterns of the words made with two non-allographed consonants are [ɔ : ø] or [ɔ : o]. For validation let us consider the examples given below with their pronunciation in International Phonetic Alphabet (IPA).

kårå [kɔr] "do"   kårå [kɔro] "do"
cålå [cɔl] "go"   cålå [cɔlo] "go"
bålå [bɔl] "ball"   bålå [bɔlo] "say"
chålå [chɔl] "feign"   gåmå [gɔm] "wheat"
påṇå [pɔn] "dowry"   hålå [ɦɔl] "hall"
hålå [ɦɔlo] "happened"   tåbå [tɔbo] "your"
någå [nɔg] "hill"   målå [mɔl] "mall"
pålå [pɔl] "moment"   dåså [dɔʃ] "ten"
sårå [ʃɔr] "move"   håtå [ɦɔto] "killed"
ghånå [gɦɔno] "thick"   råbå [rɔb] "sound"
måtå [mɔt] "opinion"   måtå [mɔto] "like"
kåmå [kɔm] "less"   pårå [pɔr] "other"
pårå [pɔro] "you wear"   kåtå [kɔto] "how much"
dålå [dɔl] "party", etc.

Pattern 2: The pronunciation patterns of words made with three non-allographed consonants is generally [ɔ : ɔ : ø] or [ɔ : ɔ : o] as the following examples illustrate.

kåtåkå [kɔtɔk] "some"   kådåmå [kɔdɔm] "step"

kåbåcå [kɔbɔc] "armour'
khåtåmå [khotɔm] "finish"
gåmåkå [gɔmɔk] "step'
cåmåkå [cɔmɔk] "surprise'
cåråṇå [cɔrɔn] "leg'
jvålånå [ɟɔlɔn] "burning'
ṭåhålå [ʈɔɦɔl] "stroll'
ḍåbålå [ɖɔbɔl] "double'
tåkhånå [tɔkhɔn] "then'
dålånå [dɔlɔn] "pressing'
dhåråṇå [dɦɔrɔn] "type'
nåråmå [nɔrɔm] "soft'
någårå [nɔgɔr] "city'
påråbå [pɔrɔb] "festival'
bådålå [bɔdɔl] "change'
båhårå [bɔɦɔr] "width'
mågåjå [mɔgɔɟ] "brain'
måråmå [mɔrɔm] "heart'
måtånå [mɔtɔn] "like'
såråmå [ʃɔrɔm] "shame'
såhårå [ʃɔɦɔr] "city'
såcålå [ʃɔcɔl] "working'
såtåtå [ʃɔtɔto] "always'
håjåmå [ɦɔɟɔm] "digest'

kåmålå [kɔmɔl] "lotus"
gålådå [gɔlɔd] "error"
gåmånå [gɔmɔn] "going"
cåpålå [cɔpɔl] "restless"
jålådå [ɟɔlɔd] "cloud"
tågårå [tɔgɔr] "kind of flower"
ṭhåmåkå [ʈhɔmɔk] "gait"
tålåbå [tɔlɔb] "summon"
dåmåkå [dɔmɔk] "guts"
dhåmåkå [dɦɔmɔk] "threat"
nådhårå [nɔdɦɔr] "succulent"
nåjårå [nɔɟɔr] "view"
pålåkå [pɔlɔk] "eyelid"
påråmå [pɔrɔm] "ultimate"
bålådå [bɔlɔd] "castrated bull"
bhåbånå [bɦɔbɔn] "house"
månånå [mɔnɔn] "think"
måråṇå [mɔrɔn] "death"
mårådå [mɔrɔd] "male"
såråṇå [ʃɔrɔn] "shelter"
såråbå [ʃɔrɔb] "vocal"
såhåjå [ʃɔɦɔɟ] "easy"
såphårå [ʃɔphɔr] "journey"
hålåphå [ɦɔlɔph] "oath", etc.

Pattern 3: The pronunciation patterns of words made with 4 consecutive non-allographed consonants are of 3 types: [ɔ : o : ɔ : ø], [ɔ : o : ɔ : o], and [ɔ : ø : ɔ : ø], as the following examples exhibit.

Type-I: [ɔ : o : ɔ : ø]

In Type-I, the pronunciation of the words mostly becomes bimoraic and trisyllabic ($CV_1CV_2CV_3$) where the first two syllables are open syllables and the last one is a closed syllable, as the following examples show.

apåcåyå	[ɔpocɔĕ]	"waste"
jånågåṇå	[ɟɔnogɔn]	"mass"
kålåråbå	[kɔlorɔb]	"noise"
jånåpådå	[ɟɔnopɔd]	"locality"
kåråtålå	[kɔrotɔl]	"palm"
jålådhårå	[ɟɔlodɦɔr]	"cloud"
såsådhårå	[ʃɔʃodɦɔr]	"moon"
hålådhårå	[ɦɔlodɦɔr]	"one who ploughs"
såmåtålå	[ʃɔmotɔl]	"plane land"
bånåcårå	[bɔnocɔr]	"living or moving in forest"
påråmåtå	[pɔromɔt]	"other's view"

jalāmaya   [jɔlomɔĕ] "flooded with water", etc.

Type-II: [ɔ : o : ɔ : o]

In Type-II, the words are pronounced in two moras and in four syllables ($CV_1CV_2CV_3CV_4$) where each syllable is an open one, as the following examples exhibit.

apāhātā   [ɔpofiɔto] "prohibited"
abāgātā   [ɔbogɔto] "aware"
abānātā   [ɔbonɔto] "lowered"
gālāgāṇḍā   [gɔlogɔnɖo] "goitre"
sārāgārā   [ʃɔɽogɔɽo] "well-versed"
tādābāsthā   [tɔdobɔstho] "as in that situation"
pādārājā   [pɔdorɔjo] "dust from legs"
annāchātrā   [ɔnnochɔtro] "place for free food"
gāṇātāntrā   [gɔnotɔntro] "democracy"
kāhātābyā   [kɔfiotɔʰbo] "something worth saying"
phālāprādā   [phɔloprɔdo] "fruitful", etc.

Type-III: [ɔ : ø : ɔ : ø]

In Type-III, the words are normally pronounced in two syllables ($CVC_1 CVC_2$) by putting a moraic break after the first closed syllable, as the following examples display.

tālāmālā   [tɔlmɔl] "stumbling"
khātāmātā   [khɔʈmɔʈ] "difficult"
jhālāmālā   [ɟfiɔlmɔl] "glittering"
cāṭāpāṭā   [cɔʈpɔʈ] "quick"
chālāchālā   [chɔlchɔl] "tearful"
jhānājhānā   [ɟfiɔnɟfiɔn] "jingling"
tārābārā   [tɔɽbɔɽ] "to be hurried"
bānābānā   [bɔnbɔn] "rotating in high speed" [6] etc.

Pattern 4: The patterns of pronunciation of those words which are formed with five or more non-allographed consonants is mostly inconsistent as these words are usually formed as compounds by way of combining two or more words. In that case, the pronunciation may vary from word to word based on various linguistic factors such as meaning, part-of-speech, etc., as the following examples show.

anābārātā   [ɔnobɔrɔto] "continuously"
dālābādālā   [dɔlbodɔl] "changing a party"
kārāmārdānā   [kɔromɔrdɔn] "hand-shake"
gāṭhānātāntrā   [gɔʈhɔntɔntro] "system of formation"
jānānātāntrā   [ɟɔnɔntɔntro] "system of reproduction"
māṅgālāmāyā   [mɔŋgɔlmɔĕ] "full of blessings"

kåråtålågåtå	[kɔrotɔlgɔto]	"under total control"
kåråkåmålå	[kɔrokɔmɔl]	"hands as lotus"
gåjågåmånå	[gɔjɔgɔmɔn]	"walking like an elephant"
cårånåkåmålå	[cɔrɔnkɔmɔl]	"legs like lotus"
cårånåtålå	[cɔrɔntɔl]	"put below one's feet"
jånåpråcålånå	[jɔnoprocɔlɔn]	"used by people"
jåbårådåkhålå	[jɔbɔrdokhɔl]	"occupied forcibly"
såṅkhåbålåyå	[ʃɔŋkhobɔlɔĕ]	"bangle of conch-shell"
mådhyåpråhårå	[moᵈdɦioprofiɔr]	"mid-day or mid-night"
såmåmånåskå	[ʃɔmomɔnɔʃko]	"of same mind"
håtåsåmpådå	[ɦɔtosɔmpɔd]	"stolen wealth"
såbåsåkåtå	[ʃɔbʃɔkɔt]	"cart that carries dead body"
såbdåcåyånå	[ʃɔbdocɔĕɔn]	"word selection"
såmåtålåsthå	[ʃɔmɔtɔlɔstho]	"placed on plane land"

Since there are irregularities in the patterns of use of the syllables within these words, it is not possible to define any regular pattern of pronunciation of these words.

5. Diacritic Signs
There are four diacritic sign in the Bengali script: *anusvār* (m), *bisarga* (h), *candrabindu* (˘), and *hasanta* (,). These are used as special orthographic symbols in writing Bengali words. With regard to their usage patterns, it is noted that although these are not allowed to be used independently within a word, these are free to be used with all the basic characters (i.e., vowels, consonants, and clusters) in formation of words with limited restriction. However, each diacritic sign has some traits of uniqueness its use in the language and I have tried to capture these traits in the following sub-sections where each diacritic sign is accounted for.

5.1 The candrbindu
The *candrbindu* is never used with a consonant or a cluster. It is usually used over the head of a vowel and a vowel allograph. The functional importance of this character is primarily context bound, because, when it is detached from the contexts of its usage, it loses its independent functional role.

Phonetically, *candrabindu* is always vocalic in pronunciation in the sense that it takes the support of a vowel to be properly pronounced in words. In general, it nasalizes a vowel or a vowel allograph with which it is attached, as the example below show.

gad [gõd] "gum"	cãd [cãd] "moon"
bãdh [bãdɦ] "dam"	chãd [chãd] "design"
bãdi [bãdi] "maid"	sãkh [ʃãkh] "conch shell"
sãs [ʃãʃ] "kernel"	gĩt [gĩt] "knot"

ṭĕk [ṭæk] "waist band"   bŏdhu [bŏdɦu] "friend"
khṹt [khũt] "fault", etc.

5.2 The bisarga

The *bisarga* is normally used immediately after a vowel, consonant or a cluster in formation of words. It is not allowed to occur at initial position of a word. Therefore it takes place at word-medial and word-final position.

Phonetically, *bisarga* is usually vocalic. That means it is pronounced with inherent [ɔ] or [o] sound in words. Moreover, it generates a slightly aspirated [ɦ] sound for the character with which it is tagged, as the following examples show.

bāḥ   [baɦ] "well"
mulataḥ   [mulɔtoɦ] "mainly"
phalataḥ   [phɔlotɔɦ] "as a result"
punaḥ   [punɔɦ] "again",
ataḥpar   [ɔtoɦpɔr] "after this" etc.

5.3 The anusvār

The *anusvār* is used after a vowel, a consonant, or a cluster within a word. Phonetically it is always non-vocalic. However, it generates a [ŋ] sound when it is attached with a vowel, a vowel allograph or a consonant, as the following examples how.

siṃ [ʃiŋ] "horn"   baṃga [bɔŋgo] "Bengal"
saṃgīt [ʃoŋgit] "song"   sambād [ʃɔŋbad] "news"
bāṃlā [baŋla] "Bengali", etc.

5.4 The hasanta,

The *hasanta* is used usually after a consonant (never with a vowel) at word-medial or word-final position. It never occurs at word-initial position. It does not have any phonetic value. That means no phonetic element is assigned to this character. However, when it is attached with a consonant, it denotes that the consonant should be non-vocalic in pronunciation, as the following examples show.

garam, [gɔrɔm] "hot"   hāḷkā [ɦalka] "light"
mahān, [mɔɦan] "great"   badal, [bɔdɔl] "change"
upal, [upɔl] "pebble"   āḷbāt, [albat] "surely"
kis̤mat, [kiʃmɔt] "fate"   kāṭh, [kaṭh] "wood"
pāḷki [palki] "palanquin"   ban̤dh, [bɔndɦ] "strike"
ul̤ṭā [ulṭa] "opposite", etc.

6. Conclusion

As mentioned in different sections of this paper, it is really difficult to determine how the non-allographed words in Bengali are to be pronounced, since participation and sequential order of the consonants are not the only criteria to be considered in pronunciation of the words. As there are several complexities, which create severe problems in compilation of pronunciation dictionaries, language teaching materials, and text-to-speech conversion, one needs to carefully investigate the patterns of pronunciation of words with reference to the contexts of usage of the consonants and the lexicosemantic information embedded within words.

In this short empirical investigation I have tried to show that there are complexities involved in forms, usages, and pronunciations of some consonants and diacritic signs in the language. In case of consonants, variations in usage and pronunciation are important issues, which cannot be ignored if we want to understand forms, usages and functions of consonants in formation of words in the language.

In fact, the information about the restriction in positional use of consonants and diacritics becomes indispensable in designing the course materials for teaching word formation strategies used in Bengali to the learners. The data, rules, patterns and examples furnished in this paper will invariable help language learners to understand how the consonants occur in words and how these are pronounced in standard Bengali speech. They will also learn from this study how the diacritic signs are used and pronounced in Bengali writing and speech. Other areas of applied and computational linguistics (e.g., pronunciation dictionary, machine learning, text-to-speech conversion, language and speech recognition, etc.) will also benefit from the data and information presented in this paper.

Notes and Comments

[1] From statistical point of view, the number of such words is so less that one can treat these examples as exceptions, since this is not the regular way of writing Bengali words (Haque 1995, Bhattacharya 2007, Dan 2007).

[2] Perhaps this change in pronunciation of *rh* is caused due to recurrent use of the word in regular Bengali speech. Since the Bengali people have a tendency to pronounce word-final non-allographed consonant as a non-vocalic one, this character, is put into this frame to be pronounced in the same manner.

[3] I have used 'å' in the words to imply that the inherent [ɔ] sound is physically present with the consonants used in formation of words but not reflected orthographically.

[4] In case of allographed words there are instances where meaning and part-of-speech control the pronunciation of words. For instance, the word *badale* is pronounced as [bɔdole] "in exchange o" when it is used as an adverb, but pronounced as [bodle] "changing" when it is used as a non-

finite verb. Similarly, *surama* is pronounced as [ʃurma] when it means "collyrium" but pronounced as [ʃurɔma] when it means "beautiful woman", etc.

[5] If we take into consideration the Bengali words made with four or five non-allographed consonants, variations will be much more. And it will be a real tough task to define the patterns of pronunciation of these words.

[6] This kind of pronunciation is noted in onomatopoeic and idiophonic words used in Bengali. At the time of pronunciation, these words are actually broken into two syllables (CVC_1-CVC_2) and the last consonant of each syllable is usually pronounced as a non-vocalic one.

References

Ansari, Sakhawat. 1999. Dhvani, bagdhvani, dhvanibijnan: ekti paryalocana (Sound, Speech Sound, and Phonetics: a Discussion). *Journal of the Institute of Modern Languages*. Dhaka, Bangladesh, Vol. 8. 21-35.

Aronoff, Mark. 1981. *Word Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.

Bhattacharya, Gauri Shankar. 2005. *Bangla Uccharaner Sahaj Path* (An Easy Lesson on Bengali Pronunciation). Kolkata: Yogamaya Prakashani.

Bhattacharya, Subhas. 2007. *Bhalo Bangla Shikhte Hale* (To Learn Good Bengali). Kolkata: Sahitya Samsad.

Dan. Mina. 2007. Mid-front semivowels in Bangla: A Case Study at the Phonology-Morphology interface. Dash. N S. Dasgupta, P.; and Sarkar, P. (Eds.) (2007). *Rainbow of Linguistics. Vol.-I*. Kolkata: T. Media Publications.117-128.

Dash, Niladri Sekhar. 2006. *Bahurupi Bangla Banan* (Multifaceted Bengali Spelling). Kolkata: Dakshabharati.

Dash, Niladri Sekhar, etal (eds.) 2007. *Rainbow of Linguistics. Vol.-I*. Kolkata: T. Media Publications.117-128.

Hai, Abdul. 1985. *Dhanibijnan O Bangla Dhvanitattva*. (Phonetics and Bengali Phonology). 4th Reprint. Dhaka: Mullick Brothers.

Haque, Enamul. 1995. *Bangla Bakdhvani : Svarup O Binyas*. (Bengali Speech Sounds: Nature and Distribution). Dhaka: Ayantika.

Hossain, K.A. 1989. *Sadharan Dhanitattver Upadan*. (Materials of General Phonology). Dhaka: Bangla Academy.

Pal, Palash Baran. 2001. *Dhvanimala Barnamala* (Speech Sounds and Alphabets). Kolkata: Pyapirus.

Sarkar, Pabitra. 1992. *Bangla Balo*. (Speak Bengali). Kolkata: Prama Prakashani.

Sarkar, Pabitra. 1992. *Bangla Banan Sanskar: Samasya o Sambhabana* (Bengali Spelling Reform: Problems and Possibilities). Kolkata: Chirayata Prakasan.

Schane, S.C. 1972. *Generative Phonology*. Chicago: Prentice Hall.

□□□