The Dhaka University Journal of Linguistics: Vol. 1 No.2 August, 2008
Page : 87-114, Published on August 2009 © Registrar, Dhaka University
ISSN-2075-309

Gemination in Bangla: An Optimality Theoretic Analysis

Somdev Kar
Senior Research Fellow, Central Institute of Indian Languages (CIIL), Mysore, India

Abstract
This paper attempts to identify and analyze different types of gemination processes in Bangla. The focus is mainly on the phonological representation of sound combinations which forms a set of valid geminates in this language. I argue for three major types of gemination processes present in modern Bangla and a stratification strategy for the relavant lexical items based on their origin (SB, NB and OB, depending on the native vs. two type of borrowings). An analysis of these gemination processes are given in the framework of optimality theory (OT). Therefore, the constraint-based analysis of OT is organized in a threefold argument structure for each stratum. The conclusion is drawn towards an understanding of gemination processes of Bangla for different categories of lexical items and their phonological formuations.

1. Introduction

In the last two decades, there has been a shift in focus in much of the studies on phonological theory, from rule-based system to sets of constraints on well-formedness principles, making way to the formation of optimality theory (McCarthy, 2001;

I am grateful to an anonymous reviewer for helping me strengthen the argumentation and for suggesting a detailed introduction to OT and its structure. I would also like to thank Prof. Hubert Truckenbrodt for his guidance in developing this work. Of course I am solely responsible for the remaining misunderstandings and lack of clarity, if any.

McCarthy & Prince, 1993a, 1993b; Paradis, 1988; Prince & Smolensky, 1997). This theory was developed as a response to a "conceptual crisis at the center of phonological thought" (Prince & Smolensky, 1993) concerning the role of output constraints. It was also (partly) inspired by the concepts of neural networks, as we can see the ideas of optimization, parallel evaluation, competition, and soft, conflicting constraints are familiar in this framework.

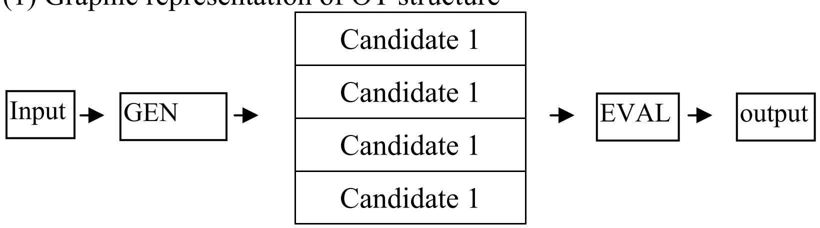
Optimality Theory (henceforth, OT) is often considered as a development of generative grammar and the successor of the harmonic grammar developed in 1990 by the trio Géraldine Legendre, Yoshiro Miyata and Paul Smolensky (Legendre, Miyata & Smolensky, 1990; Prince & Smolensky, 1993; Smolensky & Legendre, 2006 *a.o.*). In harmonic grammar, the maximal harmony is on the optimal candidate, where 'harmony' is calculated using a simple linear equation. Given a representation's scores on a set of constraints, and a set of coefficients, or weights, 'harmony' is the sum of the weighted constraint scores. From this structural framework, Alan Prince and Paul Smolensky developed the optimality theory where the constraints assign scores based on the number of violations by the candidates, then the scores are the corresponding negative integers, and the weights are positive reals (Pater, 2007). Now, OT is widely adopted by scholars not only in the area of phonology, where OT was initially developed and applied, but also in other areas of linguistic studies, such as in syntax and semantics (*see* Legendre, Grimshaw & Vikner, 2001).

2. Theoretical Framework

In a compact introduction, phonological (rather, grammatical) constraints are ranked and violable by the phonetic forms of their underlying representations in the OT structure. These constraints are minimally violated by a set of candidates (potential surface forms) and the one which incurs the least serious violations wins. The seriousness of a violation is defined in terms of hierarchies of constraints; the violations of higher-ranked constraints are most serious. An OT-style tableau using Harmony maximization as the criterion for optimality where the weights are in the top row and the rightmost column provides the harmony values for the candidates.

More elaborately, OT works in a constraint-based competition system among a possibly infinite set of candidates (at least two). In classical representation, the generation of utterances in the optimality theory involves two very important functions, viz., GEN and EVAL. From an input, GEN returns a set of unique output candidates. Among these candidates at least one could be identical to the input and the rest are somewhat modified in their structure. Then, EVAL functions to choose the optimal candidate that best satisfies a set of specially ranked constraints depending on the violation. That means, in OT the constraints are violable. The ranking process of the constraints is very crucial here, because it is the most important criterion that chooses the optimal candidate as output. EVAL chooses the out from a set of candidate starting from two to an infinite number (n). The figure in (1) bellow illustrates the process to reach an output from the input through the function of GEN and EVAL (Davenport & Hannahs, 2005).

(1) Graphic representation of OT structure



There are two types of constraints which act as EVAL: markedness and well-formedness constraints. Markedness constraints enforce well-formedness of the output candidate, prohibiting structures that are difficult to produce or comprehend, such as consonant clusters or phrases without overt heads (Arbib, 2002; Kager, 1999). These constraints usually prohibit some phenomenon or impose restrictions on the occurrence of certain segments. For instance,

a. Syllables must not have codas
b. Syllables must have onsets
c. Obstruents at coda position must not be voiced

On the other hand, faithfulness constraints enforce similarity between input and output. For instance, all morphosyntactic features

in the input to be overtly realized in the output. Kager (1999) lists some typical instances of faithfulness constraints that are available in most languages:

a. The output must present all segments present in the input
a. The output must present the leaner order of segments in the input
b. Output segments must have counterparts in the input

These constraints, both markedness and faithfulness, are ranked in a planned order. In an analysis, different markedness and faithfulness constraints usually do conflict, so the ranking of the constraints decides the right candidate as the output depending on the violation (of constraints) pattern. This ranking of the constraints is not a strict universal ranking; rather it differs from language to language. In other words, different languages have their own constraint ranking which applies to that particular language only. But, for every language, the constraint ranking is very strict. That means a candidate violating a high-ranking constraint can never be a winner by satisfying lower-ranked. Here, the other important issue is the violability of constraints. Violability ensures that the optimal candidate is not required to satisfy all constraints. It may violate a constraint and still win as the optimal candidate, if it satisfies the top-ranked candidate(s). Another way of describing EVAL is that a candidate x is optimal if and only if, for any constraint that prefers another candidate y to x, there is a higher-ranked constraint that prefers x to y (*see* Zuraw, 2003). So, the main point in this section is that OT allows the specification of a ranking among the constraints and allows lower ranked constraints to be violated in order for higher ranked constraints to be satisfied.

The constraint ranking and their interaction among the input candidates is typically showed in tableaux in the OT analysis. In such a tableau, the candidates are listed vertically while the constraints are ranked in the horizontal line. For a hypothetical language X, let us assume that constraint 1 (con_1) is satisfied by candidate 1 and on the other hand, candidate 2 satisfies constraint 2 (con_2). In the following tableau, it is shown how the optimal candidate is being chosen by EVAL through the constraint interaction.

(2) Constraint ranking for language X

input	con_1	con_2
☞ 1. $cand_1$		*
2. $cand_2$	*!	

In (2), con_1 is ranked above con_2 and hence, it is violated by the candidate 2. This is a fatal violation and for this candidate, because of the high ranking of con_1. As a result, candidate 1 becomes the optimal candidate, even after violating the low ranked con_2. A hand symbol is used in OT to indicate the optimal candidate in a tableau (see the above tableau).

Optimality theory examines several restrictions available in the phonological processes in a language and relevant constraints are formed to account for those restrictions. Syllable structure is one of the prominent topics in the research activities in the OT framework. In recent time, Féry & van de Vijver (2003) presented a collection of studies in this topic that opens up new ways of further research on several issues of syllable structure. But, so far OT is comparatively a less preferred methodology among researchers who work in Bangla phonology. There are only a few works (Das, 2002; Kar, 2009; Kothari, 2004; Vijayakrishnan, 2003 *inter alia*) done in different phonological studies in Bangla using this theoretical framework. So, there is always a lack of literature for anyone who would study this issue in Bangla. But, on the other hand, OT itself is a very innovative methodology and it is always a challenging matter to deal with.

OT is used in a large set of phonological analysis which includes certain types of sound changes as well. Bangla registers a significant amount of sound changes taking place in different types of lexical elements, mostly in among the borrowed words. A large number of Sanskrit words are integrated in Bangla lexicon over a long period of time. These words typically belong to the SB[1] stratum, of a three level stratification of the Bangla lexicon. The other two strata are the NB[2] and OB[3]. The syllabic pattern of many of these words is

changed after their integration in the Bangla vocabulary. There are different types of sound changes that are available throughout the process of borrowing from other languages including Sanskrit. One of such changes is the gemination process, where a consonantal sound reduplicates inside a word and sounds audibly longer period of time compared to its original form. Interestingly, these words retain their original orthographic mapping in the target language even after the phonological changes in the spoken form.

3. Types of Gemination

A corpus study (of Bangla lexicon) presents various types of geminates available in the Bangla vocabulary. Earlier investigations show that only a few languages allow all types of geminates, but actually there is a significant amount of work still needs to be done in the area of gemination. In this regard, some remarkable work is done in the case of voiced obstruent geminates. These studies reveal that voiced obstruent geminates are cross-linguistically disfavored, because, with long closure, it is difficult to maintain a transglottal air pressure drop sufficient to produce voicing (Hayes & Steriade, 2004; Jaeger, 1978; Kawahara, 2005; Ohala, 1983; Taylor, 1985). For instance, Luganda[4] has no nongeminate coda present in the whole lexicon (*see* McCarthy, 2003).

In a C_1C_2 structure, where C denotes a consonantal sound, the loss (or merger) of C_2 in the geminated form is subsidized as the former consonant (C_1) reduplicates to compensate the loss (or merger) of the following sound. In other cases, the structure could be C_1V_2 as well, where V_2 denotes the presence of a semi-vowel (in general, V denotes a non-consonantal sound), assuming that this cluster is located word-medially and hence V_1 is taken as the peak of the first syllable. There is a third type of gemination found in Bangla lexicon where no sound is dropped or merged, but the first consonant is reduplicated, keeping the following sound intact. This type of gemination is seen when an obstruent is followed by a liquid (/r/ or /l/) sound in the SB stratum words. Incidentally, all the above said combinations are found in Bangla vocabulary. But, depending on the sound preceded by the consonant C_1, the geminates available in Bangla could be divided into three categories as shown in (3) in the following. All the geminate cases are available in the word-medial position, where, in Sanskrit, the following conditions prevail.

(3) Gemination process in Bangla (SB) from Sanskrit

i. Obstruents followed by an semi-vowel
(C_1C_2>C_1C_2, where C_1=Obstruent and C_2=Semi-vowel)

ii. Plosives followed by a nasal /m/
(C_1C_2>C_1C_2, where C_1=Plosive and C_2 =/m/)

iii. Obstruents followed by a liquid
(C_1C_2> $C_1C_1C_2$, where C_1=Obstruent and C_2= liquid)

3.1. Gemination with semi-vowels

The very first type in (3) represents a significant number of SB words which lost their post-consonantal semi-vowels in the spoken form of Bangla (SCB or Standard Colloqual Bangla, in particular) when borrowed from Sanskrit. Note that, such changes are not visible in some other Indian languages such as Hindi, Gujarati, Marathi etc., which are, like Bangla, also closely related to Sanskrit. These languages typically retain the original phonological forms of Sanskrit words even after the bowing process is over. Sanskrit has two distinct semi-vowels (also called glides in some earlier literature): labio-dental semi-vowel [ʋ] and palatal semi-vowel [j]. Below is a set of examples where type (i) gemination process takes place with labio-dental [ʋ]. This observation is also supported by several literary works (*see* Chatterji, 1926a, 1926b, 1988; Dey, 1979; Singh, 1980 *a.o*).

(4) Gemination: Consonant followed by labiodental semi-vowel in Bangla (SB)

a. *San.* साध्वी /sad^h.ʋi/ > *Ban.* mvaŸx [sad^h.d^{h}i] ‘faithful wife’

b. *San.* पृथ्वी /pṛt^h.ʋi/ > *Ban.* c„_x [prit^h.t^{h}i] ‘earth’ (as in *‘prithviraj’*)

c. *San.* विश्वास /bis.ʋaś/ > *Ban* wekvm [biʃ.ʃaʃ] ‘faith/belief/trust’

(5) Gemination: Consonant followed by labiodental semi-vowel in Bangla (SB)

a. *San.* सत्य [satya/ > *Ban.* mZ¨ [ʃot.to] ‘truth’

b. *San.* बाल्यकाल /bālyakāl/ > *Ban* evj¨Kvj [bal.lo.kal] ‘childhood’

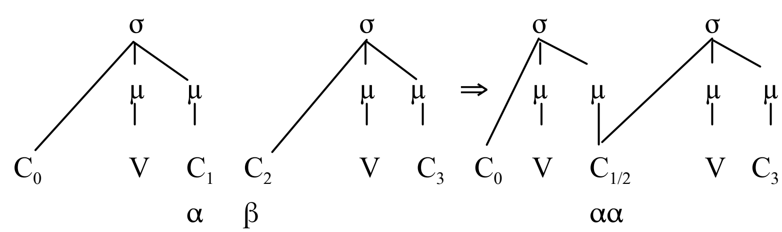
c. *San.* मृत्यु /mṛtyu/ > *Ban.* g„Zz¨[mrit.tu] ‘death’

d. *San.* बश्यता /bashyatā/ > *Ban.* ek¨Zv [bɔʃ.ʃo.ta] ‘domesticize’

e. *San.* पुण्य /punya/ > *Ban.* cY¨ [pun.no] ‘virtue’

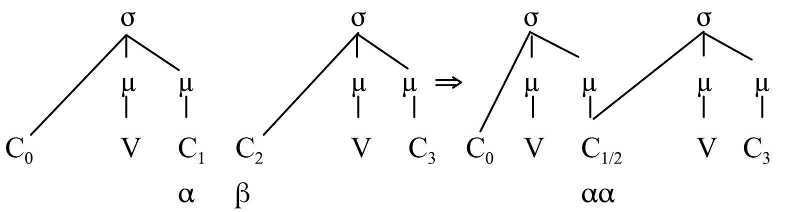
There are two possible gemination processes for this type of words that could be explained in this study. These are either a merger of C_1 and C_2 into C_1 or the deletion of C_2. It is generally assumed that the deletion process is not favored in many languages. We will consider both possibilities in the present study and see what be the outcomes. The above mentioned gemination (semi-vowel) cases could be structurally illustrated as follows. First, the more favored merger is shown in (6) below.

(6) Geminate formation 1: Merger



In this gemination process, C_1 is merged with C_2 and constitutes the geminated form C_1C_1 which is shown as $C_{1/2}$ in the above illustration. That means the merged place feature is shared by the coda of the first syllable and the onset of the following syllable. This formation would give a [C_1C_1] geminate in the surface structure. Another process of gemination is illustrated bellow where the coda of a syllable shares its position with the onset of the following syllable after the deletion of the original value of that very onset. This is not a case of merger, rather an issue where a segment of the syllable is deleted to make way for a geminate.

(7) Geminate formation 2: Deletion



As we can see in (7), the onset (C_2) of the second syllable is dropped and the coda (C_1) is shared at that position (onset). In general this type of gemination is not favored in many languages. In the OT framework, there could be a constraint favoring this type of gemination as it shows coalescence in the syllable structure. The faithfulness constraint UNIFORMITY (*see* McCarthy & Prince, 1995; *cf.* Causley, 1997; Gnanadesikan, 1995, 1997; Keer, 1999; Lamontagne & Rice, 1995; McCarthy, 2000; Pater, 1999) could be used in order to dispose of certain codas by coalescence with a following consonant (violation of UNIFORMITY). Since this constraint blocks any merger in the syllable or cross-syllabically (but allows deletion), it must be ranked lower than markedness for gemination effect to take place.

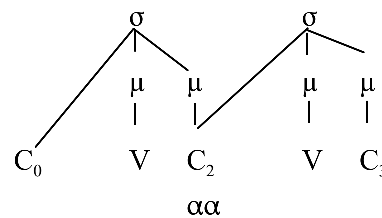
(8) UNIFORMITY (McCarthy & Prince, 1995)
No element of the output has multiple correspondents in the input. (No coalescence)

The constraint UNIFORMITY punishes the mapping of a pair geminate onto a single segment. This constraint punishes any candidate where two elements are merged (fusion) into one from the input form and form a geminate. In the following figure, this type of fusion occurred where the nearby consonants C_1 and C_2 are merged as $C_{1/2}$ in the later stage. UNIFORMITY does not like this type of merger though.

A constraint prohibiting any kind of gemination in syllable is used in some previous studies (*see* Baković, 2005; Hall, 2003; Ham, 1998; Murray & Vennemann, 1983; Shinohara, 2002). In many literary works it is called NOGEM; we will use the same name in the present study as well. This constraint requires every root node to be linked only to a mora and not to any onset. This concept is illustrated

below.

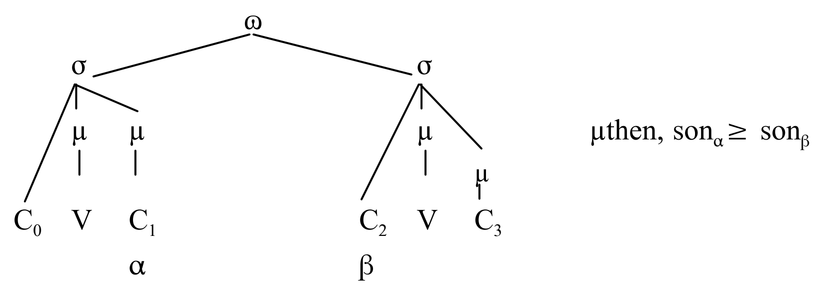
(9) No gemination structure



(10) NOGEM (Hall, 2003)
No multiple links from a root node to a higher tier.

The well known syllable contact law is a relevant issue in the current section of this work, as it maintains the sonority of the syllable edge intact across the syllable boundary (Davis, 1998; Gouskova, 2002; Hooper, 1976; Murray & Vennemann, 1983; Rose, 2000; Vennemann, 1988). In a disyllabic word, if the coda of the first syllable is α and the onset of the second syllable is β, then the sonority of β must not be greater than α. This law is illustrated in the following moraic structure in (11) and could be represented as a constraint to account for a possible sonority change (rise) from coda to the following onset (12).

(11) Syllable contact principle



(12) SYLCONTACT (Gouskova, 2002)
Sonority must not rise across a syllable boundary.

McCarthy & Prince (1995) argued for the anti-epenthesis constraint Dependency-IO (DEP-IO) which says that nothing should be inserted in the output segments that does not correspond to the input segments. That means, there should not be any epenthesis in the output candidates. On the other hand, MAX-IO requires all the input segments to be preserved in the output candidates. Hence, no deletion is allowed in the output segments.

(13) DEP-IO (McCarthy & Prince, 1995)

Output segments must have input segments.

(14) MAX-IO (Kager, 1999)

Input segments must have output correspondence.

Now, let us consider (4c) िवश्ास /bis.ʋaś/ [biʃ.ʋaʃ] 'faith/belief/trust' from the example set to account for an OT analysis based on the above mentioned gemination structures. Let us mark the consonant-semivowel cluster /$ʃ_1ʋ_2$/ as /$ʃ_αʋ_β$/ for the simplicity of explanation.

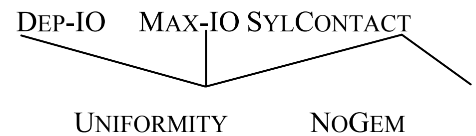
(15) Gemination (SB): with a semi-vowel as C_2

/$biʃ_αʋ_βaʃ$/	DEP-IO	MAX-IO	SYLCONTACT	UNIFORMITY	NOGEM
1. $biʃ_α.ʋ_βaʃ$			*!		
2. $bi.ʃ_αa.ʋ_βaʃ$	*!				
☞3. $bi_μʃ_{α/β}aʃ$				*	*
4. $bi_μʃ_αaʃ$		*!		*	
5. $bi.ʃ_1aʃ$		*!			

In this analysis, there are two different geminates contesting to win. The above tableau has candidate 3 and 4 with geminated forms which violate the NOGEM constraint once each. Additionally,

candidate 3 violates the no coalescence constraint UNIFORMITY for the merger of α and β. But these violations didnot rule out this candidate, since UNIFORMITY is ranked lowest in the tableau. But, candidate 4 is ruled out by the violation of the faithfulness constraint MAX-IO, and so does candidate 5. An insertion is also not tolerated in the case of candidate 2. The input form (candidate 1) has a fatal violation of constraint SYLCONTACT for the rising sonority of the β place from α. The winning candidate /$bi_μʃ_{α/β}aʃ$/ (ıekvm) [biʃʃaʃ]) is a geminated form of the corresponding non-geminated Sanskrit word and an acceptable phonological word-form in Bangla vocabulary. The constraint ranking of this OT account would be as follows.

(16) Constraint ranking for gemination (obstruent-semivowel)



The first three of the five conflicting constraints are mutually ranked while they dominate the last two gemination-related constraints NOGEM and UNIFORMITY. Again, the low ranking anti-gemination constraints NOGEM is dominated by this faithfulness constraint.

3.2. Gemination of nasal

A corpus study[5] of Bangla lexicon (corpus size: 1.6 million words, with repetition of same entries), it is evident that the non-coronal nasal /m/ occurs restrictively after a plosive. A closer investigation in the data collected from the corpus reveals that most of these cases have more than zero instances belong to a particular stratum of Bangla lexicon, viz., OB. So, NB and SB class words following this restriction more faithfully, but OB does not. Let us have a look in the following illustration of such clusters where the frequency of the clusters is shown according to their respective stratum. Below, the X mark denotes that the frequencies of the corresponding clusters are zero and the tick (✓) mark is for frequencies between one and ten.

(17) Frequency of plosive-/m/ clusters

Cluster	NB	SB	OB	Cluster	NB	SB	OB

/pm/	X	X	✓	/pʰm/	X	X	X
/tm/	X	X	X	/tʰm/	X	X	X
/ʈm/	X	X	X	/ʈʰm/	X	X	X
/km/	X	✓	X	/kʰm/	X	X	X
/bm/	X	X	X	/bʰm/	X	X	X
/dm/	X	X	✓	/dʰm/	X	X	X
/ɖm/	X	X	✓	/ɖʰm/	X	X	X
/gm/	X	✓	✓	/gʰm/	X	X	X

So, it is clear from the above illustration, that plosives are not followed by /m/ for maximum number of entries in the NB and SB strata. But, two SB strata do record a nominal number of such cases, which will be taken into the account later in this section. We can draw the general plosive-/m/ restriction in terms of a constraint prohibiting any plosive to occur before /m/ in a word-medial cluster in Bangla, which will be applied only to the NB and SB stratum.

(18) *Plosive-/m/ (Kar, 2009)

No plosive is followed by nasal /m/.

In Sanskrit, there are numerous instances where a plosive precedes /m/ at the word-medial position. In general, many such words in Bangla are borrowed from Sanskrit (hence, belong to the SB stratum). These words project an interesting behavior when integrated in the B-lexicon. They always form a geminate of the first element of the cluster, by either merging or dropping the second element. That means, C_1C_2 becomes C_1C_1 where C_1 is the plosive and C_2 is /m/. For instance,

(19) Gemination: Consonant followed by /m/ (SB)

a. *San.* पद्म /padma/ > *Bang.* cÙ /padma/ [pɔd.do] 'lotus'

b. *San.* छद्म /cʰadma/ > *Bang.* QÙ /cʰadma/ [cʰɔd.do] 'disguise'

c. *San.* आत्मा /atma/ > *Bang.* AvZ¥ /atma/ [at.ta] 'soul'

These words retain the [dm] or [tm] sound sequesnce in some other NIA languages like Hindi, Marathi etc. and of course in Sanskrit.

But, in Bangla, the sequence is changed through a gemination process. Let us take an example from the above set of SB words: (19c) AvZ¥ /ātmā/ [ãtta] 'soul'. It is presented in the format of C_1C_2 (that is, $/t_1m_2/$) in the following tableau.

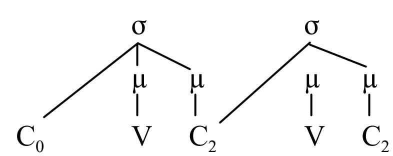
(20) No plosive + /m/ cluster (SB case)

$/at_1m_2a/$	DEP-IO	MAX-IO	*Plosive-/m/	UNIFORMITY
1. at_1m_2a			*!	
☞ 2. $a\mu t_{1/2}a$				*
3. at_1a-m_2a	*!			
4. $a\mu t_1a$		*!		
5. at_1a		*!		

In the above tableau, the markedness constraint *Plosive-/m/ rules out candidate 1 for the existence of an unwanted cluster /tm/ in the candidate. Candidates 3, 4 and 5 fatally violate faithfulness constraints either by adding sounds or deleting any part of the input. The lowest ranked constraint Uniformity is violated by candidate 2. But, this violation is the weakest among all the candidates. So, this geminated form wins.

The μ symbol in (20) represents the place sharing feature of the geminate structure as illustrated in the following figure.

(21) Geminate structure (moraic)



Now, in the presentation of the frequency of occurrences, very few cases of velar-nasal /m/ (/km/ and /gm/) clusters occur in the SB stratum. For instance, hyM¥ /jugma/ [ʤugmo] 'joint', evMx /bagmi/ [bagmi] 'speaker', i"¹´bx /rukmini/ [rukmini] 'a proper noun' etc do not allow any change in the medial cluster. A context sensitive markedness constraint requiring the nasal /m/ to preserve its nasality after a velar sound could be introduced here to account for such cases in Bangla.

(22) IDENT-NAS(vel, -m)
Input and output of a velar-nasal cluster must agree in nasality, if the velar is followed by a nasal /m/.

This constraint punishes any deviation in the [nasal] feature in the output candidate, when the input contains a velar-/m/ cluster. Let us consider the above mention SB case: evMœx /bagmi/ [bagmi] 'speaker'
(23) Velar (plosive)+/m/ cluster (SB case)

/bag$_1$m$_2$i/	DEP-IO	MAX-IO	IDENT-NAS(vel, -m)	*Plosive -/m/	UNIFORMITY
☞ 1. bag$_1$m$_2$i				*	
2. baμg$_{1/2}$a			*!		*
3. bag$_1$a-m$_2$i	*!				
4. baμg$_1$a		*!			
5. bag$_1$a		*!			

In the above tableau, candidates 3, 4 and 5 violate the highest ranking faithfulness constraints Dep-IO and Max-IO. For the input candidate (cand 1) to win, the context sensitive constraint IDENT-NAS(vel, -m) needs to be ranked above *Plosive-/m/ and Uniformity. This ranking would rule out the gemination case (candidate 2) for the violation of the nasality ([gm]>*[gg]) in this velar-nasal cluster.

Hence, the input candidate wins for the least amount of violation (only *Plosive-/m/). There would be a strict constraint ranking for the above mentioned criteria from NB and SB strata of the B-lexicon, keeping the exceptional cases of SB (velar-nasal).

Now, let us consider a word from the OB stratum: ZKgv /takma/ [tɔkma] 'marking'. This is a borrowed word from Persian. The OT analysis for this word in illustrated in the following tableau. The ranking of the markedness constraints, UNIFORMITY and *Plosive-/m/ is changed here to respect the occurrence of /km/ cluster in the OB stratum. Since, OB stratum does not have any restriction in the occurrence of /m/ after plosives; the relevant constraint prohibiting such occurrences is ranked lowest in the above tableau.
(24) No plosive + /m/ cluster (OB case)

/tɔk$_1$m$_2$a/	DEP-IO	MAX-IO	IDENT-NAS (vel, -m)	*Plosive -/m/	UNIFORMITY
☞ 1. tɔk$_1$m$_2$a				*	
2. tɔμk$_{1/2}$a			*!		*
3. tɔk$_1$a-m$_2$a	*!				
4. tɔμk$_1$a		*!			
5. tɔk$_1$a		*!			

So, the previous winner (cand 2) is ruled out in this tableau by a fatal violation of the anti-merger (no-gemination) constraint UNIFORMITY, which is ranked comparatively higher. The input candidate (cand 1) wins, because of the low ranking of *Plosive-/m/.

The conclusion of this section is that the plosives are followed by /m/ in the OB stratum, but prohibited in NB and SB stratum in Bangla. But, they need different constraint rankings (of the same set of constraint) for the respective stratum.

3.3. Gemination with liquids

In Bangla vocabulary, voiced plosives geminate before a liquid (/r/ or /l/), keeping the latter sound intact. That means the adjacent consonants C1 and C2 forms a geminate as C1C1C2. But, this rule actually extends throughout the obstruent column, not limited only to voiced obstruents. The said corpus study of Bangla shows that word initial consonant clusters with a liquid sound at the C2 position, only except one case (/ʈʰr/), all the clusters have more than zero occurrences in the Bangla lexicon. That means the obstruent-liquid (/r/) clusters occur in geminated form in all possible cases in this category. On the other hand, the (c) column records instances only for unaspirated obstruents followed by /l/. As mentioned in the previous chapter, such sound combinations where $C_1C_1C_2$ type gemination occurs are found only in the SB and OB stratum of the B-lexicon. The NB words do not allow any gemination of this sort. Actually, there are a very few words present in the NB stratum that have such sound combinations and when the gemination issue is relevant, they refuse to change their structure.

Orthographically, all these combinations are represented in the written form of Bangla using "*phala*", a distinct feature of Bangla orthography. While used as r-phala, the obstruent and the following /r/ sound creates a ligature so that the regular form of the /r/ sound in Bangla is replaced by a special r-phala symbol. For instance, para = ci, but, pra = c. Sarkar (2006) includes also "l-phala" in this type of gemination, though formally there is no evidence of "l-phala" in Bangla orthography. Let us list down some examples of this type of cluster, both from SB and OB strata.

(25) Gemination: Obstruent followed by liquid in Bangla

I. Sanskrit Borrowings (SB)

a. *Ban.* cy©/putra/ [put.tro] 'son'(<Sanskrit)

b. *Ban.* ïå /subʰra/ [ʃubʰ.bʰro] 'white/bright' (<Sanskrit)

c. *Ban.* AvcøZ /apluta/ [ap.pluto] 'inundated' (<Sanskrit)

II. Other Borrowings (OB)

d. *Ban.* mycg /suprim/ [ʃup.prim] 'supreme' (<English)

e. *Ban.* mvcøvB /saplai/ [ʃap.plai̯] 'suply' (<English)

f. *Ban.* gv`ªvmv /madrasa/ [mad.dra.ʃa] '(Islamic) school' (<Arabic)

It should be noted that this type of gemination occurs only when the adjacent consonants (αβ) are preceded by a vowel. Thus, 'Cαβ' clusters do not show such a gemination in Bangla. For instance, B‡jKwUK /ilekṭrik/ [ilek.ʈrik] 'electric' does not register such gemination at the /ʈr/ cluster, even though it belongs to the OB stratum. On the other hand, it is already mentioned that words from NB category strictly do not follow this gemination pattern at all.

(26) Word-medial obstruent-liquid clusters belonging to NB stratum

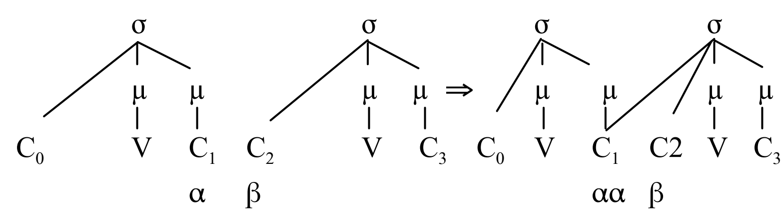
a. *Ban.* mvZiv /sãtra/ [ʃãt.ra] 'a Bengali surname'

b. *Ban.* evewi /babri/ [bab.ri] 'long curing hair-style'

c. *Ban.* kvcjv /sapla/ [ʃap.la] 'water lily'

In the following figure (27), at the gemination stage, $C_2(\beta)$ remains untouched, but $C_1(\alpha)$ is being shared by the coda of the first syllable and the first segment of the complex onset of the following syllable. SB and OB class words fall in this structure, but NB words do not allow any complex coda.

(27) Geminate formation 3: WGG



This type of gemination has a close resemblance with a very well known process of language change in Germanic language family called West Germanic Gemination (WGG). As claimed in Hall (2003), that a post-short vowel consonant is a geminate in West Germanic languages before the palatal semi-vowel (glide) /j/. The said gemination process is attested in some West Germanic languages, such as Old English (OE), Old High German (OHG), Old Saxon (OS), whereas it is absent in North Germanic (=Old Norse (ON)) or in East Germanic (=Gothic (Got.)). A set of data showing WGG (taken from Hall, 2003 and Simmler, 1974) is illustrated below.

(28) West Germanic Gemination (WGG) data (from Simmler, 1974)

East/North Germanic	**West Germanic**	**Gloss**	**Geminate**
a. *Got.* Skapjan	*OS* skeppian, *OE* scieppan	'to create'	[pp]
b. *Got.* hafjan	*OHG heffan, OE hebban*	'to lift'	[ff]
c. *ON* framja	*OHG fremmen, OE fremman*	*'to carry out'*	[mm]
d. *Got.* halja	*OS hellia, OHG hell*a	'hell'	[ll]

The contrast between the nongeminate forms in the East or North Germanic languages (first column) and the corresponding cases with geminates in the West Germanic laguages (second column) is claimed to be a historical development. It is postulated that in Early West Germanic languages, all of the above listed cases (and many more) were originally VCjV (where j is the palatal glide) and that a historical process of the form VCjV > VCCjV occurred before the West Germanic daughter languages emerged from this family.

But, one notable difference is that, in WGG, mostly voiceless obstruents are geminated when preceded by a liquid (Gaeta, 2001), while Bangla typically allows such geminates for all obstruents irrespective of their voicing. This type of sound change could be explained by the so called Preference Laws for Syllable structure (Murray & Vennemann, 1983; *cf.* Vennemann, 1988). The idea of this type of sound change was determined by the influence of marked

syllable structure as a consequence of the Germanic preference of the bimoraic stems as asserted by Murray and Vennemann. Bimoraic stem structure is also preferred by Bangla syllables. Hence, a similar mechanism may also be applied to analyze the said gemination process in the present work.

In order to omit a consonant or vowel which is not favored in a syllable structure, two faithfulness constraints could be introduced here. Both of these constraints belong to the same family of Max.

(29) Max-IO-C
[+con][6] sounds in the input must have output correspondence.
(30) Max-IO-V
[-con] sounds in the input must have output correspondence.

The faithfulness constraint MAX-IO-C requires every consonant of the input form to be preserved in the output. On the other hand, MAX-IO-V wants the same for every vowel in the input. These Max constraints will replace the general MAX-IO constraint which is used in this study so far. Additionally, we will not consider DEP-IO in this section (and thereafter), since it is not so crucial in the improvement of the analysis. Instead, we would introduce the complex onset related markedness constraint *COMPLEX^{ONS}. This constraint does not prefer consonant clusters at the onset position.

(31) *COMPLEX^{ONS} (Kager, 1999)
Onsets are simple.

Let us take a word from (25a) *Ban.* cŷ /putra/ [put.tro] 'son', which belongs to the SB stratum of Bangla vocabulary, with the same set of constraints from the earlier section. In the following table (32) the input form (cand 1) violated the SYLCONTACT constraint by allowing a word-medial cluster [tr] which registers a rising sonority pattern. This constraint punishes such phonological feature across syllable boundaries.

(32) Gemination (SB): with a liquid as C_2

[put$_1$r$_2$o]	MAX-IO-C	SYLCONTACT	*COMPLEX^{ONS}	MAX-IO-V	NOGEM
1. put$_1$.r$_2$o		*!			
☞ 2. pu$_\mu$t$_1$r$_2$o			*		*
3. pu$_\mu$t$_1$o	*!				*
4. put$_1$o	*!				

The winning candidate (cand 3: [pu$_\mu$t$_1$r$_2$o]) satisfies top ranked constraints MAX-IO-C and SYLCONTACT, but fails to satisfy *COMPLEX^{ONS} and the anti-gemination constraint NOGEM. But, the lower ranking of these constraints could not rule out candidate 3. The last two candidates simply violate the faithfulness constraint MAX-IO-C while candidate 3 violates another faithfulness constraint NOGEM. This explains the double linking issue with a liquid sound in the cluster. Now, the modified consonant ranking for gemination in Bangla would be as followed.

(33) Constraint ranking for gemination (obstruent-liquid)
MAX-IO-C, SYLCONTACT » *COMPLEX^{ONS} » MAX-IO-V » NOGEM

This constraint ranking, however, cannot account for NB category words. Hence, the constraint ranking should be changed for this stratum. Let us take a word of such a cluster from the NB stratum:

(26c) kvcjv /sapla/ [ʃap.la] ‘water lily’.

(34) No gemination (NB): with a liquid as C_2

[ʃap$_1$l$_2$a]	MAX-IO-C	*COMPLEX^{ONS}	MAX-IO-V	NOGEM	SYLCONTACT
☞ 1. ʃap$_1$.l$_2$a					*!
2. ʃa$_\mu$p$_1$l$_2$a		*!		*	
3. ʃa$_\mu$p$_1$a	*!			*	
4. ʃap$_1$a	*				

The input is the optimal candidate in the above tableau for its least amount of violation (only the lowest ranked SYLCONTACT). The first four constraints MAX-IO-C, *COMPLEX^{ONS}, MAX-IO-V and NOGEM, eventually rule out all other candidates for fatal violations. The major change in this tableau from the last one is in the ranking of the constraints. The crucially ranked *COMPLEX^{ONS} and NoGem in the previous tableau no more maintain that ranking. It is now ranked with the other Faith (Max-IO-C, MAX-IO-V) constraints. On the other hand, SYLCONTACT is now ranked crucially with the other four constraints, which was mutually ranked with Faith in the earlier tableau.

(35) Constraint ranking for gemination (obstruent-liquid): NB
MAX-IO-C, *COMPLEX^{ONS}, MAX-IO-V, NOGEM » SYLCONTACT

Conclusion

The OT analysis for the gemination process given in this study covers almost all possible gemination cases in the Bangla vocabulary. One must be careful about the stratification in Bangla lexicon, because, gemination may or may not occur in certain clusters, depending on their stratum. For instance, the constraint ranking for plosive-nasal restrictions that applies to NB and SB strata of the Bangla lexicon, does not apply on the OB stratum.

(36) Constraint ranking for plosive-nasal restriction: NB and SB strata

DEP-IO, MAX-IO, IDENT-NAS(vel, -m) » *Plosive-/m/ » UNIFORMITY

The three top-ranked faithfulness constraints are not crucially ranked relative to the other two constraints in this analysis. These constraints dominate the last two markedness constraints. But, this ranking is valid only for NB and SB strata. If the same constraints are applied in case of an instance from the OB stratum, the ranking would be changed in the latter case. These restrictions lead to a stratified analysis of the gemination issue in Bangla. From the entire discussion, we can draw a factorial typology with the relevant constraints, taking DEP-IO and MAX-IO in one group: FAITH.

(37) Rankings for strata:

a. FAITH, SYLCONTACT » *COMPLEX^{ONS}, NOGEM
 'Gemination in Obstruent-semivowel cluster (SB, OB)'

b. FAITH, SYLCONTACT » *COMPLEX^{ONS} » NOGEM
 'Gemination in Obstruent-liquid cluster (SB, OB)'

C. FAITH, *COMPLEX^{ONS}, NOGEM » SYLCONTACT
'Gemination in Obstruent-liquid cluster (NB)'

It is widely believed that OT can give a fairly strong and stable account for a structural analysis in a language. In this work, I tried to focus on the gemination patterns in Bangla and then analyze them in terms of optimality theory. The outcome of the study is quite neat.

But there are still some areas that need more attention in future. The OT approach to gemination in Bangla opens new possibilities to cultivate this area from different angles. This approach can also be used in the study of some other phonological issues in Bangla, mainly those which are analyzed in the generative model of linguistics.

Note

1. **SB:** Sanskrit Borrowing (sords directy borrowed from Sanskurti)
2. **NB:** Native Bangla (native Bangla words and words indirectly borrowed from Sanskrit, i.e., through MIA languages, such as Pali, Prakrit etc.)
3. **OB:** Other Borrowing (words borrowed from other Indian and foreign languages, such as Tamil, Gajarati, Santali, Enlish, Portuguese, Persian, French etc.)
4. Luganda belongs to the Bantu branch of the Niger-Congo language family. It is spoken by more than three million people in Uganda (Africa).
5. The corpus used in this study was developed under the TDIL program of the erstwhile DoE (Dept of Electronics, Govt. of India, *now* Ministry of Communication and Information Technology). Subsequently, the corpora were passed on to the Central Institute of Indian Languages (CIIL), Mysore and have been in their custody since then. This is available in Indian Standard Code for Information Interchange or Indian Script Code for Information Interchange (ISCII) format. However, a Romanized version of the same is used here (thanks to Prof. Gautam Sengupta and Dr. Soma Paul), with some modifications.
6. [+/-con] denotes the presence or absence of a consonantal sound.

Reference:

Arbib, M. A. (Ed.). 2002. *The Handbook of Brain Theory and Neural Networks* (2 ed.). Cambridge, MA: MIT Press.

Baković, E. 2005. Antigemination, assimilation and the determination of identity. *Phonology, 22*, 279-315.

Causley, T. 1997. *Identity and featural correspondence: the Athapaskan case.* Paper presented at the NRLS27, Amherst, Massachusetts.

Chatterji, S. K. 1926a. *The origin and development of the Bengali*

language (Vol. 1). Calcutta: University of Calcutta.
Chatterji, S. K. 1926b. *The origin and development of the Bengali language* (Vol. 2). Calcutta: University of Calcutta.
Chatterji, S. K. 1988. *Bhasa prakash bangla vyakaran*. New Delhi: Rupa.
Das, S. 2002. *Some aspects of the phonology of Tripura Bangla and Tripura Bangla English.* Unpublished PhD dessertation, Central Institute of English and Foreign Languages (CIEFL), Hyderabad.
Davenport, M. & Hannahs, S. J. 2005. *Introducing phonetics and phonology* (2 ed.). London: Hodder Arnold.
Davis, S. 1998. Syllable contact in optimality theory. *Journal of Korean Linguistics, 23*, 181-211.
Dey, P. 1979. On rule ordering in Bengali phonology. *Indian Linguistics, 40*(1), 24-34.
Féry, C. & van de Vijver, R. (Eds.). 2003. *The syllable in optimality theory*. Cambridge: Cambridge University Press.
Gaeta, L. 2001. Striving for optimality. In K. Dziubalska-Kołaczyk (Ed.), *Constraints and Preferences* (pp. 101-142). Berlin: Walter de Gruyter.
Gnanadesikan, A. (1995). *Markedness and faithfulness constraints in child phonology*.Unpublished manuscript, University of Massachusetts.
Gnanadesikan, A. 1997. *Phonology with ternary scale.* Unpublished PhD, University of Massachusetts, Amherst.
Gouskova, M. 2002. Falling sonority onsets, loanwords, and Syllable Contact. In M. Andronis, C. Ball, H. Elston & S. Neuvel (Eds.), *CLS 37: the main session. Papers from the 37th meeting of the Chicago Linguistic Society* (pp. 175-186). Chicago: The Chicago Linguistics Society.
Hall, T. A. 2003. *On the nongemination of /r/ in West Germanic twenty years later.* Paper presented at the Old World Conference in Phonology I, Leiden.
Ham, W. 1998. A new approach to an old problem: gemination and constraint reranking in West Germanic. *The Journal of Comparative Germanic Linguistics, 1*, 225-262.
Hayes, B. & Steriade, D. 2004. Introduction: the phonetic bases of phonological markedness. In B. Hayes, R. Kirchner & D. Steriade (Eds.),

Phonetically-Based Phonology (pp. 1-33). Cambridge: Cambridge University Press.
Hooper, J. 1976. *An introduction to natural generative phonology*. New York: Academic Press.
Jaeger, J. 1978. *Speech aerodynamics and phonological universals.* Paper presented at the Proceedings of the Chicago Linguistics Society, Chicago.
Kager, R. 1999. *Optimality theory*. Cambridge: Cambridge University Press.
Kar, S. 2009. *The syllable structure of Bangla in optimality theory and its application to the analysis of verbal inflectional paradigms in distributed morphology.* Tübingen: University of Tübingen (TOBIAS-lib).
Kawahara, S. 2005. *Sonorancy and geminacy*. Unpublished manuscript.
Keer, E. 1999. *Geminates, the OCP and the nature of CON.* Unpublished PhD dissertation, State University of New Jersey, New Brunswick.
Kothari, A. 2004. *Hiatus resolution in Bengali: a non-cyclic account.* Paper presented at the 24th Annual Conference of South Asian Languages Analysis (SALA), Stony Brook, New York.
Lamontagne, G. & Rice, K.1995. A correspondence account of coalescence. *University of Massachusetts Occasional Papers in Linguistics, 18*, 211-224.
Legendre, G., Grimshaw, J. & Vikner, S. (Eds.). 2001. *Optimality-theoretic syntax*. Cambridge, MA: MIT Press.
Legendre, G., Miyata, Y. & Smolensky, P. 1990. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the twelfth annual conference of the Cognitive Science Society* (pp. 388-395). Cambridge, MA: Lawrence Erlbaum.
McCarthy, J. 2000. The prosody of phase in Rotuman. *Natural Language and Linguistic Theory, 18*(1), 147-197.
McCarthy, J. 2001. *A Thematic guide to optimality theory*. Cambridge: Cambridge University Press.
McCarthy, J. 2003. Sympathy, cumulativity, and the Duke-of-York

Gambit. In C. Féry & R. van de Vijver (Eds.), *The syllable in optimality theory* (pp. 23-76). Cambridge: Cambridge University Press.

McCarthy, J. & Prince, A. 1993a. Generalised alignment. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1993* (pp. 79-153). Doedrecht: Kluwer.

McCarthy, J. & Prince, A. 1993b. *Prosodic morphology: constraint interaction and satisfaction*. Amherst: University of Massachusetts, Amherst and Rutgers Centre for Cognitive Science (RuCCS) technical report 3.

McCarthy, J. & Prince, A. 1995. Faithfulness and Reduplicative Identity. In J. Beckman, S. Urbanczyk & L. Walsh Dickey (Eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory* (pp. 249-384). Amherst, Massachusetts: GLSA.

Murray, R. & Vennemann, T. 1983. Sound change and syllable structure in Germanic phonology. *Language, 59*, 514-528.

Ohala, J. 1983. The origin of sound patterns in vocal tract constraints. In P. MacNeilage (Ed.), *The Production of Speech* (pp. 189-216). New York: Springer Verlag.

Paradis, C. 1988. On constraints and repair strategies. *The Linguistic Review, 6*(1), 71-97.

Pater, J. 1999. Austronesian nasal substitution and other NC effects. In R. Kager, H. van der Hulst & W. Zonneveld (Eds.), *The Prosody-Morphology Interface* (pp. 310-343). Cambridge: Cambridge University Press.

Pater, J. 2007. *Harmonic grammar, gradual learning and phonological gradience.* Paper presented at the Workshop on Variation, Gradience, and Frequency in Phonology, Stanford University.

Prince, A. & Smolensky, P. 1993. *Optimality theory : constraint interaction in generative grammar*.Unpublished manuscript, Rutgers University, New Brunswick and University of Colorado, Boulder.

Prince, A. & Smolensky, P. 1997. Optimality: from neural networks to universal grammar. *Science, 275*, 1604-1610.

Rose, S. 2000. Epenthesis positioning and syllable contact in Chaha.

Phonology, 17, 397-425.

Sarkar, P. 2006. *Bangla balo*. Kolkata: Dey's Publishing.

Shinohara, S. 2002. *Emergence of Universal Grammar in foreign word adaptations.* Paper presented at the CNRS UPRESA 7018, Université Paris 3.

Simmler, F. 1974. *Die westgermanische konsonantengemination im Deutschen unter besonderer berücksichtigung des Althochdeutschen*. München: Fink.

Singh, U. N. 1980. Comments on 'On rule ordering in Bengali phonology'. *Indian Linguistics, 40*(2), 91-101.

Smolensky, P. & Legendre, G. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT press.

Taylor, M. 1985. Some patterns of geminate consonants. *The University of Chicago Working Papers in Linguistics, 1*, 120-129.

Vennemann, T. 1988. *Preference laws for syllable structure and th e explanation of sound change*. Berlin: Mouton de Gruyter.

Vijayakrishnan, K. G. 2003. Weakening processes in the optimality framework. In J. M. van de Weijer, V. van Heuven & H. van der Hulst (Eds.), *The Phonological Spectrum* (pp. 241-255). Amsterdam: John Benjamins.

Zuraw, K. R. 2003. Optimality theory in linguistics. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks* (2 ed., pp. 819-822). Cambridge, MA: MIT Press.

Email Contact: somdev.kar@gmail.com