Investigating into the Patterns of Usage and Nature of Pronunciation of Some Consonant Grapheme Clusters in Bengali

Niladri Sekhar Dash
Linguistic Research Unit
Indian Statistical Institute, Kolkata, India
Email: ns_dash {at} yahoo.com

ABSTRACT— *In this paper I have made an attempt to describe the surface forms, usage patterns, and pronunciation variations of some Bengali consonant grapheme clusters, which are found as great stumbling blocks in language learning, optical character recognition, text-to-speech conversion, speech synthesis, and language processing. To make this study authentic, I have used a moderately large lexical database obtained from a written corpus of modern Bengali prose texts. I have tried to study how these consonant grapheme clusters are orthographically designed by way of combining consonant graphemes and consonant graphic variants; how these clusters are utilized in formation of words; and how these clusters are actually pronounced within words in standard Bengali speech. I have also tried to discuss how the pronunciation of individual consonant graphemes is largely affected when these characters are combined in formation of clusters; and how the pronunciation of these clusters, that are controlled by various phonological rules and processes that operate in standard Bengali speech, leave notable marks on orthographic representation of words in written texts.*

Keywords— consonants, clusters, pronunciation, Bengali

1. INTRODUCTION

The origin and usage of consonant grapheme clusters in the Bengali script vis-à-vis in written text may be traced back in Sanskrit (Devanagari script), since the patterns of formation and the nature of use of consonant grapheme clusters in modern Bengali texts, barring a few instances, is almost similar to those available in Sanskrit texts (Chattopadhyay 1974: 56). One of the basic reasons behind the formation of grapheme clusters in Bengali lie in *phonetic clusters*, which have acted as one of the main catalysts for formation of *orthographic clusters* used in the written texts of the language[1].

In general, it has been observed that in the standard Bengali speech, two consecutively occurring consonants are sometimes phonetically combined together to form a phonetic cluster for ease or simplification in pronunciation. In the later stage, most of these phonetic clusters are represented orthographically in the written text. This particular phenomenon, in return, manifests how the normal speech habit of the native Bengali speakers can have a direct impact on the orthography or script of the language.

Although the modern Bengali texts easily show the present of large number of consonant grapheme clusters formed by combining two or more consonant graphemes, there is not a single cluster, which is formed by combining two or more vowel graphemes[2]. This is quite surprising because the pronunciation of diphthongs, glides, and hiatus is very much present in the speech of the language. These phenomena could have been instrumental in formation of vowel grapheme clusters in the language.

In a simple estimation there are more than 450 consonant grapheme clusters, which are found to be used in modern Bengali written texts. Out of these, ± 350 clusters are made of two consonant graphemes; ± 90 clusters are made of three consonant graphemes; and ± 10 clusters are made of four consonant graphemes. It is perhaps true that the Sanskrit language did not posses so many varieties of consonant grapheme clusters as these are found in the modern Bengali texts (Pal 2001: 183). It also observed that the modern Bengali texts contain many new consonant grapheme clusters, which are never used before – neither in Sanskrit nor in old Bengali texts (Dash 2006: 52).

With regard to pronunciation variations of consonant grapheme clusters in Bengali, it is observed that some clusters show unique pronunciation variations, which are different from individual pronunciation of consonant graphemes combined to form clusters. Due to certain linguistic or non-linguistic factors, pronunciations of some consonant graphemes are highly affected, which are eventually reflected in pronunciation of clusters in words where these clusters are used. An attempt is made in this paper to identify the factors and contexts which are responsible for causing variations in pronunciation of consonant graphemes in context-bound situations. The study carries importance not only in the area of language teaching, text processing, machine learning, pronunciation dictionary compilation, and text-to-speech conversion, but also throws some lights on the pronunciation behaviour of the native Bengali speakers.

Some discussions about the surface forms of consonant grapheme clusters in general are presented in Section 2; the usage patterns of consonant clusters is referred to in Section 3; general rules of pronunciation of consonant clusters is presented in Section 4; pronunciation patterns of those consonant clusters, which reveal variations, are presented in Section 5; and application relevance of this study is reported in Section 6.

2. SURFACE FORMS OF CONSONANT GRAPHEME CLUSTERS

It has been noted by many scholars that the basic shapes of Bengali consonant graphemes undergo change in three different ways at the time of consonant cluster formation (Sarkar 1993, Nath 2004, Mukhopadhyay 2005, Dash 2006: 61): (i) in case of some consonant clusters, the original shapes of participating consonant graphemes are totally altered; (ii) in case of some other consonant clusters, shapes of the participating consonant graphemes are partially modified; and (iii) for the rest of the consonant clusters, shapes of the participating consonant graphemes remain mostly unaffected. Based on this observation, it is possible to classify Bengali consonant clusters into three broad types, as stated below:

(a) **Fully transparent consonant clusters:** Those consonant clusters are formed in such a way where the original shapes of the participating consonant graphemes have always remained unaffected. Therefore, just by looking at the surface forms of the clusters one can easily identify which consonant graphemes are actually combined together to generate final surface forms, e.g., *kk, kv, kl, gn, gv, gl, jv, dd, db, nv, nn, pn, pl, śl, sv, hv, hl,* etc. In estimation, there are ± 150 clusters, where shapes of both the consonant graphemes are normally unaffected.

(b) **Partially transparent consonant clusters:** In case of these clusters the shape of one of the participating consonant graphemes remains usually unaffected, while the shape of the other consonant grapheme is grossly modified, e.g., *ky, gr, br, ghr, rk, ñj, jñ, hn,* etc. In a simple count, there are ± 60 clusters, where the shape of the first consonant grapheme is heavily changed, and there are ± 100 clusters, where the shape of the second consonant grapheme is severely affected.

(c) **Completely non-transparent consonant clusters:** In case of these clusters the basic shapes of the participating consonant graphemes are changed to such an extent that it is simply impossible to identify, by observing the final surface forms, which two consonant graphemes are actually combined together to generate the final forms, e.g., *kṣ, kt, kr, ñc, ṅg, tt, tr, hm*, etc. There are near about 8 clusters of this type in the present Bengali script (Biswas 2005, Mukhopadhyay 2005).

In general, language learners can face difficulties in recognizing the partially transparent and completely non-transparent clusters due to built-in complexities involved in the surface forms of these character combinations (Dash 2007: 32). Also, these clusters can create great deal of technical problems in case of developing systems for character recognition, character encoding, script recognition and machine learning of scripts. To overcome these problems, people have suggested for taking initiative for simplification of shapes of completely non-transparent clusters so that these becomes more transparent and easily recognizable (Sarkar 1993, Nath 2004, Mukhopadhyay 2005, Chakrabarti 2008). To materialize the suggestions of experts, the Bangla Akademi of West Bengal has modified shapes of some clusters as well as has used these modified versions of the clusters in writing and publishing Bengali texts (Sarkar, Mukhopadhyay and Dasgupta 2003). Even then, in the printed Bengali documents available today, both the transparent and non-transparent shapes of these clusters are found to be used in regular frequency.

3. USAGE PATTERNS OF CONSONANT GRAPHEME CLUSTERS IN BENGALI

A simple look into the modern Bengali written text databases can reveal the presence of large number of consonant clusters that have occurred in all kinds of writing belonging to various disciplines of human knowledge. However, it is interesting to note that all the clusters are not entitled to occur at all the positions within words. Out of the total number of clusters available in the language script, only a limited few, due to Morpheme Structure Rules (Hyman 1975: 105-108) and Sequence Redundancy Rules (Schane 1921: 40) are entitled to take place at word-initial position, while other clusters can occur at word-medial and word-final positions. Interestingly, the clusters, which are allowed to take place at word-initial position, are also allowed to take place at medial and final position, which, however, is not true to those clusters, which are allowed to occur at word-medial and word-final positions only. In a simple count, there are nearly 107 consonant clusters, which can occur at word-initial position only in Bengali. These clusters are presented below in 8 (eight) broad groups with examples obtained from the corpus. The list includes clusters which are inherited from Sanskrit as well as clusters borrowed from foreign languages like Arabic, Persian, Russian, and English, etc.

Type 1: Velar Consonant Group (18)

ky	(kyāblā)	[kæbla]	"idiot"	kr	(kramik)	[kromik]	"serial"
kry	(krabk)	[kræb]	"crab"	kv	(kvacit)	[kotʃit]	"rarely"
kl	(klānta)	[klanto]	"tired"	kly	(klyāp)	[klæp]	"clap"
kṣ	(kṣati)	[khoti]	"loss"	kṣm	(kṣmā)	[kha]	"earth"
khy	(khyāpā)	[khæpa]	"mad"	khr	(khrīṣṭa)	[khriʃʈo]	"Christ"
gy	(gyās)	[gæʃ]	"gas"	gr	(grām)	[gram]	"village"
gry	(gryānḍ)	[grænḍ]	"grand"	gy	(gvāliyar)	[goaliɔr]	"Gowalior"
gl	(glāni)	[glani]	"defame"	gly	(glyāḍ)	[glænḍ]	"glad"
ghy	(ghyām)	[gɦæm]	"pride"	ghr	(ghrāṇ)	[gɦran]	"smell"

Type 2: Palatal Consonant Group (6)

cy	(cyuti)	[tʃuti]	"fault"	chy	(chyāblā)	[tʃhæbla]	"rascal"
jñ	(jñānī)	[gẽni]	"wise"	jv	(jvar)	[ɟɔr]	"fever"
jy	(jyā)	[ɟæ]	"chord"	jhy	(jhyā̃tā)	[ɟɦæ̃ʈa]	"broom"

Type 3: Retroflex Consonant Group (8)

ṭy	(ṭyārā)	[ʈæra]	"squint"	ṭr	(ṭrām)	[ʈram]	"tram"
ṭry	(ṭryāp)	[ʈræp]	"trap"	ṭhy	(ṭhyālā)	[ʈhæla]	"push"
ḍy	(ḍyākrā)	[ɖækra]	"sly"	ḍr	(ḍrām)	[ɖram]	"drum"
ḍry	(ḍryāg)	[ɖræg]	"drag"	ḍhy	(ḍhyāmnā)	[ɖɦæmna]	"lewd"

Type 4: Dental Consonant Group (14)

ty	(tyakta)	[tɔkto]	"left"	tr	(trās)	[traʃ]	"fear"
tv	(tvak)	[tɔk]	"skin"	try	(tryakṣar)	[trɔᵏkhor]	"tri-syllabic"
thy	(thyābṛā)	[thæbṛa]	"flat"	thr	(thro)	[thro]	"throw"
dy	(dyuti)	[duti]	"glow"	dr	(druta)	[druto]	"fast"
dv	(dvīp)	[dip]	"island"	dvy	(dvyarthak)	[dɔrthok]	"ambiguous"
dhy	(dhyān)	[dɦæn]	"meditation"	dhr	(dhruba)	[dɦrubo]	"sure"
dhv	(dhvani)	[dɦoni]	"sound"	ny	(nyasta)	[nɔsto]	"entrusted"

Type 5: Labial Consonant Group (21)

py	(pyāc)	[pætʃ]	"patch"	pr	(pragati)	[progoti]	"progress"
pry	(pryākṭis)	[prækṭiʃ]	"practice"	pl	(plāban)	[plabon]	"flood"
ply	(plyān)	[plæn]	"plan"	phy	(phyāsād)	[phæsad]	"trouble"
phr	(phrak)	[phrɔk]	"frock"	phry	(phryāgmenṭ)	[phrægmenʈ]	"fragment"
phl	(phlo)	[phlo]	"flow"	phly	(phlyāṭ)	[phlæʈ]	"flat"
by	(byakti)	[bekti]	"person"	br	(brata)	[broto]	"oath"
bry	(bryānḍ)	[brænḍ]	"brand"	bl	(blāuj)	[blauɟ]	"blouse"
bly	(blyāk)	[blæk]	"black"	bhy	(bhyālā)	[bɦæla]	"good"
bhr	(bhram)	[bɦrɔm]	"error"	bhr	(bhlādimir)	[bɦladimir]	"Vladimir"
my	(myāc)	[mætʃ]	"match"	mr	(mriyamāṇ)	[mrioman]	"glum"
ml	(mlān)	[mlan]	"gloomy"				

Type 6: Liquid Consonant Group (2)

ry	(ryāpar)	[ræpar]	"wrapper"	ly	(lyājā)	[læɹa]	"tail"

Type 7: Sibilant Consonant Group (34)

śv	(śvās)	[ʃaʃ]	"breath"	śm	(śmaśru)	[ʃõˈʃru]	"beard"
śy	(śyāmal)	[ʃæmol]	"green"	śr)	(śram)	[ʃrɔm]	"labour"
śl	(ślīl)	[ʃlil]	"polite"	ṣṭ	(ṣṭap)	[ʃʈɔp]	"stop"
ṣṭh	(ṣṭhīban)	[ʃʈhibɔn]	"spit"	sk	(skul)	[ʃkul]	"school"
sky	(skyām)	[ʃkæm]	"scam"	skr	(skru)	[ʃkru]	"screw"
skry	(skryāp)	[ʃkræp]	"scrap"	skh	(skhalan)	[ʃkhɔlon]	"fall"
sṭ	(sṭal)	[ʃʈɔl]	"stall"	sty	(styāmp)	[ʃʈæmp]	"stamp"
sṭy	(sṭreṭ)	[ʃʈreʈ]	"straight"	sṭry	(sṭryāp)	[ʃʈræp]	"strap"
st	(stan)	[stɔn]	"breast"	str	(strī)	[stri]	"wife"
sth	(sthir)	[sthir]	"still"	sn	(snān)	[snan]	"bath"

sny	(snyāks)	[snækʃ]	"snacks"	sp	(sparśa)	[ʃpɔrʃo]	"touch"
spy	(spyāniś)	[ʃpæniʃ]	"Spanish"	spr	(spriṁ)	[ʃpriŋ]	"spring"
spry	(spryāṁ)	[ʃpræŋ]	"sprang"	spl	(spliṭ)	[ʃpliṭ]	"split"
sply	(splyāś)	[ʃplæʃ]	"splash"	sph	(sphaṭik)	[ʃphoṭik]	"crystal"
sv	(svar)	[ʃɔr]	"voice"	sm	(smar)	[ʃmɔr]	"god"
sy	(syāndan)	[ʃændon]	"filtration"	sr	(sraṣṭā)	[ʃrɔʃṭa]	"creator"
sl	(sleṭ)	[ʃleṭ]	"slate"	sly	(slyāb)	[ʃlæb]	"slab"

Type 8: Glottal Fricative Consonant Group (4)

hy	(hyāpā)	[ɦæpa]	"trouble"	hr	(hrad)	[ɦrɔd]	"lake"
hv	(hvān)	[ɦian]	"self"	hl	(hlād)	[ɦilad]	"joy"

On the other hand, in Bengali script, there are large numbers of consonant grapheme clusters, which can occur only at word-medial and/or word-final positions, but cannot take place at word-initial position. Only a few examples are presented below.

pt	(saptāha)	[ʃɔptaɦo]	"week"	kt	(baktā)	[bɔkta]	"speaker"
lṭ	(ulṭā)	[ulṭa]	"reverse"	lp	(alpa)	[ɔlpo]	"few"
ṇḍ	(aṇḍa)	[ɔnḍo]	"egg"	lk)	(śulka)	[ʃulko]	"tax"
nd	(manda)	[mɔndo]	"bad"	mb	(ambar)	[ɔmbor]	"sky"
nt	(anta)	[ɔnto]	"end"	pn	(svapna)	[ʃɔpno]	"dream"

These consonant clusters cannot occur at word-initial position because the organs of speech, which are deployed in pronunciation of sounds, are not volatile enough to produce such combinations of sounds at the beginning of a syllable or word.

4. PRONUNCIATIONS OF CONSONANT CLUSTERS: GENERAL RULES

During last few decades several studies have reported about overall pronunciation patterns of consonant clusters in standard Bengali speech (Chattopadhyay 1974, Hai 1985, Hossain 1989, Bhattacharya 1992, Sarkar 1992, Haque 1993, Sarkar 1993, Haque 1995, Ray 1997, Sakhawat 1999, Bhattacharya 2000, Pal 2001, Nath 2004, Biswas 2005, Mukhopadhyay 2005, Bhattacharya 2006, Dan 2006, Dash 2006, Dan 2007, Haldar 2007).

From these studies it is observed that in their pronunciation, for most of the consonant clusters, the participating consonant graphemes usually follow the regular sequences of their orthographic occurrence in words. That means, the orthographic sequence of C_1C_2 of a consonant cluster is normally retained in its pronunciation, where the C_1 is first pronounced followed by C_2. For instance, in case of the cluster <kl>, the consonant grapheme <k> is first pronounced followed by <l>, as in (śuklā) [ʃukla] "white".

However, in case of some clusters, this order is not always restored in accordance with the sequence of orthographic occurrence of the consonant graphemes. That means, while orthographically a cluster contains C_1C_2, in pronunciation it yields C_2C_1. This reverse process is observed in case of clusters like <hṇ>, <hn>, <hm>, <hl>, etc, which are discussed in following subsections. The sequential order of the consonant graphemes is reversed or modified due to several phonological processes, such as, deletion, addition, substitution, and displacement of sounds, etc. Also, there are some other issues relating to pronunciation of clusters, which are discussed in the following paragraphs. In this section, I have discussed pronunciation of those clusters which record variations – therefore those clusters, which do not show any notable variation in pronunciation, are not considered here. The pronunciation of each cluster is presented within a word as well as supported with *International Phonetic Alphabets* (IPA) for better comprehension.

In general, at word-final position, all non-allographed clusters are vocalic in pronunciation with [o] sound. This is, however, true to most of the *Tatsama, Tadbhaba* and *Deśi* (i.e., of native origin) consonant clusters, as the following examples show:

(asta)	[ɔsto]	"set"	(masta)	[mɔsto]	"big"
(kilṣṭa)	[kliʃṭo]	"tired"	(śiṣṭa)	[ʃiʃṭo]	"rest"
(kāṣṭha)	[kɑʃṭho]	"wood"	(kalaṅka)	[kɔlɔŋko]	"defame"
(binyasta)	[biⁿnɔsto]	"arrayed"	(patra)	[pɔˈtro]	"leaf"
(citra)	[ciˈtro]	"picture"	(bikalpa)	[bikɔlpo]	"other"
(paṛanta)	[pɔɽɔnto]	"falling"	(tapta)	[tɔpto]	"hot"
(salajja)	[ʃɔlɔˈɟo]	"shy"	(manda)	[mɔndo]	"bad"
(basanta)	[bɔsonto]	"spring"	(dhasta)	[dɦɔsto]	"exhausted"

(gandha)	[gɔndɦo]	"smell"	(nagna)	[nɔgno]	"naked"
(jabda)	[ɟɔbdo]	"tamed"	(stabdha)	[stɔbdɦo]	"silent"
(phalanta)	[phɔlɔnto]	"bearing"	(jvalanta)	[ɟɔlɔnto]	"burning"
(barañca)	[bɔrɔnco]	"rather"	(taraṅga)	[tɔrɔŋgo]	"wave"
(bandha)	[bɔndɦo]	"closed"	(spaṣṭa)	[ʃpɔʃʈo]	"clear"
(samagra)	[ʃɔmɔgro]	"all"	(prānta)	[pranto]	"end"

Only one or two exceptions are found to occur in case of place names where the word-final consonant grapheme of the cluster <ñj> is pronounced as open-ended one, such as, (bāligañj) [baligɔɲɟ] "Bullygunj", (ṭāligañj) [ʈaligɔɲɟ] "Tollyganj", (rānigañj) [ranigɔɲɟ] "Raniganj", etc.

However, in case of those words that are borrowed from Persian, Arabic, Hindi, English, and other languages, the word-final non-allographed consonant cluster is usually non-vocalic in pronunciation. In these cases, at least, the impact of the foreign pronunciation is easily identified in normal Bengali speech habits, as the following examples can exhibit:

(hind)	[ɦind]	"Hindi"	(pasand)	[pɔʃɔnd]	"liking"
(bandh)	[bɔndɦ]	"strike"	(dost)	[dost]	"friend"
(behest)	[beɦest]	"heaven"	(mast)	[mɔst]	"stupor"
(putra)	[putr]	"son"	(karm)	[kɔrm]	"work"
(dharm)	[dɦɔrm]	"religion"	(julm)	[ɟulm]	"force"
(basant)	[bɔʃɔnt]	"spring"	(hemant)	[ɦemɔnt]	"Autumn"
(cost)	[ʧost]	"pure"	(janm)	[ɟɔnm]	"birth"
(taraṅ)	[tɔrɔŋ]	"wave"	(kalaṅk)	[kɔlɔŋk]	"defame"
(bālb)	[balb]	"bulb"	(sans)	[ʃɔnʃ]	"sons'"
(nārs)	[narʃ]	"nurse"	(philm)	[philm]	"film"
(gālph)	[galph]	"gulf"	(pārṭ)	[parʈ]	"part"
(śārṭ)	[ʃarʈ]	"shirt"	(pyānṭ)	[pænʈ]	"pant"
(baks)	[bɔkʃ]	"box"	(seks)	[ʃekʃ]	"sex"
(belṭ)	[belʈ]	"belt"	(ṭyānk)	[ʈænk]	"tank"
(ānṭ)	[anʈ]	"aunt"	(ṭenṭ)	[ʈenʈ]	"tent"
(lyānḍ)	[lænɖ]	"land"	(byānḍ)	[bænɖ]	"band"
(gryānḍ)	[grænɖ]	"grand"	(glyānḍ)	[glænɖ]	"gland"
(hyānḍ)	[ɦænɖ]	"hand"	(hājbyānḍ)	[ɦaɟbænɖ]	"husband"
(kanḍākṭ)	[kɔnɖɑkʈ]	"conduct"	(kasṭ)	[kɔʃʈ]	"cost"
(besṭ)	[besʈ]	"best"	(vesṭ)	[bɦeʃʈ]	"vest"
(cesṭ)	[ceʃʈ]	"chest"	(bresṭ)	[breʃʈ]	"breast"
(ṭesṭ)	[teʃʈ]	"test"	(gesṭ)	[geʃʈ]	"guest"
(peśenṭ)	[peʃenʈ]	"patient"	(pesṭ)	[peʃʈ]	"paste"
(posṭ)	[poʃʈ]	"post"	(resṭ)	[reʃʈ]	"rest"
(rosṭ)	[roʃʈ]	"roast"	(ṭosṭ)	[ʈoʃʈ]	"toast"
(risṭ)	[riʃʈ]	"wrist"	(hosṭ)	[ɦoʃʈ]	"host", etc.

For these types of words, the non-allographed consonant clusters, even if they occur at word-final position, are not pronounced in the same manner as regular Bengali and Sanskrit consonant clusters are pronounced.

One interesting observation in made in case of the consonant cluster <ndh> in the word (bandh). Here the non-allographed word final cluster <ndh> is pronounced open-ended [bɔndɦo], if the word is used in the sense of closing something [e.g., (kāl skul bandha thākbe) "the school will remain closed tomorrow"]. On the other hand, the cluster <ndh> is pronounced as closed one [bɔndɦ], if the word is used in the sense of strike [e.g., (kāl ekṭi rājnaitik dal bandh ḍekeche) "a political party has called strike tomorrow], as the following examples show.

(1) (kāl skul bandha thākbe)   (bandha) [bɔndɦo]
     [kal ʃkul bɔndɦo thakbe]
     "The school will remain closed tomorrow"]
(2) (kāl ekṭi rājnaitik dal bandh ḍekeche)   (bandh) [bɔndɦ]
     [kal ekʈi raɟnoitik dɔl bɔndɦ ɖekeʧhe]
     "A political party has called strike tomorrow"

5. VARIATION IN PRONUNCIATION OF SOME CONSONANT CLUSTERS

The consonant clusters, which show notable variations in their pronunciation, are discussed in the following subsections with examples obtained from the Bengali written text corpus.

5.1 The Cluster <kṣ>

The cluster <kṣ> records two pronunciation variations. When it occurs at the initial position of a word, it is pronounced as [kh]. In these cases, the actual pronunciation of the second consonant grapheme [i.e., <ṣ>] is entirely lost, while that of the first consonant grapheme [i.e., <k>] is slightly aspirated. There is no exception to this pattern of pronunciation in the Bengali speech, as the following examples show.

(kṣati)	[khoti]	"loss"	(kṣata)	[khɔto]	"wound"
(kṣobh)	[khobɦ]	"anger"	(kṣay)	[khɔĕ]	"erosion"
(kṣār)	[khar]	"alkali"	(kṣīpra)	[khiᵖpro]	"agile"
(kṣudhā)	[khudɦa]	"hunger"	(kṣīr)	[khir]	"latex", etc.

On the other hand, when this particular consonant cluster <kṣ> occurs at the word-medial or word-final position, pronunciation of the second consonant grapheme [i.e., <ṣ>] is fully lost, while that of the first consonant grapheme [i.e., <k>] is fully aspirated with partial germination of its original sound, as the following examples can show.

(akṣa)	[ɔᵏkho]	"orbit"	(bakṣa)	[bɔᵏkho]	"chest"
(rakṣita)	[roᵏkhito]	"stored"	(akṣāṃśa)	[ɔᵏkhaŋʃo]	"latitude"
(pakṣī)	[poᵏkhi]	"bird"	(kakṣa)	[kɔᵏkho]	"chamber"
(cakṣu)	[ʧoᵏkhu]	"eye"	(pakṣe)	[poᵏkhe]	"for"
(rakṣā)	[roᵏkha]	"save"	(gabākṣa)	[gɔbaᵏkho]	"window"
(sakṣam)	[ʃɔᵏkhɔm]	"able"	(sākṣī)	[ʃaᵏkhi]	"witness", etc.

Also, there is no exception to this pattern of pronunciation in the standard speech habits of the speakers of the language.

5.2 The Cluster <ñc>

In case of the consonant cluster <ñc>, the pronunciation of the first consonant grapheme [i.e., <ñ>] is lost generating a dental nasal sound [n], while the pronunciation of the second consonant grapheme [i.e., <c>] remains mostly unchanged, as the following examples can exhibit:

(añcal)	[ɔnʧol]	"area"	(āñcalik)	[anʧolik]	"local"
(kañci)	[konʧĩ]	"twig"	(cañcal)	[cɔnʧol]	"restless"
(cañcu)	[conʧu]	"beck"	(tañcak)	[tɔnʧɔk]	"traitor"
(bañcanā)	[bɔnʧona]	"deprivation"	(bañcita)	[bɔnʧito]	"deprived"
(barañca)	[bɔrɔnʧo]	"rather"	(mañca)	[mɔtʧco]	"platform"
(pañcam)	[pɔnʧɔm]	"fifth"	(kāñcan)	[kanʧɔn]	"gold"
(sañcay)	[ʃɔnʧɔĕ]	"storage"	(sañcita)	[ʃonʧito]	"stored"
(pañcāś)	[pɔnʧaʃ]	"fifty"	(sañcārī)	[ʃɔnʧari]	"flier", etc.

Not a single variation in pronunciation of these clusters is observed in the Bengali language irrespective to its occurrence at any position within a word.

5.3 The Cluster <ñch>

In case of consonant cluster <ñch>, almost similar pattern of pronunciation is observed as it is noted in case the consonant cluster <ñc>. Here, the pronunciation of the first consonant grapheme [i.e., <ñ>] is lost giving place to a dental nasal sound [n], while the pronunciation of the second consonant grapheme [i.e., <(ch>] remains mostly unaffected, such as the followings:

(uñchabṛtti)	[unʧhobriᵗti]	"begging"	(bāñchā)	[banʧha]	"desire"
(lāñchanā)	[lanʧhona]	"harassment", etc.			

This pronunciation pattern is universally true as far as the standard Bengali pronunciation habits are concerned.

5.4 The Cluster <jñ>

The consonant cluster <jñ> exhibits two pronunciation variations. At word-initial position, both the consonant graphemes lose their respective original pronunciations to generate a [g] sound, as found in the following examples:

(jñān)	[gæn]/[gæ̃n]	"knowledge"	(jñānī)	[gæni]/[gæ̃ni]	"wise"
(jñāta)	[gæto]/[gæ̃to]	"known"	(jñānata)	[gænoto]/[gæ̃noto]	"in sense"
(jñānadā)	[gænoda]/[gæ̃noda]	"knowledge giver"	(jñātārthe)	[gætarthe]/[gæ̃tarthe]	"for learning", etc.

In case of normal pronunciation, however, the vowel of the first syllable is not nasalized, while in standard pronunciation it is expected that it should be nasalized.

On the other hand, if this consonant cluster is used at word-medial or word-final position, pronunciations of both the consonant graphemes are lost to generate a [g] sound, and this [g] sound is doubled while its following vowel sound is nasalized, such as the followings:

(bijñān)	[bi^{g}gæ̃n]	"science"	(bijñānī)	[bi^{g}gæ̃ni]	"scientist"
(ajñān)	[ɔ^{g}gæ̃n]	"senseless"	(abhijñān)	[obɦi^{g}gæ̃n]	"proof"
(abhijñatā)	[obɦi^{g}gõta]	"experience"	(abhijña)	[obɦi^{g}gõ]	"experienced"
(dharmajña)	[dɦormɔ^{g}gõ]	"religious-minded"	(bijña)	[bi^{g}gõ]	"wise", etc.

No variation of pronunciation is observed in standard Bengali speech.

5.5 The Cluste <ñj>

In case of the consonant cluster <ñj>, pronunciation of the first consonant grapheme [i.e., <ñ>] is entirely lost to generate a dental nasal [n] sound, while pronunciation of the second consonant grapheme [i.e., <j>] remains unchanged, as shown in the examples below.

(añjan)	[ɔnɟɔn]	"eye-salve"	(añjali)	[ɔnɟoli]	"prayer"
(gañjanā)	[gɔnɟona]	"rebuke"	(jañjāl)	[jɔnɟal]	"garbage"
(bañjul)	[bonɟul]	"curved"	(mañjirā)	[monɟira]	"anklet"
(mañju)	[monɟu]	"beautiful"	(mañjarī)	[mɔnɟori]	"bud"
(mañjil)	[monɟil]	"palace"	(mañjur)	[monɟur]	"sanction"
(rañjanā)	[rɔnɟona]	"dyer"	(rañjita)	[ronɟito]	"coloured"
(khañjani)	[khɔnɟoni]	"tambourine"	(iñjin)	[inɟin]	"engine"
(pañjikā)	[ponɟika]	"almanac"	(bhañjan)	[bɦɔnɟɔn]	"removal", etc.

Not a single variation is observed in pronunciation of this cluster in the standard Bengali speech habits.

5.6 The Clusters <tm>, <dm> and <kṣm>

In case of the consonant clusters <tm>, <dm>, and <kṣm>, pronunciation of the second consonant grapheme [i.e., <m>] is entirely lost, while pronunciation of the first consonant grapheme or cluster [i.e., <t>, <d>, <kṣ>] remains mostly unchanged and the vowel which immediately follows them is nasalized, as the following examples show:

(ātmā)	[a^{t}tã]	"soul"	(mahātmā)	[mɔɦa^{t}tã]	"great"
(padma)	[pɔ^{d}dõ]	"lotus"	(lakṣmī)	[lo^{k}khĩ]	"Goddess of wealth"
(sūkṣma)	[ʃu^{k}khõ]	"fine"	(pakṣmal)	[pɔ^{k}khɔl]	"eye lids", etc.

There is hardly any exception to this pattern of pronunciation in standard Bengali speech.

5.7 The Clusters <śm> and <sm>

The consonant clusters <śm> and <sm> show two pronunciation variations in standard Bengali speech. At word-initial position, the second consonant grapheme [i.e., <m>] loses its own pronunciation, while pronunciation of the first consonant grapheme [i.e., <ś> or <s>] remains mostly unchanged and their immediately following vowel sound is nasalized, as the followings example show:

(śmaśān)	[ʃɔ̃ʃan]	"burning ghat"	(śmaśru)	[ʃõʃru]	"beard"
(smaraṇ)	[ʃɔ̃rɔn]	"memorization"	(smar)	[ʃɔr]	"god"
(smārak)	[ʃãrɔk]	"memento"	(smṛti)	[ʃr̃ĩti]	"memory", etc.

Some exceptions in pronunciation, are however, noted in case of a few words where both the consonant graphemes are distinctly pronounced in their respective original sounds, such as, (smitā) [ʃmita] "smiling", (smita) [ʃmito] "smiled", etc.

On the other hand, at word-medial or word-final position, for the consonant clusters <śm> and <sm> in some specific words, pronunciation of the first consonant grapheme [i.e., <ś> and <s>] as well as pronunciation of the second consonant grapheme [i.e., <m>] is distinctly perceived without any alternation, as the following examples reveal:

(asmitā)	[oʃmita]	"egoism"	(susmitā)	[ʃuʃmita]	"smiling"
(kāśmīr)	[kaʃmir]	"Kashmir"	(kuṣmāṇḍa)	[kuʃmanḍo]	"pumpkin"
(parasmaipadī)	[pɔrɔʃmoipodi]	"parasite"	(tasmin)	[toʃmin]	"your"
(ýasmin)	[ɹoʃmin]	"whose", etc.			

Some exceptions to this pattern are, however, found in case of some words where the first consonant grapheme [i.e., <ś> and <s>] is partly doubled in pronunciation, while its following consonant grapheme [i.e., <m>] loses its pronunciation to nasalize its immediately succeeding vowel sound, such as, (rasmi) [roʃʃĩ] "rays", (grīṣma) [griʃʃõ] "summer", (bismay) [biʃʃɔĕ] "surprise", (bismita) [biʃʃĩto] "surprised", etc.,

5.8 The Cluster <hm>

In case of the consonant cluster <hm>, although the actual pronunciations of first and second consonant graphemes remain usually unchanged, the actual orthographic sequence of the consonant graphemes (C_1C_2) is reversed in pronunciation (C_2C_1). As a result, the second consonant grapheme [i.e., <m>] is pronounced first followed by pronunciation of the first consonant grapheme [i.e., র (h)], as the following examples exhibit:

(brahma)	[bromɦo]	"Brahma"	(brāhmī)	[bramɦi]	"Brahmi"
(brahmottar)	[bromɦoᵗtor]	"after Brahma"	(brāhmaṇ)	[bramɦon]	"Brahmin", etc.

The uniqueness of pronunciation of the particular cluster also asks for careful treatment of the cluster in language teaching and text-to-speech conversion.

In case of other clusters made with the labial nasal consonant grapheme <m> as second member [e.g., <gm>, <nm>, <lm>, etc.], the sequential order and original pronunciation of the consonant graphemes remain mostly unaltered, as shown below:

(bāgmī)	[bagmi]	"orator"	(ýugma)	[ɹugmo]	"dual"
(manmatha)	[mɔnmɔtho]	"Cupid"	(janma)	[ɹonmo]	"birth"
(unmād)	[unmad]	"insane"	(unman)	[unmɔn]	"restless"
(gulma)	[gulmo]	"shrub"	(balmīk)	[bolmik]	"ant-hill"
(śālmalī)	[ʃalmoli]	"teak", etc.			

No variation to this pattern of pronunciation is observed in the standard Bengali speech.

5.9 The Consonant Clusters <hṇ> and <hn>

Similar to the consonant cluster <hm>, in case of the consonant clusters <hṇ> and <hn>, although actual pronunciation of the consonant graphemes remains mostly unchanged, the actual orthographic sequences of consonant graphemes (C_1C_2) are reversed in pronunciation (C_2C_1). As a result, the second consonant grapheme [i.e., <ṇ> or <n>] is pronounced first followed by the pronunciation of the first consonant grapheme [i.e., <h>], as the following example show:

(cihna)	[cinɦo]	"sign"	(aparāhṇa)	[ɔporanɦo]	"afternoon"
(sāyāhna)	[ʃaĕanɦo]	"evening"	(cihnita)	[cinɦito]	"marked"
(bahni)	[bonɦi]	"fire"	(jāhṇabī)	[ɹanɦobi]	"an Indian river", etc.

This is a unique type of pronunciation of the cluster and it demands careful treatment at the time of language teaching and text-to-speech conversion.

5.10 The Cluster <hl>

In case of the consonant cluster <hl>, the original sequence of occurrence of the consonant graphemes is reversed in pronunciation. Here also, the actual pronunciation of the first and the second consonant grapheme becomes almost reverse to the orthographic sequence of the consonant graphemes. As a result, the second consonant grapheme [i.e.,

<l>] is pronounced first followed by the pronunciation of the first consonant grapheme [i.e., <h>], as the following examples show:

(āhlād)	[alɦad]	"delight"	(āhlādita)	[alɦadito]	"delighted"
(āhlādi)	[alɦadi]	"over-indulged"	(kahlār)	[kɔlɦar]	"white water-lily"
(prahlād)	[prolɦad]	"a person name"	(pahlabī)	[pɔlɦobi]	"an Indian race", etc.

There is also no exception to this pattern of pronunciation of this cluster in the standard pronunciation of the language. Other consonant clusters found to be used in Bengali writings as such do not show much variation in sequential order of consonant graphemes in pronunciation. Moreover, they do not exhibit any kind of deviation from the original pronunciation of the individual consonant graphemes.

It is to be understood that in normal speech of native Bengali speakers, many consonant sounds are combined together to generate *phonetic clusters*, which are not represented orthographically in written texts. If these phonetic clusters are converted into orthographic clusters, the number of orthographic clusters will be large and more varied in the language. The trailer of the trend is reflected in modern Bengali written text documents where we come across instances of formation of new orthographic clusters, as shown in the following examples[3]:

<tl>	(kātlā)	[katla]	"carp"	<lt>	(bālti)	[balti]	"bucket"
<ṭl>	(jaṭlā)	[ɹɔʈla]	"gathering"	<tn>	(phātnā)	[phatna]	"bait"
<spl>	(splinṭār)	[ʃplinʈar]	"splinter"	<bhl>	(bhlādimir)	[bɦladimir]	"Vladimir"
<ḍl>	(ḍlāusi)	[ḍlauʃi]	"Dalhousie", etc.				

6. CONCLUDING REMARKS

In this paper I have tried to show that there are complexities involved in from, usage, and pronunciation of some of the Bengali consonant grapheme clusters. Variations in usage and pronunciation are important features that cannot be ignored if we try to understand forms, usage, and functions of these clusters in word formation as well as in pronunciation in the language. To make this study authentic, I have accessed and referred to a corpus that contains samples of texts from more than eighty five subject areas and disciplines with a total strength of more than five million (i.e., fifty lakhs) of words.

From analysis of data it is understood that information about restriction in positional use of clusters becomes indispensable for designing course materials for learning word formation strategies used in the language. At certain contexts, it becomes impossible to determine how some consonant clusters will be pronounced rightly, as participation and sequential order of the consonant graphemes are not the only criteria to be considered in pronunciation of the clusters. Since so many complexities are bound to evoke many problems in compilation of pronunciation dictionary, guidebook preparation, language teaching, and text-to-speech conversion, we need to deal with these issues very carefully with close reference to the contexts of occurrence of the characters as well as on the lexicosematic information embedded within the words.

This moderately elaborate discussion presented here will help the language learners to understand how consonant grapheme clusters occur within words in Bengali texts and how these are pronounced in standard Bengali speech. Also they will learn how these are structurally formed[4]; which clusters are allowed to take place at which positions of a word; and which clusters should be uttered in which manner to produce acceptable pronunciation. This description has also strong relevance for the people who are working in the area of speech technology for the Bengali language. They can utilize necessary information and examples from this study to develop a system that can generate artificial Bengali speech nearly similar to the standard pronunciation of words.

Although efforts are made to standardize pronunciation of some Bengali consonant clusters (Hai 1985, Sarkar 1992, Haque 1995, Ray 1997, Ansari 1999, Dan 2006, Bhattachraya 2006) these are still far from being uniform, as Bengali language users continue to have different pronunciation resulting in variations of pronunciation of some clusters. Since this is inevitable in a living language having several regional varieties, a workable standard is, none-the-less, expected that will work as a benchmark for formal and official purposes. The data, analysis, and information given in this paper may work in this direction to empower language users with necessary knowledgebase to have preliminary ideas about form, function, and pronunciation of clusters used in the language.

7. ACKNOWLEDGEMENT

February 2014. I humbly and happily acknowledge that all the comments and suggestions of the audience are accepted and properly implemented in this revised version of the paper.

8. NOTES AND COMMENTS

[1] Scholars have identified *phonetic clusters* as *ear clusters* and *orthographic clusters* as *eye clusters*, since *ear clusters* are perceived in speech while *eye clusters* are visualised in written documents (Sarkar 1993: 25).

[2] This is because the vowels, due to their inherent phonetic entity, cannot combine together to form *vowel clusters*, as it naturally happens to the consonants. As a result, the vowels, when they are placed in situations where they can converge, usually conjoin to generate diphthongs, glides or hiatus. Also they can converge, in such situations, to generate a vowel having different set of features, as observed in case of *sandhi* and euphonic combinations of vowel sounds.

[3] This tendency is becoming regular in Bengali writing due to use of computer in writing texts and desk-top publishing in Bengali. We shall not be surprised if, in near future, some of these clusters are included in existing list of Bengali clusters. This will not only increase the number of clusters in the language but also will create more problems with regard to their standard utterance in standard speech.

[4] Because of difference in occurrence and function of consonant graphic variants, development of automatic character recognition system in Bengali has become quite problematic. Similar problem is faced by other Indian scripts also where consonant graphic variants are used as essential part of writing system.

9. REFERENCES

- Ansari, Sakhawat. "dhvani, bāgdhvani, dhvanibijñān: ekṭi paryālocanā (phones, phonemes, phonetics: an analysis)". *Journal of the Institute of Modern Languages* (Dhaka, Bangladesh), vol. 3, no. 1, pp. 22-46, 1999.
- Bhattacharya, Gauri Shankar, *Svar o bākriti (voice and patterns of speech)*. Kolkata: Sahityalok, 2006.
- Bhattacharya, Subhas, *Bāṅglā ucchāraṇ abhidhān (Bengali Pronunciation Dictionary)*. Kolkata: Sahitya Sansad, 1992.
- Bhattacharya, Subhas, *Bāṅgālir Bhāṣā (The language of the Bengali people)*. Kolkata: Ananda Publishers, 2000.
- Biswas, Naren, "Bāṅglā bānān o ucchāraṇ: samasyā o samādhāner upāy (Bengali spelling and pronunciation: problems and ways for solution)", *Bāṅglābhāṣā mātṛbhāṣā: Aikatān Gabeṣanāpatra,* vol. 1, no. 1, pp. 79-86, 2005.
- Chakrabarti, Shailen, *Natun bāṅglā bānān abhidhān (New Bengali spelling dictionary)*. Kolkata: Deb Sahitya Kutir, 2008.
- Chattopadhyay Suniti Kumar, *Bāṅgālā bhāṣātattver bhumikā (Introduction to Bengali Linguistics)*. Kolkata: Calcutta University Press, 1974.
- Dan, Mina, "Sañjanani dvanisannibeś o bāṅglā ardhasvar [y] (generative phonotactics and Bengali semivowel 'y')", *Sahitya Parisat Patrika,* vol. 113, no. 3, pp. 34-50, 2006.
- Dan, Mina, "Mid-front semivowels in Bangla: a case study at the phonology-morphology interface". In, Dash, Niladri Sekhar, Probal Dasgupta and Pabitra Sarkar (Eds.) *Rainbow of Linguistics. Vol.-I.* Kolkata: T. Media Publications, pp. 117-128, 2007.
- Dash, Niladri Sekhar, *Corpus-Based Computational Analysis of the Bangla Language. Unpublished Doctoral Dissertation*, University of Calcutta, Kolkata, 2001.
- Dash, Niladri Sekhar, "Methods in madness of Bengali spelling: a corpus-based empirical investigation". *South Asian Language Review*, vol. 14, no. 2, pp. 63-92, 2005.
- Dash, Niladri Sekhar, *Bahurūpī bāṅglā bānān (multifaceted Bengali spelling)*. Kolkata: Dakshabharati. 2006.
- Dash, Niladri Sekhar, "Frequency-based analysis of words and morphemes in Bengali text corpus". *Indian Journal of Linguistics,* vol. 25, no. 26, pp. 223-253, 2007.
- Hai, Abdul, *Dhanibijñān o bāṅglā dhvanitattva (Phonetics and Bengali Phonology)*. 4th Reprint. Dhaka: Mullick Brothers, 1985.
- Haldar, Narayan, *Bāṅglā bhāṣā prasaṅga: bānān kathan likhanrīti (in the context of Bengali language: spelling, speech and writing system)*. Madanpur, Nadia: Shankari Prakashan, 2007.
- Haque, A.F.M Daniul, "Bāṅgālir bāṅglā ucchāraṇ saṃkrānta cintā: aitihāsik paryālocanā (the thoughts of Bengali about Bengali pronunciation: historical analysis)". *Proceedings of the Seminar of the Asiatic Society of Bangladesh*, 14-15 April, Dhaka, 1993.
- Haque, Enamul, *Bāṅglā bākdhvani : svarup o binyās (Bengali speech sounds: identity and distribution)*. Dhaka: Ayantika, 1995.

- Hossain, Kamrul A., *Sādhāraṇ dhanitattver upādān (Components of general phonology).* Dhaka: Bangla Academy, 1989.
- Hyman, Larry, *Phonology: theory and analysis*. New York: Holt, Rinehart and Winston, 1975.
- Mukhopadhyay, Ashoke, "Bāṅglāy ýuktabarṇa o kichu prāsaṅgik praśna (conjuncts in Bengali and some related questions)". *Bāṅglā Bai,* September Issue, pp. 1-2, 2005.
- Nath, Mrinal, "Duhkhini barṇamālā (unfortunate alphabet)". *Ebang Mushayera.* Vol. 11, no. 1, pp. 62-73, 2004.
- Pal, Palash Baran, *Dhvaniālā barṇamālā (Phonology and Alphabet).* Kolkata: Pyapirus, 2001.
- Ray, Punya Sloka, *Bengali Language Handbook.* Calcutta: Bangla Akademi, 1997.
- Sarkar, Pabitra, Amitabha Mukhopadhyay and Prasantakumar Dasgupta (Eds.), *Akādemi bānān abhidhān (Academy spelling dictionary).* 4th Edition. Kolkata: Bangla Akademi, 2003.
- Sarkar, Pabitra, *Bāṅglā bānān saṃskār: samasyā o sambhābanā (Bengali spelling reform: problems and possibilities).* Kolkata: Chirayata Prakasan, 1992.
- Sarkar, Pabitra, "Bāṅglā bhāṣār ýuktabyñjan (consonant conjuncts of Bengali language)". *Bhasha,* vol. 6, no. 1, pp. 23-45, 1993.
- Schane, S.A. *Generative phonology*. Chicago: Prentice Hall, 1972.