

UNIVERSITY OF CALIFORNIA

Los Angeles

Similarity Avoidance
in Bengali Fixed-Segment Reduplication

A thesis submitted in partial satisfaction
of the requirements for the degree Master of Arts
in Linguistics

by

Sameer ud Dowla Khan

2006

The thesis of Sameer ud Dowla Khan is approved.

Bruce Hayes

Colin Wilson

Kie Ross Zuraw, Committee Chair

University of California, Los Angeles

2006

TABLE OF CONTENTS

§	SECTION TITLE	PAGE
1	Introduction	1
1.1	Fixed-Segment Reduplication in Bengali	2
1.1.1	Bengali Echo Fixed-Segment Reduplication	2
1.2	Fixed-Segment Reduplication Cross-Linguistically	7
1.3	Previous Studies on Identity Avoidance and Similarity Avoidance in Bengali Fixed-Segment Reduplication	11
1.4	Identity Avoidance and Similarity Avoidance in Non-Reduplicative Processes Cross-Linguistically	14
1.4.1	Measuring Similarity Avoidance in Arabic Triliteral Roots	16
1.4.2	Measuring Similarity Avoidance in Muna Roots	20
1.4.3	Measuring Similarity Avoidance in Bengali Roots	25
2	Predictions	26
2.1	Language Background: Bengali বাংলা	27
2.1.1	Dialect Background: Common East Bengali	27
2.2	Applying the Natural Classes Metric to Bengali Fixed-Segment Echo Reduplication	33
2.2.1	Predicted Fixed Segment /t/-Use in Reduplicants of Native Bengali Bases	37
2.2.2	Predicted Fixed Segment /t/-Use in Reduplicants of Borrowed English Bases	40
2.3	Applying Relativized OCP Constraints to Bengali Fixed-Segment Echo Reduplication	44
2.4	Comparing Consonant Cooccurrence and Bengali Fixed-Segment Echo Reduplication	50
2.5	Feature Weighting	55
3	Current Study	58
3.1	Methods	58
3.1.1	Subjects	58
3.1.2	Stimuli	59
3.2	Results	64
3.2.1	Reduplicants of Native Words	64
3.2.2	Reduplicants of English Borrowings	70
3.2.3	Reduplicants of Sanskrit and English Borrowings with Initial Clusters	73
3.3	Analysis	73
3.3.1	Shared Natural Classes Metric	73
3.3.2	Relativized OCP Constraints	81
3.3.3	Correlation with Consonant Cooccurrence	87
3.3.4	Feature Weighting	90
4	Conclusions and Further Investigation	96

5	Appendices	105
5.1	Appendix A: Total Reduplication Types in Bengali	105
5.2	Appendix B: Non-Echo Fixed-Segment Reduplication Types in Bengali	106
5.3	Appendix C: Feature Specifications Used	112
5.4	Appendix B: Stimulus Sets I and II (Native Words)	113
5.5	Appendix C: Stimulus Set III (English Loanwords)	115
5.6	Appendix D: Stimulus Set IV (Cluster-Initial Loanwords)	117
5.7	Appendix E: Pairwise t-test Results	119
6	References	121

TABLE OF FIGURES

LABEL	DESCRIPTION	PAGE
Figure 1.	<i>Fixed segment /t/ use in reduplicants of native Bengali words with different initial consonants, in data collected during a pilot survey</i>	6
Figure 2.	<i>The Natural Classes Metric, as introduced in Pierrehumbert (1993)</i>	16
Figure 3.	<i>Similarity of selected consonants in Arabic as determined by the shared natural classes metric</i>	17
Figure 4.	<i>Calculation for O/E ratio of cooccurrence rate of /tʰ/ and /θ/ in Arabic roots</i>	18
Figure 5.	<i>Adjacent consonant cooccurrence in Arabic roots</i>	19
Figure 6.	<i>Consonant cooccurrence rates within Muna (V)CVCVCV roots</i>	20
Figure 7.	<i>Tableau illustrating the relativized OCP constraint hierarchy in (11) of the Coetzee & Pater (2005) study of consonant cooccurrence in Muna roots, showing which root-types would be expected given the relativized OCP constraints.</i>	23
Figure 8.	<i>Tableau showing how some outputs can win by being faithful to an indexed input, despite being sub-optimal as seen in Figure 7</i>	24
Figure 9.	<i>Rough sketch of the consonantal inventory of Common East Bengali</i>	29
Figure 10.	<i>The vocalic inventory of Common East Bengali</i>	32
Figure 11.	<i>Similarity values of selected Common East Bengali consonants to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch et al. shared natural classes metric, as calculated by FeaturePad and Similarity</i>	34
Figure 12.	<i>Similarity values of selected Common East Bengali coronals to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch et al. shared natural classes metric, as calculated by FeaturePad and Similarity</i>	35
Figure 13.	<i>Similarity values of selected Common East Bengali coronal obstruents to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch, et al. shared natural classes metric, and as calculated by FeaturePad and Similarity</i>	36
Figure 14.	<i>Similarity scores of native Bengali consonants to apico-alveolar /t/. Same as Figure 11</i>	38
Figure 15.	<i>Dissimilarity scores of native Bengali consonants to apico-alveolar /t/</i>	39
Figure 16.	<i>Rough sketch of the consonantal inventory of “Bengali English”</i>	41
Figure 17.	<i>Dissimilarity values of selected Bengali consonants to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch, et al. shared natural classes metric, and as calculated by FeaturePad and Similarity. This is the same graph seen in Figure 15</i>	43
Figure 18.	<i>Dissimilarity values of selected English consonants to the voiceless alveolar stop /t/, as measured by the Frisch, et al. shared natural classes metric, and as calculated by FeaturePad and Similarity</i>	43
Figure 19.	<i>Dissimilarity values of selected “Bengali English” consonants to the voiceless alveolar stop /t/, as measured by the Frisch, et al. shared natural classes metric, and as calculated by FeaturePad and Similarity</i>	44

Figure 20. <i>Relativized OCP-predicted dissimilation pattern given the constraint hierarchy in Table 1</i>	49
Figure 21. <i>Calculation for O/E ratio of cooccurrence rate of initial /k/ and medial /t/</i>	53
Figure 22. <i>Observed / Expected Cooccurrence Rates of selected Standard Kolkata Bengali phonemes with /t/, using data collected from Mallik et al. (1998)</i>	54
Figure 23. <i>Percentage of fixed segment /t/-use in reduplicants of native words of varying initial consonants, when presented via audio recording format</i>	65
Figure 24. <i>Percentage of fixed segment /t/-use in reduplicants of native words of varying initial consonants, when presented via written test format</i>	69
Figure 25. <i>Percentage of fixed segment /t/-use in reduplicants of English borrowings of varying initial consonants, when presented via audio recording format</i>	71
Figure 26. <i>Percentage of fixed segment /t/-use in reduplicants of English borrowings of varying initial consonants, when presented via written test format</i>	72
Figure 27. <i>Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format, versus the predicted fixed segment /t/-use as per the Frisch, et al. model of similarity based on shared natural classes, measured across all Bengali phonemes</i>	74
Figure 28. <i>Fixed segment /t/-use predicted by Frisch et al shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory</i>	75
Figure 29. <i>Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format, versus the predicted fixed segment /t/-use as per the Frisch, et al. model of similarity based on shared natural classes, measured across all Bengali phonemes, excluding the phoneme /h/</i>	76
Figure 30. <i>Fixed segment /t/-use predicted by Frisch et al. shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed</i>	77
Figure 31. <i>Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format, versus the predicted fixed segment /t/-use as per the Frisch, et al. model of similarity based on shared natural classes, measured across Bengali coronals only</i>	78
Figure 32. <i>Fixed segment /t/-use predicted by Frisch et al. shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the coronal consonants only</i>	79
Figure 33. <i>Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format, versus the predicted fixed segment /t/-use as per the Frisch, et al. model of similarity based on shared natural classes, measured across Bengali coronal obstruents only</i>	80
Figure 34. <i>Fixed segment /t/-use predicted by Frisch, et al. shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured</i>	

	<i>across the coronal obstruents only</i>	80
Figure 35.	<i>Nonsense candidates subjected to violations of relativized OCP constraints arranged in order of allowing fewest to most exceptions</i>	83
Figure 36.	<i>Relativized OCP-predicted dissimilation pattern graphed against the observed pattern. Predicted dissimilarity is calculated as the percentage of constraints not violated by the cooccurrence of the consonant with fixed segment /t/</i>	84
Figure 37.	<i>Fixed segment /t/-use predicted by a relativized OCP constraint hierarchy graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed</i>	85
Figure 38.	<i>Relativized OCP-predicted dissimilation pattern graphed against the observed pattern, measuring coronals only</i>	86
Figure 39.	<i>Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format, versus the predicted fixed segment /t/-use as determined by consonant cooccurrence within CVCV roots</i>	88
Figure 40.	<i>Fixed segment /t/-use predicted by root-internal consonant cooccurrence graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed</i>	89
Figure 41.	<i>Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format, versus the predicted fixed segment /t/-use as determined by feature weighting as described in Table 3. Similarity measured across the entire phoneme inventory</i>	92
Figure 42.	<i>Fixed segment /t/-use predicted by feature weighting graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory</i>	93
Figure 43.	<i>Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format, versus the predicted fixed segment /t/-use as determined by feature weighting as described in Table 3. The data point for /h/-initial bases is removed</i>	94
Figure 44.	<i>Fixed segment /t/-use predicted by feature weighting graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed</i>	94

TABLE OF EQUATIONS

LABEL	DESCRIPTION	PAGE
	Equation 1. <i>Similarity as a decreasing function of feature disparity</i>	56
	Equation 2. <i>The probability using fixed segment /t/ in the reduplicant of a C1-initial base n times out of m trials is given by this binomial formula</i>	57

TABLE OF TABLES

LABEL	DESCRIPTION	PAGE
	Table 1. <i>Possible relativized OCP constraint hierarchy for Common East Bengali fixed-segment reduplication similarity avoidance</i>	47
	Table 2. <i>Bengali consonants that would receive a violation from each relativized OCP constraint, when cooccurring with /t/</i>	48
	Table 3. <i>Feature weights predicted by the feature-weighting metric to be greater than default 0.1</i>	91

ABSTRACT OF THE THESIS

Similarity Avoidance
in Bengali Fixed-Segment Reduplication

by

Sameer ud Dowla Khan

Masters of Arts in Linguistics

University of California, Los Angeles, 2006

Professor Kie Ross Zuraw, Chair

As described in Yip (1998), Lidz (2001), and Nevins & Vaux (2003), many languages employ morphological processes requiring a reduplicant to be minimally phonologically distinct from its base, typically achieved by replacing one segment in the reduplicant with a fixed segment. In Common East Bengali, the fixed-segment /t/ replaces the initial consonant of the reduplicant; however, when the base itself begins with /t/ or a consonant similar to /t/, replacing the original segment with the fixed segment /t/ in the reduplicant may not be sufficient to create a significant distinction between base and reduplicant. In an experiment carried out on speakers of this

language, it was found that the more similar a base's initial consonant is to /t/, the more likely it is that the reduplicant will not begin with the fixed segment /t/; instead, speakers tend to use alternate fixed segments such as /f/ or /m/ to fulfill phonological dissimilarity between base and reduplicant. On what metric are these speakers judging the similarity of base-initial consonants to /t/? This study suggests that the metrics presented in three analyses of phonological similarity in Arabic, Muna, and Bengali roots cannot accurately describe the patterns observed in Common East Bengali fixed-segment reduplication. Instead of measuring similarity by counting the number of phonological features shared across consonants, this study explores an analysis that involves assigning greater weights to features such as [voice] and [distributed], as they distinguish consonants judged similar to /t/ from those judged dissimilar from /t/ more reliably than other features.

1 INTRODUCTION

Cross-linguistically, the reduplicants in fixed-segment reduplication (e.g. English *doctor-schmactor*) must often be significantly distinct from the base. This phenomenon is typically seen in fixed-segment echo reduplication, where a phonological distinction between the base and its reduplicant is typically manifest in the differentiation of the initial consonants of the two words; the first consonant of the reduplicant is normally a fixed segment instead of an exact copy of the base-initial consonant. However, in cases where the use of a particular fixed segment does not produce a significant distinction between base and reduplicant (e.g. English **schmooze-schmooze*), languages react in different ways. While some languages allow echo base-reduplicant pairs where the fixed initial consonant of the reduplicant is homophonous with the initial consonant of the base, other languages attempt to avoid this echo base-reduplicant homophony.

This study examines the fixed-segment echo reduplication patterns of a dialect of Bengali, in order to better investigate (a) how phonologically or phonetically distinct the fixed segment must be from the base-initial consonant, (b) on what metric speakers gauge this phonological or phonetic similarity, and (c) the applicability of this metric cross-linguistically.¹

¹ Many thanks to Farida Amin Khan, my primary Bengali consultant; Kie Ross Zuraw, Colin Wilson, and Bruce Hayes, my thesis committee and statistics consultants; the

1.1 FIXED-SEGMENT REDUPLICATION IN BENGALI

Fixed-segment reduplication (henceforth fixed segment reduplication or fixed segment reduplication) is described in Yip (1998) as a morphophonological process in which identity is avoided between a base and its reduplicant. Specifically, Yip describes fixed segment reduplication as “reduplication accompanied by a small change such that the two halves are not quite identical,” (Yip 1998; 5). Bengali has five regular patterns of fixed segment reduplication: reciprocal action/state (with one pattern for native words and another for Sanskrit and foreign loans), monosyllabic, emphatic verb, and echo reduplications.

The current study examines in detail the Echo Reduplication construction, defined as a near-total reduplication of the base, where the only base segment not copied in the reduplicant is a fixed segment²; while they are normally phonologically possible words in the language, echo reduplicants typically have no meaning on their own³, while their corresponding bases do; this combination of features often sets echo reduplication apart from other types of reduplication.

UCLA Phonology Seminar attendees; and of course, the thirty subjects who participated in this study.

² Echo reduplication patterns in Dravidian languages involve the use of a fixed syllable, made up of a velar onset and high front vowel nucleus. See Keane (2001) for a detailed account.

³ Keane (2001) discusses possible counterexamples to this generalization in Telugu (Bhaskararao 1977), Malayalam (Malten 1989), and Turkish (Fitzpatrick-Cole 1994).

1.1.1 BENGALI ECHO FIXED-SEGMENT REDUPLICATION

The extremely productive echo reduplication construction in Bengali uses a Fixed Segment (fixed segment), normally /t/⁴, as shown in the following examples:

- | | | |
|-----|---|---|
| (1) | g ^h uʃi ⁵ | ‘punch’ |
| | g ^h uʃi k ^h aæa | ‘having gotten punched’ |
| | g ^h uʃi k ^h aæa kaṇṭt̪ese. | ‘Having gotten punched, he’s crying.’ |
| | g ^h uʃi t̪uʃi k ^h aæa kaṇṭt̪ese. | ‘Having gotten punched and whatnot, he’s crying.’ |
| | g ^h uʃi k ^h aæa t̪aæa kaṇṭt̪ese. ⁶ | ‘Having gotten punched and whatnot, he’s crying.’ |
| (2) | boj | ‘book’, ‘books’ |
| | uni boj p̪æen na. | ‘He doesn’t read books.’ |
| | uni boj t̪oj p̪æen na. | ‘He doesn’t read books or anything.’ |

1.1.1.1 VARIANTS OF THE FIXED SEGMENT IN BENGALI FIXED SEGMENT ECHO REDUPLICATION

The vast majority of Bengali speakers prefer /t/ as their default fixed segment in echo reduplication; however, there are some individuals who prefer other consonants as a possible fixed segment. In the Common East Bengali dialect⁷, the most

⁴ All fixed segments in this study are typed in **boldface**.

⁵ Excluding those expressly identified otherwise, all Bengali examples used in this study can be attributed to my primary consultant, Farida Amin Khan.

⁶ Any difference in meaning between these last two sentences is negligibly subtle.

⁷ This dialect is described more fully in Section 2.1.1.

prevalent consonantal fixed segments after /t/ are the labials /m/, /f/, /b/, and /p/, and the breathy affricate /dʒʱ/ (sometimes in variation with the fricative /z/):

- | | | |
|-----|--|---|
| (3) | ʌjgga ⁸
ʌjgga m ajgga | ‘having gotten angry’
‘having gotten angry and all’ |
| (4) | tʃujʃʃa
tʃujʃʃa f ujʃʃa | ‘having sucked’
‘having sucked and whatnot’ |
| (5) | dana
dana b ana | ‘wing’, ‘wings’
‘wings and other such things’ |
| (6) | silka
silka p ilka | ‘peel’, ‘peels’
‘peels, et cetera’ |
| (7) | najmma
najmma dʒʱ ajmma | ‘having gotten down’
‘having gotten down and everything’ |

Many speakers prefer to preserve the initial consonant, and instead replace the vowel of the first syllable to /u/ or /a/, serving as a vocalic fixed segment.

- | | | |
|-----|---------------------------|--|
| (8) | saʒʌ
saʒʌ s uʒʌ | ‘having let go’
‘having let go and such’ |
| (9) | tupi
tupi t api | ‘hat’, ‘hats’
‘hats and all that comes with them’ |

⁸ These examples were collected in the experiment described in Section 3.

1.1.1.2 IDENTITY AVOIDANCE AND SIMILARITY AVOIDANCE IN BENGALI FIXED SEGMENT ECHO
REDUPLICATION

In a pilot study carried out on five speakers of the Common East Bengali dialect of Bengali, it was found that speakers used four variants of the fixed segment (/t/, /f/, /m/, and /u/). With so many variants at their disposal, these five speakers tended to avoid using a variant that would create a reduplicant that would be homophonous to its base. Instead, they typically resorted to any of the other segment variants (10):

(10) t̄ometo	‘tomato’, ‘tomatoes’
t̄ometo f̄ometo	‘tomatoes and other such things’
t̄ometo m̄ometo	‘tomatoes and other such things’
t̄ometo t̄umeto	‘tomatoes and other such things’
(*t̄ometo t̄ometo) ⁹	

It seems as though these speakers are not just replacing segments in the reduplicant with just any fixed segment variant. These speakers are intentionally using variants that result in dissimilarity between base and reduplicant; for the purposes of this study, the phenomenon shall be called **Identity Avoidance**.

Even more interestingly, the speakers in pilot survey did not restrict their use of backup fixed segments /f/, /m/, and /u/ to situations where the initial consonant of

⁹ Please note that speakers that do not exhibit Similarity Avoidance would prefer this form. Standard Kolkata Bengali speakers would presumably prefer *t̄ometo t̄ometo* to the variants with fixed segments *f*, *m*, or *u*.

the base was /t/. In fact, /t^h/-initial words also showed a strong tendency of fixed segment /f/-, /m/-, and /u/-use, while /d/-initial, /t̪/-initial, and /t̪^h/-initial words showed less of a tendency. Clearly, these speakers are not just avoiding total identity of base and reduplicant, but extending this to a more gradient **Similarity Avoidance**, where even relatively similar but not identical base-reduplicant pairs such as *[t^haj̪ʃʃa taj̪ʃʃa] are avoided with preference given to pairs such as [t^haj̪ʃʃa faj̪ʃʃa] or [t^haj̪ʃʃa maj̪ʃʃa]. Figure 1 graphs fixed segment /t̪/-use in the reduplicants of Bengali words presented in pilot survey, with the initial consonant of the base strung along the x-axis:

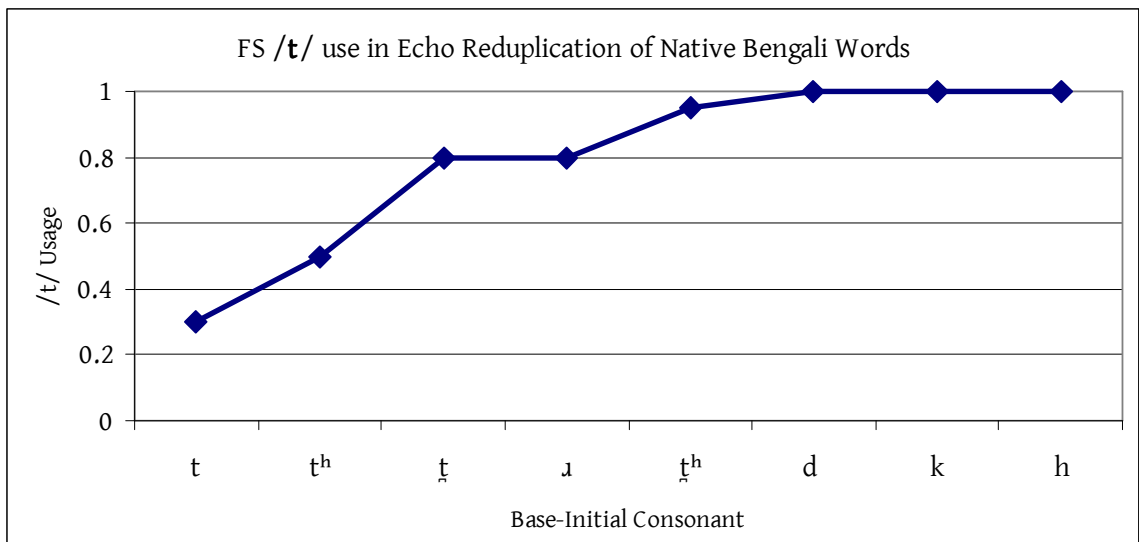


Figure 1. Fixed segment /t̪/ use in reduplicants of native Bengali words with different initial consonants, in data collected during a pilot survey. A score of 1 indicates use of fixed segment /t̪/ in every base of that type. A score of 0.5 would indicate use of fixed segment /t̪/ in half of the bases of that type. Usage rates are collapsed across the five participants in the survey.

As suggested earlier, fixed segment /t̪/ use is not only subject to identity avoidance (where /t̪/-initial bases are extremely unlikely to allow fixed segment /t̪/-

initial reduplicants), but to similarity avoidance, where bases starting with consonants presumably considered more similar to /t/ (e.g. /t^h/, /t̪/, etc.) allow fewer fixed segment /t/-reduplicants than bases starting with consonants considered less similar to /t/ (e.g. /k/, /h/, etc.). This current study aims to investigate the following questions:

1. *On what basis do Bengali speakers measure consonant similarity?*

2. *Is this measurement of similarity universal or language-specific?*

Before attempting to answer these questions, let's examine other studies on fixed-segment reduplication cross-linguistically (Section 1.2), fixed-segment echo reduplication in Bengali (Section 1.3), and similarity avoidance in non-reduplicative constructions in three languages (Section 1.4).

1.2 FIXED-SEGMENT REDUPLICATION CROSS-LINGUISTICALLY

Identity avoidance and similarity avoidance are common cross-linguistically in fixed-segment reduplication. Identity avoidance, as defined here, is the avoidance of total homophony between the base and reduplicant, while similarity avoidance is more general – the avoidance of homophony or near-homophony between the base and reduplicant. Because fixed-segment reduplication normally results in a reduplicant that differs from its base in just one segment (the fixed segment), it is not unlikely that speakers associate the dissimilarity of base and reduplicant with the construction.

When a base includes a segment that is homophonous to the fixed segment it is in correspondence with,¹⁰ speakers of different languages react in different ways.

Many cases of identity- and similarity avoidance are documented in Nevins & Vaux (2003) and Yip (1998) among others; some are summarized below, illustrating the four manners in which languages can avoid pronouncing a reduplicant that is homophonous to its base:

- A. Allowing base-reduplicant homophony.
(Seen in West Bengali echo reduplication)
- B. Maintaining a paradigmatic gap.
(Seen in Turkish echo reduplication)
- C. Using an established backup fixed segment.
(Seen in Farsi, Hindi, Kamrupi, and Kashmiri echo reduplication)
- D. Choosing the best candidate from an established set of fixed segments.
(Seen in Turkish emphatic reduplication)
- E. Choosing from a large and possibly undefined set of backup fixed segments.
(Seen in English and Kannada echo reduplication)

Language Type A: *Western Bengali dialects* - The default fixed segment is reduplicant-initial /t/; total base-reduplicant identity is not avoided.

(11) *tebil*¹¹ 'table', 'tables'

¹⁰ The theory that fixed segments are in correspondence with base material is not universally-accepted; see McCarthy & Prince (1995) for a counterargument.

tebil tebil

‘tables and such’

Language Type B: Turkish (echo) - The default fixed segment is reduplicant-initial /m/; total base-reduplicant identity is avoided by simply never uttering a reduplicated form of potentially-homophonous base-reduplicant pairs. No backup fixed segments have been attested.

(12) kitap ‘book’
 kitap **mitap** ‘books and such’

(13) masa ‘table’
 *masa **masa** (there is no echo reduplicant for ‘table’)

Language Type C: Farsi - The default fixed segment is reduplicant-initial /m/; total base-reduplicant identity is avoided by using the backup fixed segment /p/ when approached by potentially-homophonous base-reduplicant pairs.

(14) pærde¹² ‘curtain’, ‘curtains’
 pærde **mærde** ‘curtains and such’

(15) mive ‘fruit’, ‘fruits’
 mive **pive** ‘fruits and such’ (*mive **mive**)

¹¹ Note that the phoneme transcribed in this study as (post)alveolar /t/ is transcribed in Nevins & Wagner (2001) as retroflex /t̺/. Thus, the default fixed segment for this dialect of Bengali is considered to be /t̺/ in Nevins & Wagner (2001).

¹² Shademan, p.c.

The Farsi method of identity avoidance is also seen in Hindi, where default fixed segment /v/ is replaced with backup fixed segment /ʃ/ in reduplicants of /v/-initial bases (Nevins & Wagner 2001), in Kamrupi, where default fixed segment /s/ is replaced with backup fixed segment /t/ in reduplicants of /s/-initial bases (Fitzpatrick & Nevins 2002), and in Kashmiri, where default fixed segment /v/ is replaced with backup fixed segment /p/ (Koul n.d.) in reduplicants of /v/-initial bases.

Language Type D: Turkish (*emphatic*) - Total base-reduplicant identity is avoided by using one of four possible reduplicant-final fixed segments (/p/, /s/, /m/, and /r/). No one fixed segment is clearly the default variant; speakers choose the fixed segment that is least similar to any consonant in the base. In this construction, Turkish speakers appear to be applying a form of similarity avoidance. Examples (16) - (19) are drawn from Yip (1998). See Nevins & Wagner (2001) for an analysis of this phenomenon.

- | | | |
|--------|--------------------|-------------------------------------|
| (16) | kara
kap kara | ‘black’
‘jet black’ |
| (17) | belli
bes belli | ‘obvious’
‘unmistakably obvious’ |
| (18) | siki
sim siki | ‘tight’
‘extremely tight’ |
| (19) | top
tor top | ‘round’
‘fully round’ |

Language Type E: English - Total base-reduplicant identity is avoided by using one of many possible backups to default fixed-segment /ʃm/. Possible alternatives range widely; for example, 64% of respondents to the survey conducted in Nevins & Vaux (2003) felt no output was possible for the /ʃm/-reduplication of *schmooze*, while 31% preferred some form without initial *schm-* (many forms were attested), and only 5% preferred the non-dissimilated *schmooze-schmooze*.

- | | | |
|--------|----------------------|---|
| (20) | tʰɹki
tʰɹki ʃmɹki | ‘turkey’
‘turkey schmurkey’ |
| (21) | ʃmuz
ʃmuz ʃmuz | ‘schmooze’
‘schmooze schnooze’ (??ʃmuz-ʃmuz) |

It is noted in Lidz (2001) that, like the English speakers interviewed in Nevins & Vaux (2003), four Kannada consultants provided different responses when faced with a potentially-homophonous base-reduplicant pair. While two speakers replaced the default fixed segment (or more precisely, fixed syllable) /gi/ with /bi/ or /vi/ in reduplicants of /gi/-initial bases, another speaker felt that identity was permissible, while yet another speaker stated that many Kannada speakers would simply not reduplicate bases where homophony with the reduplicant would be predicted.

1.3 PREVIOUS STUDIES ON IDENTITY AVOIDANCE AND SIMILARITY AVOIDANCE IN BENGALI FIXED-SEGMENT REDUPLICATION

Although there have been many previous studies on fixed-segment reduplication in Bengali, including Trivedi (1990), Fitzpatrick-Cole (1994) and (1996), Jha *et al.* (1997), and Nevins & Wagner (2001), among others, no previous study (to my knowledge) has discussed the issue of similarity avoidance or identity avoidance in Bengali. In fact, the Bengali data in Nevins & Wagner (2001) include cases of homophonous base-reduplicant pairs, as shown in example (11), drawn from that study and repeated below in (22):

- | | | |
|--------|-----------------------------|--|
| (22) | tebil
tebil tebil | ‘table’, ‘tables’
‘tables and such’ |
|--------|-----------------------------|--|

The base [tebil] is reduplicated with the fixed segment /t/, giving [tebil **tebil**], which illustrates total homophony of base and reduplicant. It is essential to note that the dialect studied in Nevins & Wagner (2001) is presumably Standard Kolkata Bengali, the Bengali dialect most often encountered in the literature. A base-reduplicant pair such as the one in (22) would be largely unacceptable to speakers of eastern dialects such as Common East Bengali. A more acceptable reduplication of the base [tebil] ‘table’ in Common East Bengali is given in (23):

- | | | |
|--------|--|--|
| (23) | tebil
tebil tubil ¹³
*tebil tebil ¹⁴ | ‘table’, ‘tables’
‘tables and such’ |
|--------|--|--|

¹³ In the experiment discussed in Section 0, more speakers (53.3%) judged [tebil **tubil**] as the most grammatically acceptable base-reduplicant pair variant than any other variant for the base [tebil].

The Ray *et al.* (1966) study of the Standard Kolkata Bengali dialect briefly mentions the assortment of fixed segments in echo reduplication. The construction in Standard Kolkata Bengali is described as having “an obligatory initial consonant, usually /t/, less commonly /p^h/, rarely /ʃ/¹⁵ (p 58),” as shown in the following examples:

- | | | |
|--------|---|--|
| (24) | g ^h ora
g ^h ora tora | ‘horse’
‘a horse or the like’ |
| (25) | ʌumal
ʌumal p ^h umal | ‘handkerchief’
‘a handkerchief or the like’ |
| (26) | mota
mota ʃota | ‘fat’
‘stoutish’ |

Ray *et al.* also identify one vocalic fixed segment, referred to as “an obligatory first vowel /a/” (p 58), providing an example repeated here in (27).

- | | | |
|--------|--|-----------------------------------|
| (27) | b ^h ul
b ^h ul b ^h al | ‘error’
‘an error or the like’ |
|--------|--|-----------------------------------|

¹⁴ In the experiment discussed in Section 0, only two speakers (out of 15) produced the string [tebil **tebil**] as a reduplicant of [tebil]. It may be appropriate to mention that of these two speakers, one is a fluent speaker of a western dialect of Bengali (in addition to Common East Bengali), and the other later decided in the written test that [tebil **tebil**] was ungrammatical.

¹⁵ These consonants are transcribed in Ray *et al.* (1966) as /T/, /ph/, and /S/, respectively.

Bykova (1981) also describes echo reduplication in Standard Kolkata Bengali as “incompletive reduplication” (p 105), where fixed segments /t/, /b/, /ʃ/, and /a/ are mentioned in four examples copied below.

- | | | |
|--------|---|---|
| (28) | lutʃi
lutʃi tutʃi | ‘luchi (a flat cake fried in clarified butter)’
‘luchi etc.’ |
| (29) | tʃakɪ
tʃakɪ bakɪ | ‘job’, ‘jobs’
‘jobs and such.’ |
| (30) | budʒ ^h e
budʒ ^h e ʃudʒ^he | ‘having understood’
‘having understood and whatnot’ |
| (31) | tʃup
tʃup tʃap ¹⁶ | ‘silence’, ‘silent’
‘totally silent’ |

While they do mention the multiplicity in possible fixed segments, both the Bykova (1981) and Ray *et al.* (1966) studies fail to describe the specific distribution of the five fixed segments they mention (*i.e.* /t/, /p^h/, /b/, /ʃ/, and /a/), thus avoiding the question of what triggers the choice of one variant over another. And while identity avoidance in fixed segment reduplication has been documented in other languages, as mentioned earlier, similarity avoidance is apparently not discussed in depth in the

¹⁶ This construction is, in my opinion, not an example of echo reduplication, but of another type of fixed segment reduplication. See Appendix B for a discussion of this construction (monosyllabic fixed segment reduplication) and other fixed segment reduplication types in Bengali.

literature on Bengali or any other language with productive fixed segment reduplication patterns.

Although similarity avoidance in fixed-segment reduplication has not been thoroughly investigated in many languages, similarity avoidance in non-reduplicative patterns is a well-known phenomenon. To better understand how similarity avoidance works cross-linguistically, I shall discuss analyses of similarity avoidance in non-reduplicative processes in Section 1.4.

1.4 IDENTITY AVOIDANCE AND SIMILARITY AVOIDANCE IN NON-REDUPLICATIVE PROCESSES CROSS-LINGUISTICALLY

Identity avoidance (the avoidance of totally homophonous segments) and similarity avoidance (the avoidance of segments judged relatively similar to each other) are not restricted to fixed-segment reduplication. The effects of the prohibition of adjacent identical segments and/or features, the Obligatory Contour Principle (OCP), as described in Leben (1973) and McCarthy (1986), often serve as a manifestation of identity- and similarity avoidance in countless constructions. The narrowest version of the OCP disallows strings of adjacent segments when the segments are totally identical (*i.e.* identity avoidance). A broader version of the OCP that forbids any string of

adjacent¹⁷ *homorganic* consonants (*i.e.* similarity avoidance) has been proposed to describe the relationship between OCP and consonant cooccurrence within roots in the lexicons of many languages.

Although the tendency for homorganic consonants to be underrepresented in Arabic trilateral roots has been noted in many studies, including Greenberg (1950), McCarthy (1988) and (1994), Pierrehumbert (1993), Padgett (1995), and Frisch, Pierrehumbert, and Broe (2004), only this last study, Frisch *et al.* (2004), examines the Arabic data in the framework of a gradient OCP that is sensitive to how certain consonants pattern with respect to others. Section 1.4.1 describes a metric of consonant similarity used to derive the OCP effects in Arabic roots, while Section 1.4.2 describes an extension of the OCP to describe similar phenomena in the Western Austronesian language Muna. Section 1.4.3 examines consonant cooccurrence restrictions in Bengali roots, in order to see if there is a correlation with similarity avoidance in Bengali echo fixed segment reduplication.

1.4.1 MEASURING SIMILARITY AVOIDANCE IN ARABIC TRILITERAL ROOTS

The vast majority of Arabic roots are thought to be made up of three consonants each; these three letters make up the trilateral root ([dʒɪðr] in Arabic). In order to form

¹⁷ Note that ‘adjacent’ is to be taken in an abstract sense here, as the consonants in Arabic trilateral roots do not always occur adjacent to one another on the surface. They are more typically separated by vowels as determined by the ‘pattern’ of the alternation.

real words, the three consonants of the trilateral root fit into patterns ([wazn], plural [awza:n], in Arabic) of vowels and consonants. Many linguists have noted that the consonants that make up the Arabic trilateral root are very rarely homorganic; while roots with heterorganic consonants such as /d^ɕ-χ-m/ (the root of [d^ɕaχm] ‘huge’) and /ð-h-b/ (the root of [ðaha:b] ‘going’) are very common, roots with homorganic consonants such as /f-h-m/ (the root of [ʔafham] ‘I understand’) and /s-n-d/ (the root of [istina:d] ‘dependence’) are remarkably atypical. It appears that similarity of consonants within a trilateral root is being avoided in the lexicon. But what does “similarity” mean in this sense? Similarity, according to Pierrehumbert (1993) and as measured in Frisch *et al.* (2004), is defined by the number of shared natural classes between phonemes within a major place of articulation, as shown in Figure 2 below.

$$\text{Similarity} = \frac{\text{Shared natural classes}}{\text{Shared natural classes} + \text{Non-shared natural classes}}$$

Figure 2. *The Natural Classes Metric of Similarity, as introduced in Pierrehumbert (1993). Taken from Frisch, Pierrehumbert, and Broe (2004), example (7).*

For example, as shown in Figure 3 below, the similarity score of the Arabic consonant pair {b, b} is 1, as the two consonants are identical and thus share all the same natural classes. The similarity score for the consonant pair {b, m} is 0.5, as they share some (but not all) natural classes, while the consonant pair {b, t} has a similarity score of 0.0, as they are not even found in the same place category. This stipulation to not measure the similarity of consonants across major place features is crucial in Frisch *et al.*'s analysis, as it predicts that labial stops and coronal stops, for example, share no

similarity with respect to this phenomenon, and can thus freely cooccur. This topic will be addressed again later.

	b	f	m	t	d	t ^ʕ	d ^ʕ	θ	ð	s	z	s ^ʕ	z ^ʕ	ʃ	k
b	1	.38	.5	0	0	0	0	0	0	0	0	0	0	0	0
f	.38	1	.22	0	0	0	0	0	0	0	0	0	0	0	0
m	.5	.22	1	0	0	0	0	0	0	0	0	0	0	0	0
t	0	0	0	1	.42	.26	.17	.21	.14	.32	.19	.12	.08	.12	0
d	0	0	0	.42	1	.16	.3	.13	.21	.17	.32	.08	.13	.07	0
t ^ʕ	0	0	0	.26	.16	1	.4	.24	.14	.14	.09	.41	.21	.13	.21
d ^ʕ	0	0	0	.17	.3	.4	1	.13	.23	.09	.14	.2	.42	.07	.11
θ	0	0	0	.21	.13	.24	.13	1	.45	.4	.24	.45	.24	.37	0
ð	0	0	0	.14	.21	.14	.23	.45	1	.25	.44	.24	.44	.21	0
s	0	0	0	.32	.17	.14	.09	.4	.25	1	.44	.35	.2	.3	0
z	0	0	0	.19	.32	.09	.14	.24	.44	.44	1	.2	.35	.17	0
s ^ʕ	0	0	0	.12	.08	.41	.2	.45	.24	.35	.2	1	.42	.33	.11
z ^ʕ	0	0	0	.08	.13	.21	.42	.24	.44	.2	.35	.42	1	.17	.07
ʃ	0	0	0	.12	.07	.13	.07	.37	.21	.3	.17	.33	.17	1	0
k	0	0	0	0	0	.21	.11	0	0	0	0	.11	.07	0	1
g	0	0	0	0	0	.11	.24	0	0	0	0	.06	.13	0	.38
q	0	0	0	0	0	.36	.17	0	0	0	0	.17	.09	0	.32

Figure 3. Similarity of selected consonants in Arabic as determined by the natural classes metric. Drawn from Table 3 in Frisch, Pierrehumbert, and Broe (2004).

Frisch *et al.* compared the similarity scores they deduced through their natural classes metric to the cooccurrence rate of different consonant pairs in Arabic trilateral roots using the measurement of Observed Cooccurrence divided by Expected Cooccurrence (O/E). An example of the calculation of the O/E cooccurrence rate for the consonant pair {t^ʕ, θ} is provided in Figure 4 below.

Observed {t^ʕ, θ} Cooccurrence in roots

$$\begin{array}{c}
 \text{Total roots} \\
 \hline
 \frac{\text{Observed /t'/ Occurrence in roots}}{\text{Total roots}} \times \frac{\text{Observed /\theta/ Occurrence in roots}}{\text{Total roots}}
 \end{array}$$

Figure 4. Calculation for O/E ratio of cooccurrence rate of /t'/ and /\theta/ in Arabic roots.

Scores of one (1.00) indicate that the observed rate of cooccurrence and the expected rate of cooccurrence of a pair of consonants is equal, suggesting that they are neither over- nor underrepresented in the lexicon. Scores of zero (0.00) indicate that there were no observed cooccurrences of the two sounds. Scores above one indicate that the observed cooccurrence of the two sounds is greater than that predicted by the patterns in the rest of the lexicon; the cooccurrence of such a pair of consonants would be considered overrepresented in the lexicon. Conversely, scores below one indicate that the observed cooccurrence of the two sounds is less than that predicted by the patterns in the rest of the lexicon; the cooccurrence of such a pair of consonants would be considered underrepresented in the lexicon.

Figure 5 summarizes the cooccurrence rates between different consonant types. Labials were found to never cooccur with other labials (cooccurrence rate: 0.00) while coronal sonorants were found to cooccur with dorsals more frequently than expected (cooccurrence rate: 1.48). Gutturals and dorsals cooccurred at close to the expected number (cooccurrence rate: 1.07).

		Adjacent								
		Labial	Coronal obstruents				Dorsal	Guttural		Cor son
		b f m	t d t ^ʕ d ^ʕ	θ ð s z s ^ʕ z ^ʕ ʃ	k g q	χ ʁ	ħ ʕ h ʔ	l r n		
Labial	b f m	0.00	1.37	1.31	1.15	1.35	1.17	1.18		
Cor obs	t d t ^ʕ d ^ʕ		0.14	0.52	0.80	1.43	1.25	1.23		
	θ ð s z s ^ʕ z ^ʕ ʃ			0.04	1.16	1.41	1.26	1.21		
Dorsal	k g q				0.02	0.07	1.04	1.48		
Guttural	χ ʁ					0.00	0.07	1.39		
	ħ ʕ h ʔ						0.06	1.26		
Cor son	l r n							0.06		

Figure 5. *Adjacent consonant cooccurrence in Arabic roots. Copied from Frisch et al. (2004).*

Ignoring cases of identical consonants,¹⁸ an inverse relationship between the similarity of consonants and their observed cooccurrence rates is seen. The more similar two consonants in the same major place category are judged to be (according to the number of natural classes shared between them), the less often they are encountered within the same root.

1.4.2 MEASURING SIMILARITY AVOIDANCE IN MUNA ROOTS

Van den Berg's (1989) grammar of Muna includes data on the cooccurrence rates of consonants within roots, finding that consonants that fall under the same natural class tend to not cooccur within a root, save for cases in which the two consonants are totally identical. In this sense, the pattern is not unlike the Arabic.

¹⁸ Traditionally, cases of identical consonants in trilateral roots are also accounted for using the concept of the OCP. Adjacent identical consonants are treated as one underlying consonant linked to two positions in the word, thus avoiding the issue of cooccurrence altogether. Thus, the two [r]'s in [marar-na:] 'we passed' are interpreted as one /r/ linked to two consonantal positions. The trilateral root of [marar-na:] would thus be considered /m-r/.

Coetzee & Pater (2005) extend the Frisch *et al.* metric of similarity based on shared natural classes to the consonants of Muna, comparing the similarity scores of consonant pairs to their corresponding cooccurrence rates in roots of the shape (V)CVCVCV. Their findings show that, indeed, pairs of homorganic non-identical consonants are underrepresented in Muna roots, as shown in Figure 6.

	Labial			Dorsal			Coronal		
	Obs	Exp	O/E	Obs	Exp	O/E	Obs	Exp	O/E
Labial	156	477.8	0.33						
	bafu, mopi, mafaka								
Dorsal	875	770.9	1.14	63	210.6	0.30			
	meko, bage, bugi			kangia, kagala, kagu ^u di					
Coronal	2741	2179.0	1.26	1766	1522.5	1.16	1400	1751.5	0.80
	bida, budo, topina			kona, kona, si ^u kele			lani, dato, detiki		

Figure 6. Consonant cooccurrence rates within Muna (V)CVCVCV roots. Copied from Table 2 in Coetzee & Pater (2005).

The correlation of cooccurrence and the shared features of place between consonants in Muna is, much like the Arabic phenomenon, remarkable on its own. However, what may be even more worthy of note is the fact that where cooccurrence of two homorganic consonants does occur, additional features such as [nasal] and [voice] can distinguish between a consonant pair like {d, n}, which is observed at a rate of 0.25, and a pair like {d, t}, which is observed at a rate of 0.60. It would appear from the cooccurrence ratings that a pair of segments that differ only in voicing is judged more dissimilar than a pair of segments that differ only in nasality. These and other

findings in Coetzee & Pater (2005) point out some important distinctions between the Arabic pattern and the Muna pattern, as Frisch *et al.* analyze coronal obstruents as one set separate from the coronal sonorants (thus largely disregarding voicing as a factor in OCP-Place), while Coetzee & Pater find voiced coronals acting as one set separate from voiceless coronals, thus reducing the weight of the oral-nasal distinction in OCP-Place.

Coetzee & Pater maintain that only positing one general OCP-Place constraint in an Optimality Theoretic (Prince & Smolensky 1993/2004) approach would require speakers of Muna to accept a rather large number of exceptions to the generalization, when the data actually makes it quite clear that these “exceptions” can be more easily analyzed by positing more fine-grained constraints to distinguish obstruents from sonorants, voiced stops from their voiceless counterparts, and so on. Coetzee & Pater’s explanation is that the Muna grammar includes more than just one OCP-Place constraint or just one OCP(voiced) constraint. Instead, relativized OCP constraints such as OCP-Dorsal (α Voice) and OCP-Dorsal (α Continuant) are ranked over the more general OCP-Dorsal, with lexically-specific Faithfulness constraints interspersed between these markedness constraints. Actual Muna roots that would normally be predicted to be extremely rare are indexed to, and thus protected by, a lexically-specific constraint. The more exceptional the word type, the higher its corresponding faithfulness constraint is ranked.

Instead of having each OCP constraint deal with one feature each, these relativized constraints assume that some features are bundled with other features when

measuring similarity. Positing both OCP-Dorsal (α Voice) and OCP-Dorsal (α Continuant) instead of individual OCP constraints would predict, for example, that having two adjacent dorsals, two adjacent continuants, two adjacent stops, two adjacent voiced consonants, or two adjacent voiceless stops would be acceptable, but having two voiced dorsals, two voiceless dorsals, two dorsal stops, or two dorsal continuants would be penalized. OCP constraints incorporating even more features are allowed in the Coetzee & Pater framework, with their Muna analysis including a constraint OCP-Coronal (α Voice, β Continuant). Hypothetical OT tableaux illustrating the effect of using relativized OCP constraints are given in

Figure 7 and

Figure 8 below. In

Figure 7, instead of assessing an input and output, the constraints assign violations to potential roots in the language; those candidate types that violate the highest-ranked constraints will be underrepresented in the lexicon. For example, while roots such as /dida/, /dita/, and /disa/¹⁹ would be underrepresented, roots such as /diɾa/ would be robustly attested, regardless of the particular number of features shared between the consonants in each root.

Figure 8 illustrates how lexically-specific faithfulness constraints can allow certain exceptions to the distribution otherwise expected.

¹⁹ Please note that although the phonemes used in these candidate words are actual sounds of Muna, the words themselves are totally hypothetical; I do not claim to have any knowledge of actual Muna roots.

	OCP(COR, -Son, αCont)	OCP(COR, +Sonorant)	FAITH-L1	OCP(COR, -Sonorant)	FAITH	OCP(COR)
/disa/				*!		*
/dita/	*!			*		*
/dida/	*!			*		*
/dina/						*!
/diɣa/						

Figure 7. Tableau illustrating the relativized OCP constraint hierarchy in (11) of the Coetzee & Pater (2005) study of consonant cooccurrence in Muna roots, showing which root-types would be expected given the relativized OCP constraints.

/disa-L1/	OCP(COR, -Son, αCont)	OCP(COR, +Sonorant)	FAITH-L1	OCP(COR, -Sonorant)	FAITH	OCP(COR)
☞ [disa]				*		
[dita]	*!		*	*	*	*
[dida]	*!		*	*	*	*
[dina]			*!		*	*
[diɣa]			*!		*	

Figure 8. Tableau showing how some outputs can win by being faithful to an indexed input, despite being sub-optimal as seen in Figure 7.

While both Frisch *et al.* and Coetzee & Pater use phonological features as the basis for their similarity metrics, Frisch *et al.* argue that within each major place, the raw number of natural classes shared between consonants is their measure of similarity, while Coetzee & Pater argue that within each major place, certain features can be thought of as working together to make the similarity of certain consonants more salient than others, oftentimes regardless of the number of features shared, while exceptions to these generalization can be derived using lexically-specified faithfulness constraints.

Coetzee & Pater demonstrate that while the Frisch *et al.* metric of similarity based on shared natural classes is effective in demonstrating the OCP effects in Arabic roots, and to some degree, in Muna roots²⁰, implementing an Optimality Theoretic (OT) analysis with increasingly relativized OCP constraints ranked above the general OCP(Place) can better describe the peculiarities of the Muna data. They suggest that language-specific details can be factored in using OT-style violable constraints. While this theory can be considered somewhat weaker than the strict Frisch *et al.* metric, as it requires additional language-specific details, using relativized OCP constraints gives the OT approach more flexibility in describing languages that appear to be weighing phonetic features differently.

²⁰ Coetzee & Pater measure the r^2 -value for the regression analyses on the correlation between cooccurrence rate and Frisch, et al.-style similarity based on shared natural classes to be only .22.

1.4.3 MEASURING SIMILARITY AVOIDANCE IN BENGALI ROOTS

Frisch *et al.* (2004) examines the correlation between consonant cooccurrence within roots and the number of natural classes shared between the cooccurring (or *not* cooccurring) consonants. If consonant cooccurrence is restricted cross-linguistically by some form of the OCP, then we can predict that speakers' grammars correlate consonant cooccurrence with consonant similarity cross-linguistically. Since Frisch *et al.* used their interpretation of consonant similarity to describe cooccurrence restrictions, one can presume that cooccurrence restrictions can say something about consonant similarity. In other words, the O/E ratio of consonants within roots can make predictions about which consonants are considered more similar in that language.

Mallik *et al.* (1998) investigates the frequencies of every word, root, prefix, suffix, syllable, cluster, phoneme, and grapheme printed in selected editions of seven leading Bengali-language newspapers and in selected works of popular Bengali-language juvenile literature printed over the course of the year 1982. The findings of Mallik *et al.* (1998) can make predictions as to which phonemes have restrictions set against their cooccurrence with /t/, by examining which consonants are underrepresented in words that include the phoneme /t/. Assuming this cooccurrence restriction is applied in Common East Bengali fixed segment reduplication, consonants that have a low O/E ratio in words that include /t/ will be considered similar to /t/, thus triggering

similarity avoidance from fixed segment /t/ in reduplicants of words beginning with that consonant.

2 PREDICTIONS

The theories presented in Section 1.4 make predictions as to what base-initial consonants Bengali speakers will consider overly similar to the default fixed segment /t/. I shall first discuss the background of the language and specific dialect studied here and then discuss how each theory can make predictions on the fixed segment echo reduplication patterns of this dialect.

2.1 LANGUAGE BACKGROUND: BENGALI বাংলা

One of the easternmost Indo-European languages, Bengali (native name: বাংলা [banla]) is an Indic tongue spoken by over 171 million people in the People's Republic of Bangladesh and the neighboring Indian states of West Bengal, Assam, and Tripura. According to the Summer Institute for Linguistics Ethnologue Survey, it is the fifth most widely-spoken language worldwide. The international importance of Bengali is evident in its use as a lingua franca across the eastern Indian subcontinent, with 40 million second-language speakers (Gordon 2005).

2.1.1 DIALECT BACKGROUND: COMMON EAST BENGALI

The two dialects of Bengali presented to the participants in this study are Choltibhasha (literally ‘current language’), the official language of Bangladesh and the national language of Indian Bengal, and Common East Bengali, spoken primarily by inhabitants of central Bangladesh east of the Padma River. Choltibhasha is based on the local speech of the towns near Kolkata²¹, India, and is used in all formal and official discourse in both Bangladesh and Indian Bengal. Loosely based on the local dialect of towns outside Dhaka, Bangladesh, Common East Bengali is used in everyday situations across the whole of Bangladesh and parts of India as it is easily understood by speakers of other local dialects. Common East Bengali is not taught in schools or heard often via the media, save for some television and radio shows depicting everyday life in Bangladesh. There is an enormous dearth of information both on and in Common East Bengali, as most Bengali speakers only recognize the Shadhubhasha (literally ‘saintly language’, henceforth High Literary Bengali) and Choltibhasha (henceforth Standard Kolkata Bengali) dialects as appropriate for written and formal communication or for linguistic study. Therefore, all written materials used in this study were translated into English, Standard Kolkata Bengali, and Common East Bengali. As the focus of this study is based on the alternations in Common East Bengali, all following language information is specific to this dialect.

2.1.1.1 PHONEMIC INVENTORY

²¹ Formerly known as Calcutta.

Stops and affricates in most dialects of Bengali contrast in a four-way distinction of voiceless unaspirated, voiceless aspirated, voiced unaspirated, and voiced aspirated (breathy). Most fricatives are voiceless and all sonorants are voiced in Bengali.

The phonemic inventory of Common East Bengali is not fixed; any two speakers can have slightly different inventories given their exposure to other dialects, as no speaker of Bengali is truly “monodialectal”. Some consonants, in particular, can have a number of pronunciations depending on the individual speaker and the register in which the utterance is made. For example, there exist two phonemes /tʰ/ and /s/ for many speakers, especially in more formal situations. This is particularly true for those who are particularly familiar with the standard dialects of Bengali. However, a large number of speakers will pronounce both phonemes as [s] in almost all phonological contexts, especially in more informal situations. The majority of speakers, however, have only one phoneme /tʰ/, which can alternate between [s] and [tʰ], either in free variation or with some degree of allophony. Because of this variation, the charts provided below may include more phonemes and/or allophones than any one speaker may have; the charts are only meant as a guide and not as an accepted fact for all speakers of this fluid dialect.

2.1.1.1.1 *Consonants*

CONSONANTS	Bilabial	Labio-Dental	Lamino-Dental	Apico-Alveolar	Apico-Postalveolar	Lamino-Postalveolar	Velar	Glottal
<i>Stops</i>	p b b ^h		ṭ ḍ ṭ ^h ḍ ^h		ṭ ḍ ṭ ^h ḍ ^h		ḳ ɡ̣ ḳ ^h ɡ̣ ^h	
<i>Affricates</i>						tʃ dʒ tʃ ^h dʒ ^h		
<i>Fricatives</i>		f		s z		ʃ		h
<i>Approximants</i>				l	ɹ			
<i>Nasals</i>	m			n			ŋ	

Figure 9. Rough sketch of the consonantal inventory of Common East Bengali.

The stops listed as “apico-postalveolar” in

Figure 9 have been described as “retroflex” in Ramaswami (1999) and several other sources, “retroflex alveolar” in Ray *et al.* (1966:6), “pronounced at a lower position [than retroflex or cerebral], approaching the alveolar region” in Chatterji (1970:xxxiii), “not true retroflex; the tip of the tongue is slightly curved back at the point of articulation, which is a little further back than for that of the usual English sounds of ‘t’ and ‘d,’” in Haldar (1986:22), and “simply alveolar as they are articulated in the alveolar region and not in the retroflex or cucuminal [sic] region” in Tunga (1995:139), adding that “true retroflex stops are found only in Dravidian and Munda languages”. Palatographic evidence presented in Hai (1960) indicates that these stops are “alveolo-retroflex”, as they are realized “with the tip of the tongue curled back against the point of contact at the alveolar ridge (Hai 1960:186).”

What are labeled in

Figure 9 as “lamino-postalveolar” affricates and fricatives are characterized as “palatal affricates” in Ramaswami (1999), “apico-dental hissing sibilants freely varying with affricated plosives” in Ray *et al.* (1966:81), “palatal...made with the front of the tongue” in Chatterji (1970:xxxii), “dental affricates” in Haldar (1986:26), and “dental-palatal plosives” or “dorso-alveolar” affricates in Tunga (1995:131). The palatographic evidence presented in Hai (1960) indicates that the articulation of these consonants is dorso-alveolar, adding that they “are pronounced with the tip of the tongue down and the blade of the tongue spread out well against [the] teeth ridge and not against the palate (Hai 1960:185).”

The status of the apico-alveolar fricatives in Bengali dialects is a debated topic. Free variation and allophony between affricates and these fricatives occurs in the speech of some speakers, while for other speakers, several phonemic distinctions are maintained. Common East Bengali [z] is an allophonic variant (for most speakers) of /dz/ and /dzʰ/. Similarly, Common East Bengali [s] is an allophonic variant (for most speakers) of /tʃ/ and /tʃʰ/. There are, however, separate phonemes /s/ and /z/ for other speakers, apart from allophones of the affricates.

The aspirated sonorants /mʰ/, /nʰ/, /ɳʰ/, /ɽʰ/, and /lʰ/ are extremely rare and are not listed in the phoneme inventory chart; as described in Bykova (1981:26) they “occupy the peripheral part of the Bengali phonological system. Their existence...is due mainly to the contacts between the Bengali phonological system proper and those of

other languages.” The status of the aspirated nasals /m^h/ and /n^h/ is considered in Esposito *et al.* (2005) to be intermediate between phonemic singletons and clusters of consonants.

There is a small number of significant differences between the inventory in Figure 9 and the phonemic inventory of Standard Kolkata Bengali. Many of the differences are one-to-one correspondences across the dialects; Common East Bengali /f/, for example, corresponds to Standard Kolkata Bengali /p^h/. Other differences, however, involve collapses in phonemic distinction across dialects; Standard Kolkata Bengali phonemes /ɹ/ and /ɹ̥/, for example, are collapsed in Common East Bengali as two allophones of /ɹ/. These cross-dialectal differences can be very salient for some speakers, and depending on the situation, a speaker may be very conscious of his or her particular pronunciation. For the purposes of this study, audio recordings and written materials were provided in both Common East Bengali and Standard Kolkata Bengali.

2.1.1.1.2 Vowels

VOWELS	<i>Front</i>	<i>Central</i>	<i>Back</i>
<i>High</i>	i		u
<i>High-Mid</i>	e		o
<i>Low-Mid</i>	æ ²²		ɔ
<i>Low</i>		a	

Figure 10. *The vocalic inventory of Common East Bengali.*

²² This vowel is often transcribed /ɛ/ in the literature on Bengali phonology.

Both Standard Kolkata Bengali and Common East Bengali have seven contrastive vowels, distinguished by four levels of tongue height and three positions of tongue backness. Vowel qualities are less disparate than consonantal realizations across the dialects, although western dialects such as Standard Kolkata Bengali tend to make use of more vowel harmony processes than eastern dialects such as Common East Bengali. This does not increase or decrease the size of the vocalic phonemic inventory, but it does mean that vowels in certain words may not be consistent across dialects. Standard Kolkata Bengali has twice as many vowels as Common East Bengali when nasalization is factored in, as nasalization can be a phonemically distinguishing feature for all seven vowel qualities. Nasalized vowels in western dialects are most often the result of historically oral vowels followed by nasal consonants, a sequence type that is still preserved in eastern dialects.

2.2 APPLYING THE NATURAL CLASSES METRIC TO BENGALI FIXED SEGMENT ECHO

REDUPLICATION

As seen in example (10) in Section 1.1.1.1, Bengali echo fixed segment reduplication typically involves replacing the initial consonant in the reduplicant with the fixed segment, which is most often /t/. This default fixed segment /t/ is replaced by backup fixed segments in circumstances where the base already begins with /t/, a situation that would otherwise create a homophonous base-reduplicant pair.

According to the Frisch *et al.* (2004) theory, the consonants that share the largest number of natural classes are to be considered the most similar to each other. If the Bengali speakers that participated in the pilot survey are choosing fixed segments based on this measure of similarity, we can make a prediction as to what consonants will be considered more similar to /t/. Using the programs FeaturePad and Similarity²³ to apply the natural classes metric (Frisch, Pierrehumbert, and Broe – or Frisch *et al.* – metric) to the Bengali phoneme inventory²⁴, consonant similarity to /t/ is predicted to follow as shown in Figure 11.

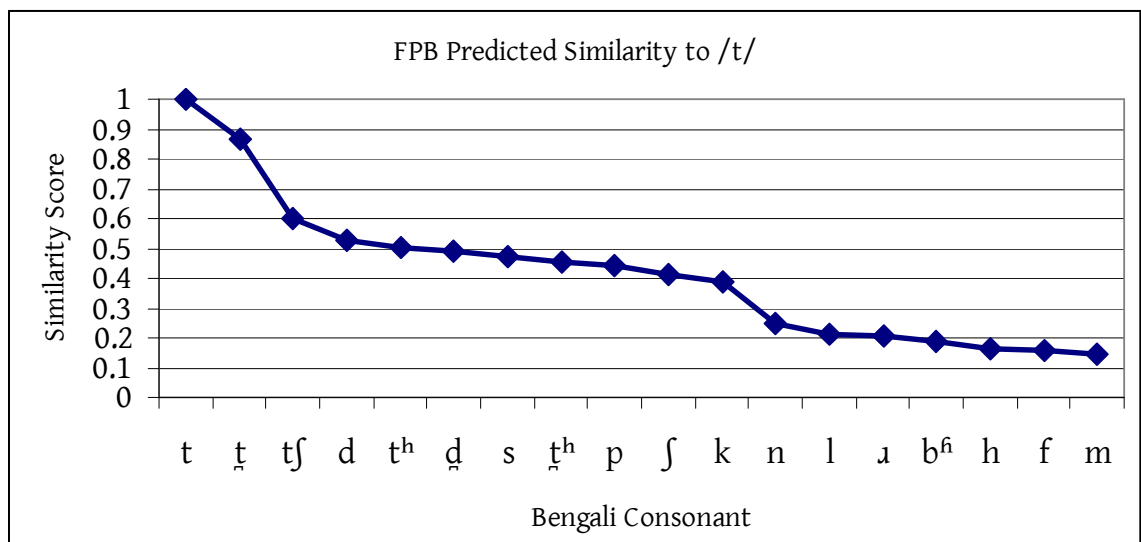


Figure 11. Similarity values of selected Common East Bengali consonants to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch *et al.* shared natural classes metric, and as calculated by FeaturePad and Similarity.

²³ Zuraw (1998) and (n.d.), respectively.

²⁴ The feature specifications of each Bengali phoneme can be found in Appendix C (Section 5.3).

The Frisch *et al.* metric calculates that the most similar consonants in Bengali to apico-alveolar²⁵ /t/ are (in order of decreasing similarity) apico-alveolar /t/ (similarity score 1.0), dental /t̪/ (0.87), affricate /tʃ/ (0.60), voiced alveolar /d/ (0.53), aspirated /tʰ/ (0.50), and so on. If Bengali speakers interpret this metric of similarity as a strong basis for deciding when to apply similarity avoidance in fixed segment echo reduplication, the prediction is that /t/-initial bases would accept the fewest reduplicants with fixed segment /t/, while /t̪/-initial bases would accept some reduplicants with fixed segment /t/, and /tʃ/-initial bases would accept even more, and so on. Bases beginning with /h/, /f/, or /m/ would accept most, if not all, reduplicants with fixed segment /t/, as they are measured to be the least similar consonants to /t/.

The Frisch *et al.* (2004) study, in fact, does not measure similarity across all major places of articulation, however. Instead, it measures similarity only within each major place grouping; thus Figure 12 graphs consonant similarity to /t/ within the coronals – a large set that includes lamino-dentals, apico-alveolars, apico-(post)alveolars, and lamino-(post)alveolars. In the measurements for Arabic, Frisch *et al.* found that it was best to measure coronal obstruents and coronal sonorants separately,

²⁵ Apico-alveolar was chosen over apico-postalveolar for the articulatory description of /t/, as the latter has no equivalent in traditional feature representations of coronal stops ([–distributed, +anterior] represents apico-alveolar, [+distributed, +anterior] represents lamino-postalveolar, and [–distributed, –anterior] represents retroflex). Many other combinations of consonants were considered; this configuration (lamino-dental, apico-alveolar, and lamino-postalveolar) was chosen both because its predictions are closest to the Bengali data when only coronals are measured and because of its close match with the palatographic evidence presented in Hai (1960). See Appendix A for a complete feature specification list of the phonemes of Common East Bengali.

claiming that they belong to two very distinct groups when examining the cooccurrence data; thus, Figure 13 graphs similarity to /t/ within Common East Bengali coronal obstruents, thus removing /n/, /l/, and /ɹ/ from the set.

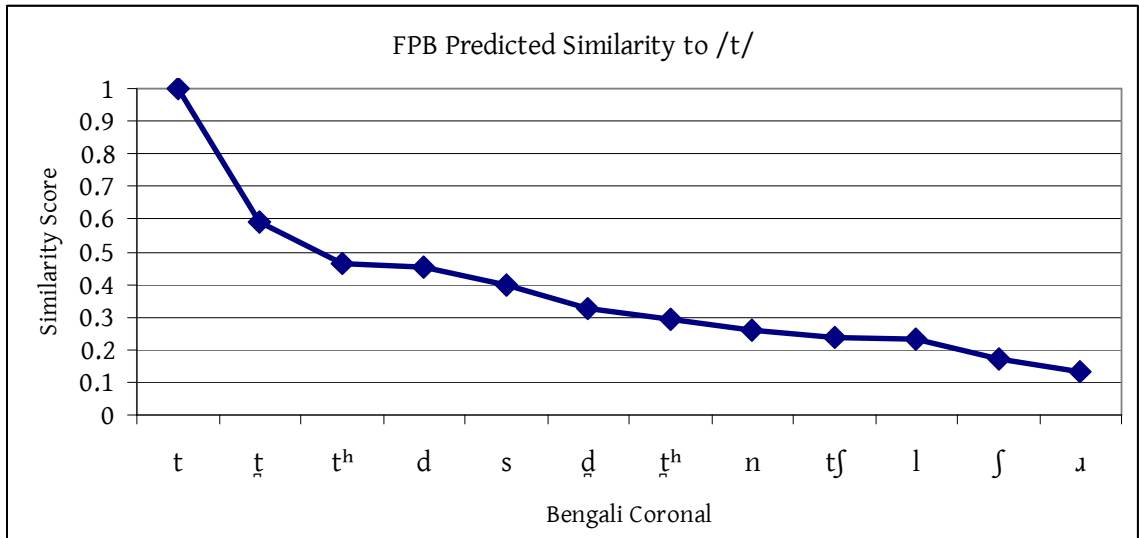


Figure 12. Similarity values of selected Common East Bengali coronals to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch et al. shared natural classes metric, and as calculated by FeaturePad and Similarity.

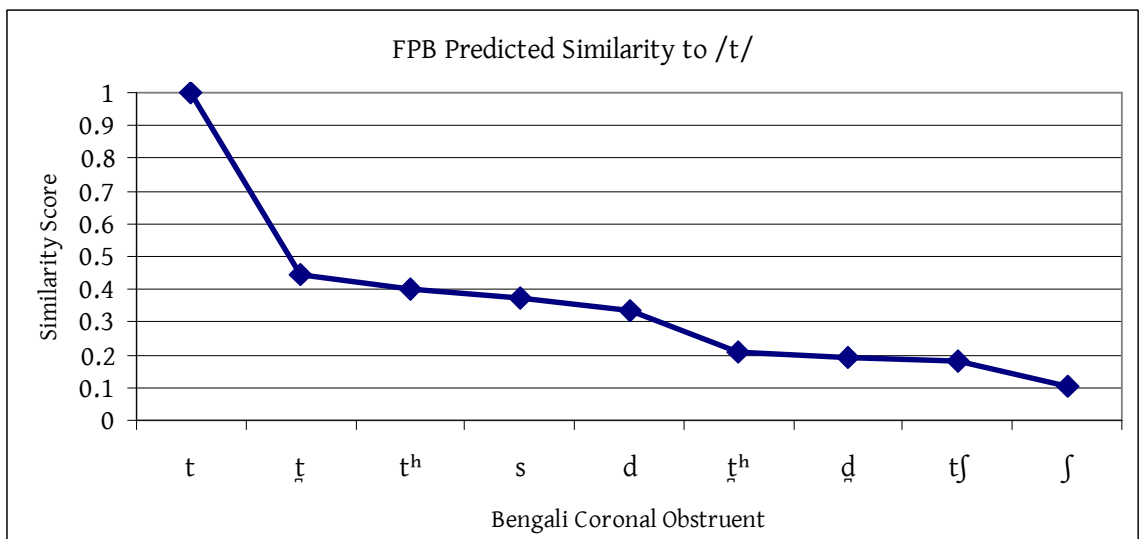


Figure 13. *Similarity values of selected Common East Bengali coronal obstruents to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch et al. shared natural classes metric, and as calculated by FeaturePad and Similarity.*

It may at first glance be surprising to see that the order of consonants across the similarity calculations differs depending on the size of the set measured. For example, when measuring similarity across all Bengali phonemes (as shown in Figure 11), /s/ is considered less similar to /t/ (similarity score 0.47) than /d/ (0.53), /t^h/ (0.50), and /ɖ/ (0.49) are to /t/. However, when measuring similarity only within the coronal place of articulation (as shown in Figure 12), /s/ is considered more similar to /t/ (similarity score 0.40) than /ɖ/ is to /t/ (0.33). Furthermore, in the similarity measurements of just the set of coronal obstruents (as shown in Figure 13), /s/ is considered more similar to /t/ (similarity score 0.38) than either /d/ (0.33) or /ɖ/ (0.19) is to /t/. Although similarity scores fall overall when the set of phonemes measured shrinks, some pairs of consonants (e.g. {t, s}) will be promoted through these changes to be considered more similar than other pairs (e.g. {t, d}) in measurements of larger sets of phonemes.

This decrease in similarity can arise due to the change in inventory size, as a natural class loses its significance when it includes within it the entire inventory of phonemes in a given set. In the example of {d,t}, measuring similarity across the entire Bengali inventory places both phonemes in the natural classes [+coronal] (the class of coronals), [-sonorant] (the class of obstruents), and [-delayed release] (the class of stops), among several others. However, when measuring similarity within the coronal set only, there is no longer a natural class [+coronal], leaving {d,t} with one natural

class less than in the larger set. When restricting similarity measurements to coronal obstruents, the natural class [-sonorant] also vanishes, and the pair {d,t} is once again considered further dissimilar. While some phoneme pairs fall sharply in similarity when decreasing inventory size, others fall at a slower rate or simply do not fall, allowing the pair {s,t} to become more similar in measurements of smaller inventories.

2.2.1 PREDICTED FIXED SEGMENT /t/-USE IN REDUPLICANTS OF NATIVE BENGALI BASES

In their discussion of OCP effects in Arabic, Frisch *et al.* examine the relation between the number of shared natural classes (Shared Natural Classes) between any pair of consonants and the observed cooccurrence of those two consonants in trilateral roots. Using this Shared Natural Classes metric, a prediction can be made about fixed segment /t/ use in Bengali echo reduplication. Those bases that begin with consonants sharing a high number of natural classes with /t/ will have a higher tendency not to have reduplicants that begin with /t/, in order to maintain similarity avoidance. The following graph (copied from Figure 11) illustrates the similarity score between /t/ and each consonant of the Bengali inventory.

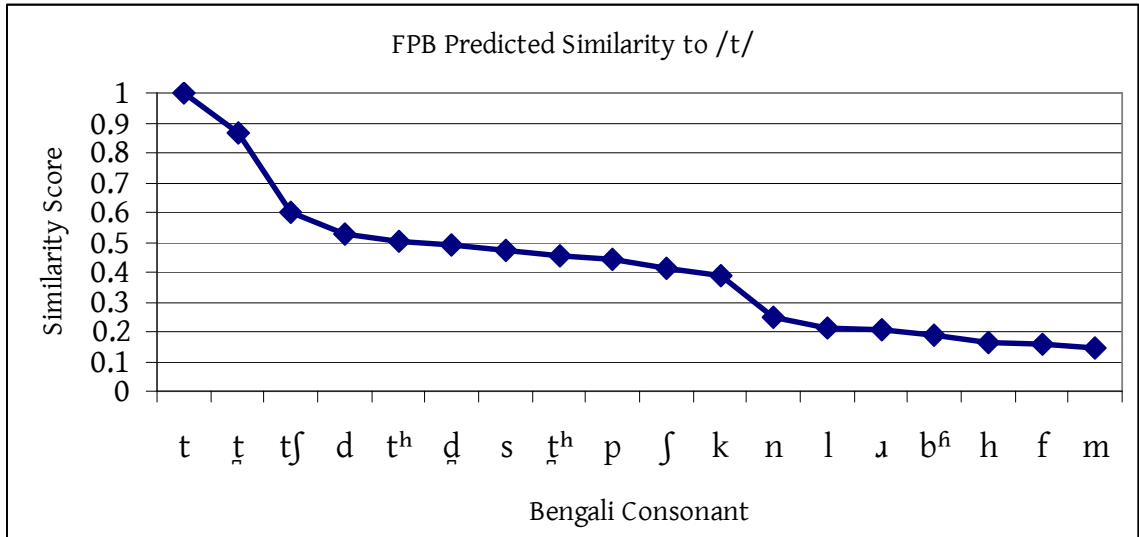


Figure 14. Similarity scores of native Bengali consonants to apico-alveolar /t/. Same as Figure 11.

Subtracting this score from one gives us what we shall call the Dissimilarity Score for each consonant to /t/. This measurement is provided below:

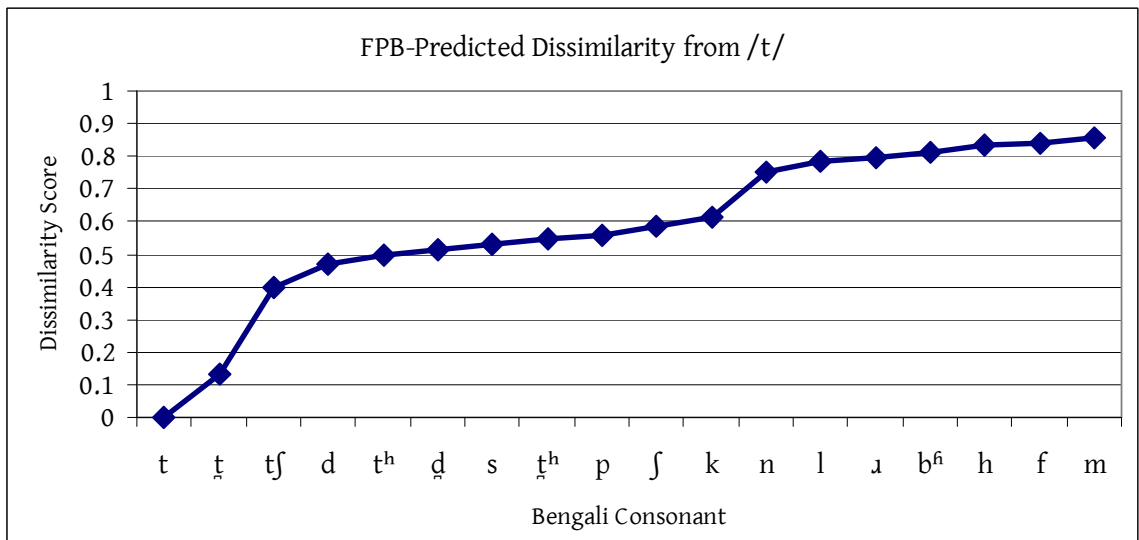


Figure 15. Dissimilarity scores of native Bengali consonants to apico-alveolar /t/.

Figure 15 shows that consonants such as alveolar /t/, dental /t̪/, and affricate /tʃ/ should be considered relatively similar to alveolar /t/, with dissimilarity scores of 0, 0.13, and 0.40, respectively. All other consonants, including all voiced consonants, aspirated consonants, continuants, and non-coronals will be considered relatively dissimilar from alveolar /t/, with dissimilarity scores all above 0.5.

If similarity avoidance in Bengali fixed segment echo reduplication works on a gradient scale with respect to similarity to /t/, then we can assume that Figure 15 can also be interpreted as a scale of predicted fixed segment /t/-use in reduplicants of words starting with the consonants marked along the x-axis. Thus, /t/-initial words are predicted to be reduplicated with fixed segment /t/ the least, with dental /t̪/-initial words allowing some reduplicants with fixed segment alveolar /t/, and /tʃ/-initial words allowing more. Words starting with other consonants are predicted to allow reduplicants with fixed segment /t/ much more often, with /h/-initial words predicted to boast the most reduplicants with /t/ as their fixed segment.

2.2.2 PREDICTED FIXED SEGMENT /t/-USE IN REDUPLICANTS OF BORROWED ENGLISH BASES

English words have been borrowed into the Bengali language for centuries; numerous loanwords have become totally incorporated into the Bengali lexicon and have replaced native words or older loanwords from Sanskrit, Persian, Arabic, or Portuguese. Many loanwords from English and other languages have become well

enough established in the lexicon to undergo fixed segment echo reduplication. To investigate whether borrowings would be treated the same as native words of similar structure, English loanwords were included in the current study.

Except for a small set of loans that have been obscured by centuries of reanalysis and phonological change (e.g. [haʃpaʈal] ‘hospital’, [tali] ‘tile’, [gelaʃ] ‘drinking glass’, etc.), most English words are borrowed into Bengali using a simple one-to-one correspondence between English and Bengali phonemes. The set of phonemes used to pronounce English words in Bengali is a subset of the phonemes used to pronounce native words.

Figure 16 illustrates the subset of the Bengali phonemic inventory as used for English borrowings. In this study, this phoneme set will henceforth be considered a feature of “Bengali English”.

CONSONANTS	Bilabial	Labio-Dental	Lamino-Dental	Apico-Alveolar	Apico-Postalveolar	Lamino-Postalveolar	Velar	Glottal
Stops	p b (p ^h)(b ^h)		t̪ d̪		t d		k g	
Affricates						tʃ dʒ (tʃ ^h)		
Fricatives		f v		s z		ʃ		h
Approximants				l	ɹ			
Nasals	m			n			ŋ	

Figure 16. Rough sketch of the consonantal inventory of “Bengali English”. Parentheses surround allophonic variations of “Bengali English” phonemes, as described in Footnote 26.

Most of the consonantal phonemes of English have rough Bengali equivalents used in “Bengali English”. Some of the important distinctions between the phonemic inventory of “Bengali English” and that of English as spoken by native English speakers are as follows:

- 1) English /θ/ is borrowed into “Bengali English” as /t̪ʰ/,
- 2) English /ð/ is borrowed into “Bengali English” as /d̪/,
- 3) English /z/ and /dz/ are collapsed in “Bengali English” as /dʒ/,
- 4) English /z/ is borrowed into “Bengali English” as /z/ or /dʒ/,²⁶
- 5) English /s/ is borrowed into “Bengali English” as /s/, /ʃ/, or /tʃʰ/,
- 6) English /f/ is borrowed into “Bengali English” as /f/ or /pʰ/, and
- 7) English /v/ is borrowed into “Bengali English” as /v/ or /bʱ/.

Considering the fact that English words are borrowed into Bengali using native Bengali phonemes to represent the English sounds, it would be predicted that English words rendered in “Bengali English” should show the same pattern of relation between fixed segment /t̪/-use and similarity avoidance as native Bengali words. If, however, Bengali speakers access the native English phonemic inventory when confronted by an English borrowing, then the similarity values, and consequently, the predicted fixed segment /t̪/-use could be very different from the native Bengali pattern. Another possibility is that speakers use neither the native Bengali similarity values nor the native English similarity values to determine the fixed segment /t̪/-use for borrowed

²⁶ The variation in the pronunciation of the borrowed English fricatives /f/, /v/, /s/, and /z/ is dependent on the status of the phoneme in the individual speaker’s Bengali phonology. For example, many speakers pronounce /s/ as [s] in clusters and [ʃ] elsewhere, although they also have a separate /ʃ/ phoneme. These speakers would likely apply this distribution to English words that include /s/.

English bases, but that they use calculations based solely on the phonemic inventory of “Bengali English” – the restricted subset of Bengali phonemes applied in borrowings from English. The corresponding dissimilarity values (based on the Frisch *et al.* model) and predicted fixed segment /t/-use for the three predictions are shown below.²⁷

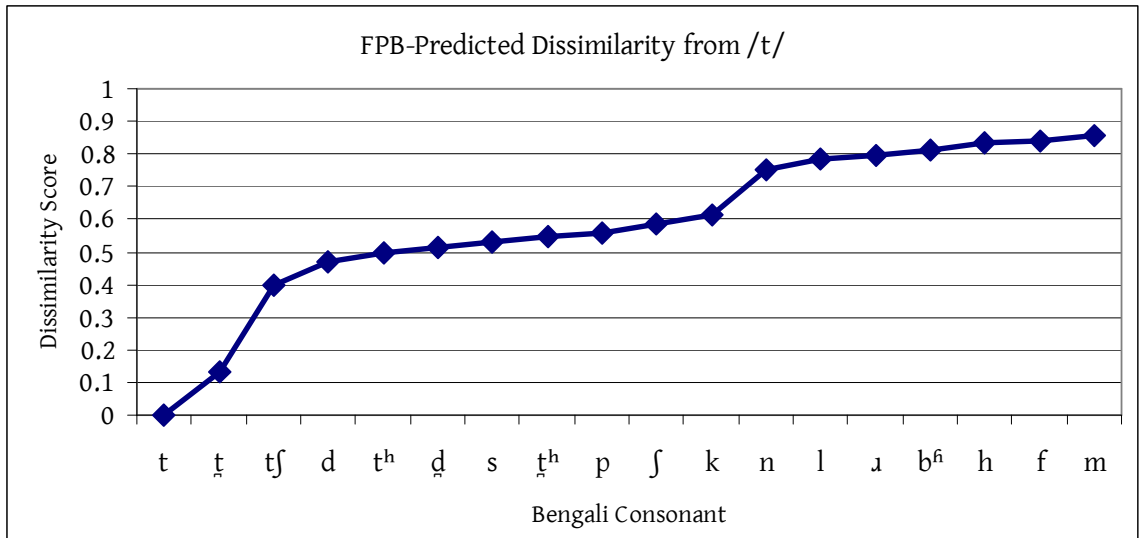


Figure 17. Dissimilarity values of selected Bengali consonants to the voiceless unaspirated alveolar stop /t/, as measured by the Frisch *et al.* shared natural classes metric, and as calculated by FeaturePad and Similarity. This is the same graph seen in Figure 15.

²⁷ Dissimilarity is measured by subtracting the similarity score from one.

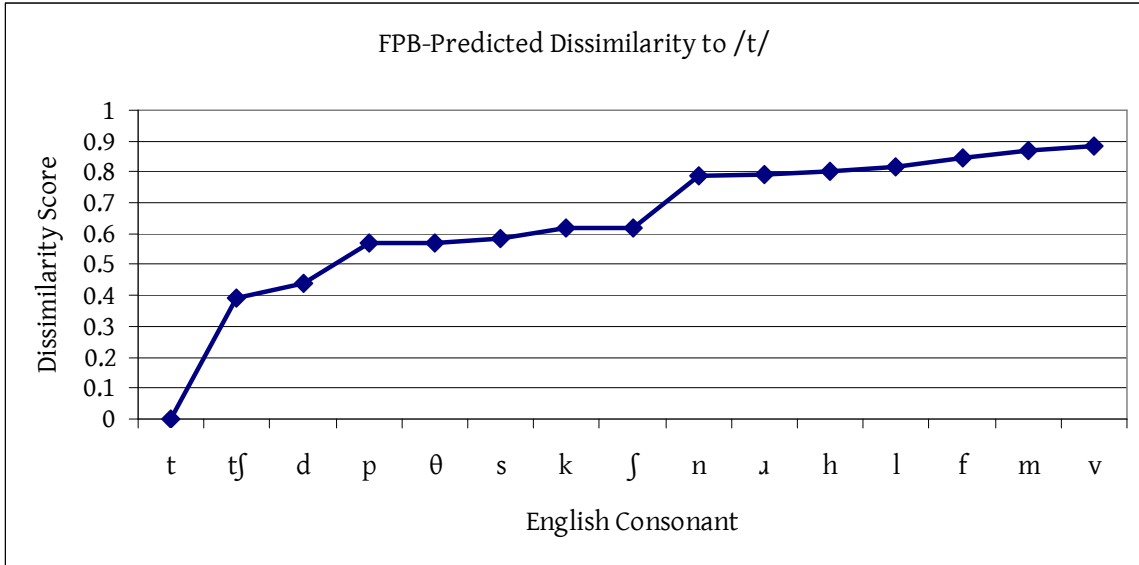


Figure 18. Dissimilarity values of selected English consonants to the voiceless alveolar stop /t/, as measured by the Frisch et al. shared natural classes metric, and as calculated by FeaturePad and Similarity.

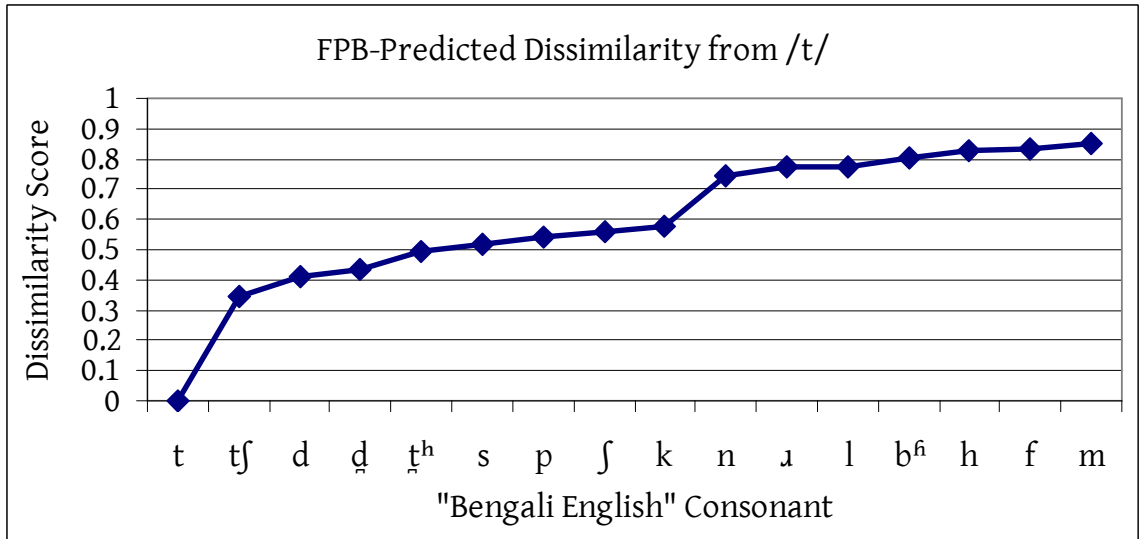


Figure 19. Dissimilarity values of selected “Bengali English” consonants to the voiceless alveolar stop /t/, as measured by the Frisch et al. shared natural classes metric, and as calculated by FeaturePad and Similarity. The phonemes of “Bengali English” are a subset of the Bengali inventory.

2.3 APPLYING RELATIVIZED OCP CONSTRAINTS TO BENGALI FIXED SEGMENT ECHO

REDUPLICATION

Coetzee & Pater's analysis of Muna makes use of OCP constraints relativized within each major place of articulation for features such as [sonorant] and [nasal] to describe the cooccurrence restrictions within (VCV)CVCV roots. Coetzee & Pater propose ranking more basic OCP constraints such as OCP-COR below more complex OCP constraints such as OCP-COR ([-sonorant], [α continuant]), as it was found that the restriction against two coronals within a word is not as strong as the restriction against two coronal obstruents agreeing in continuancy.

In the Bengali case, /t/-initial bases are expected to allow exceptionally few reduplicants with fixed segment /t/, while bases starting with consonants similar to /t/ (e.g. /t^h/, /t̪/, /d/, etc.) will allow more fixed segment /t/-reduplicants. Bases starting with consonants judged very dissimilar from /t/ (e.g. /m/, /b^h/, etc.) will allow the most fixed segment /t/-reduplicants. It can be assumed that the more feature combinations a base-initial consonant shares with /t/, the fewer fixed segment /t/-reduplicants it will allow. Of course, all of the constraints will be violable, and some will have no effect on the candidate, to allow some instances of /t/-initial bases cooccurring with fixed segment /t/-reduplicants, etc. This fact separates the Muna analysis from the Bengali analysis: while exceptions to Muna root-internal cooccurrence can be explained with the use of lexically-specific faithfulness constraints, Bengali similarity avoidance in

fixed segment reduplication cannot be explained in this way. Bengali speakers do not all agree on which bases allow fixed segment /t/ and which do not, and they are not even internally-consistent either, as the same base can take more than one reduplicant at times. Any adaptation of the relativized OCP in Bengali fixed segment reduplication will have to make use of stochastic OT or other such theory, in which constraint rankings are only tendencies, allowing for some exceptions.

A very conservative application of Coetzee & Pater's theory to the Bengali fixed segment reduplication data would posit an OCP constraint against the cooccurrence of any two phonemes that share all the features of /t/ (e.g. [-continuant], [-sonorant], [-spread glottis], etc.). The next constraint down in the hierarchy would assign a violation to every base beginning with a consonant that shares all features with /t/ minus one (e.g. the [spread glottis] feature). A lower-ranked constraint would assign a violation to every base beginning with a consonants that shares all features with /t/ minus two, and so on, until the general OCP-coronal constraint, which assigns a violation to any base-reduplicant pair where the base-initial consonant is coronal.

Each of the constraints described above would allow for exceptions, with the higher-ranked constraints allowing far fewer exceptions than those ranked at the bottom of the hierarchy. Table 1 illustrates a possible constraint set given these parameters. Note how each higher-ranked OCP constraint assigns a violation to each

candidate in which yet another feature is shared between the base-initial consonant and the fixed segment /t/.²⁸

OCP-COR (α s.g., α voi, α dist, α d.l., α ant, α son, α nas, α lat)²⁹ »

OCP-COR (α voi, α dist, α d.l., α ant, α son, α nas, α lat) »

OCP-COR (α dist, α d.l., α ant, α son, α nas, α lat) »

OCP-COR (α d.l., α ant, α son, α nas, α lat) »

OCP-COR (α ant, α son, α nas, α lat) »

OCP-COR (α son, α nas, α lat) »

OCP-COR (α nas, α lat) »

OCP-COR (α lat) »

OCP-COR

Table 1. *Possible relativized OCP constraint hierarchy for Common East Bengali fixed segment reduplication similarity avoidance.*

²⁸ There are of course many combinations of features that could have been considered in an appropriate constraint hierarchy. The particular combinations of features used in this hierarchy was determined to be the best fit with the data, as seen in Section 3.3.2.

²⁹ Abbreviations used: COR = coronal, s.g = spread glottis., voi = voice, dist = distributed, d.l. = delayed release, ant = anterior, son = sonorant, nas = nasal, and lat = lateral.

Applying these constraints to Bengali fixed segment reduplication, certain consonants would receive not only more but also more severe violations when cooccurring with /t/. The consonants that would receive violations from each constraint are shown in

Table 2 below.

Each constraint below...	...assigns one violation to each consonant below when cooccurring with /t/
OCP-COR (α s.g., α voi, α dist, α d.l., α ant, α son, α nas, α lat)	t
OCP-COR (α voi, α dist, α d.l., α ant, α son, α nas, α lat)	t, t ^h
OCP-COR (α dist, α d.l., α ant, α son, α nas, α lat)	t, t ^h , d, d ^h
OCP-COR (α d.l., α ant, α son, α nas, α lat)	t, t ^h , d, d ^h , ṭ, ṭ ^h , ḍ, ḍ ^h
OCP-COR (α ant, α son, α nas, α lat)	t, t ^h , d, d ^h , ṭ, ṭ ^h , ḍ, ḍ ^h , s, z
OCP-COR (α son, α nas, α lat)	t, t ^h , d, d ^h , ṭ, ṭ ^h , ḍ, ḍ ^h , s, z, tʃ, tʃ ^h , dʒ, dʒ ^h , ʃ
OCP-COR (α nas, α lat)	t, t ^h , d, d ^h , ṭ, ṭ ^h , ḍ, ḍ ^h , s, z, tʃ, tʃ ^h , dʒ, dʒ ^h , ʃ, ɹ
OCP-COR (α lat)	t, t ^h , d, d ^h , ṭ, ṭ ^h , ḍ, ḍ ^h , s, z, tʃ, tʃ ^h , dʒ, dʒ ^h , ʃ, ɹ, n
OCP-COR	t, t ^h , d, d ^h , ṭ, ṭ ^h , ḍ, ḍ ^h , s, z, tʃ, tʃ ^h , dʒ, dʒ ^h , ʃ, ɹ, n, l

Table 2. Bengali consonants that would receive a violation from each relativized OCP constraint, when cooccurring with /t/.

Pairing each Bengali consonant with fixed segment /t/ and applying the constraint hierarchy in Table 1 would predict that base-initial /t/ would cooccur with fixed segment /t/ the least, as this pair violates the highest-ranked OCP constraint. Base-initial /t^h/ (*i.e.* the segment that shares all features with /t/ except for [spread glottis]) would cooccur with fixed segment /t/ the “next least”, as this pair violates the second-highest-ranked OCP constraint. The hierarchy would not assign violations to fixed segment /t/ in cooccurrence with /p/- or /k/-initial bases, as OCP constraints are by the Coetzee & Pater definition (and by most traditional definitions) restricted to major place features.

If each constraint is assigned a value measuring its effectiveness in this hierarchy, with the highest-ranked constraint receiving a value of 9 and the lowest-ranked constraint receiving a value of 1, we can calculate which consonants would be considered more similar to (*i.e.* would violate more OCP constraints when cooccurring with) /t/. Similarity would be calculated by dividing the numerical value of the highest constraint violated by the consonant by the numerical value of the highest constraint value possible (9), and multiplying this by 100. The similarity of {t,t} would thus be $(9/9) \times 100 = 100$, and the similarity score for {l,t} would be $(1/9) \times 100 = 11.1$. Subtracting the similarity score from 100 produces the dissimilarity score; Figure 20 below measures the dissimilarity score for each consonant with /t/.

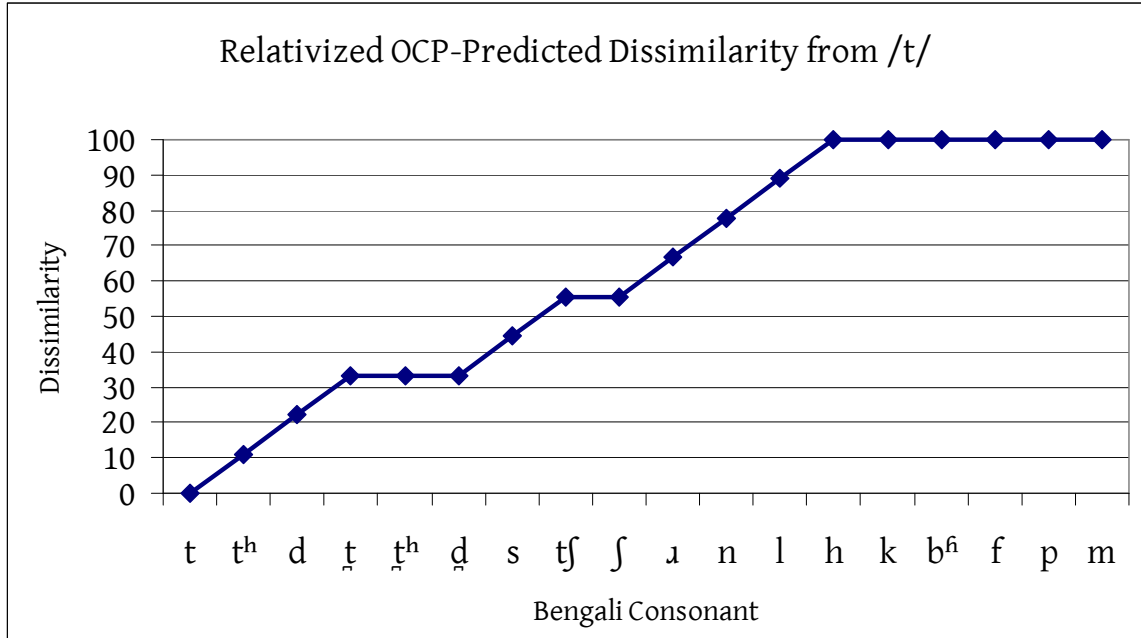


Figure 20. Relativized OCP-predicted dissimilation pattern given the constraint hierarchy in Table 1. Dissimilarity is measured crudely by assigning a numerical value (n) to the constraints in Table 1. When cooccurring with /t/, each consonant gets a dissimilarity score by subtracting its similarity score ($100 \times (n/9)$) from 100, where (n) is the value of the highest-ranked constraint the pair violates. The predicted acceptance of fixed segment /t/ would be roughly equivalent to the dissimilarity measurements here.

Note that given the hierarchy in Table 1, all dental consonants would be considered equally dissimilar from /t/, as they all violate the same exact OCP constraints. The voicing and aspiration distinctions would be irrelevant, as the features associated with those distinctions (*i.e.* [voice] and [spread glottis]) are only relevant in constraints that already specify for tongue tip configuration (*i.e.* the [distributed] feature), which distinguish apico-alveolars from lamino-dentals. Similarly, the lamino-(post)alveolar affricate /tʃ/ and fricative /ʃ/ are considered equally dissimilar from /t/, as the feature distinguishing them (*i.e.* [continuant]) is already irrelevant at the point in the hierarchy where the highest-ranked constraint they do not violate is placed.

2.4 COMPARING CONSONANT COOCCURRENCE AND BENGALI FIXED SEGMENT ECHO

REDUPLICATION

Both the Frisch *et al.* (2004) study and the Coetzee & Pater (2005) study relate consonant similarity based on features and consonant cooccurrence within roots; if the two are directly related universally, consonant cooccurrence in a particular language can make predictions about the similarity metric of that language. Thus, given the Mallik *et al.* (1998) study on the frequency of words, roots, morphemes, phonemes, and other structures in the Bengali language, consonant cooccurrence measurements can be made for each pair of consonants; these measurements can be considered the observed rate of cooccurrence (O), which can be compared to the expected rate of cooccurrence (E), calculated from the overall occurrence of the two sounds in those word positions, independent of the other consonant(s).

Because /t/-initial stems are relatively rare in Bengali,³⁰ consonant cooccurrence was calculated only in words where /t/ served as the intervocalic onset of the second

³⁰ Although coronal stops are normally considered rather unmarked universally, Bengali /t/ is the modern reflex of what was once retroflex /ʈ/, often considered a marginal phoneme in older Indic languages. Thus, inherited /t/-initial roots are relatively rare in the language, especially compared with the large number of /t/-initial borrowings from English and other languages. As described in Chatterji (1970), medial /t/ is more common, because of a historical process that fused rhotics and other approximants with dental /ʈ/ to create retroflex /ʈ/, which later fronted to alveolar /t/. See Steriade (1995) for a perceptual account for the underrepresentation of word-initial /t/.

(and final) syllable of the root. Only roots composed entirely of two open syllables were considered, in order to avoid the effects of codas and complex onsets. Thus, in a root composed of a [CVtV] string, the cooccurrence rate of {C, t} was measured and compared to occurrence rates of C-initial words and /t/-medial³¹ words overall. As Mallik *et al.* considered “word” to refer to whatever string appeared in print between two blanks, many instances of the same word are found, each with a separate case ending. For example, the root /g^hɔt/- ‘occur’ is listed 41 times, with 23 verbal conjugations and 18 nominal inflections and derivations. For consistency, a number of guidelines were followed when counting /t/-medial and /C/-medial words.

1. Forms that only differed in their verbal conjugation suffixes (*e.g.* /i/, /o/, /e/, *etc.*) were not considered separate entries in the calculations.

2. In general, any word in which the second syllable was wholly contained in a separate morpheme from the root was excluded from the counts. This eliminated words in which the syllable-initial /t/ was actually the initial consonant of one of the definite articles /ta/, /ti/, and /to/, as all nouns can presumably take these suffixes regardless of the consonants within the word. Furthermore, words in which the second syllable began with a case marker (*e.g.* /ke/, /ṭe/, *etc.*), a negative particle (*e.g.* /na/, /ni/, *etc.*), or an enclitic (*e.g.* /ṭo/, /dʒe/, *etc.*) were also eliminated.

³¹ Here, “medial” is used only to refer to the first consonant of the second syllable.

3. Different pronunciations of the same word were considered as one entry. For example, [nowko] and [nowka] both mean ‘boat’, with the former being associated with colloquial Kolkata pronunciation, and the latter being associated both with highly formal speech and with many rural dialects.

The calculations performed included counting the number of instances in which each Bengali consonant occurred word-initially in words of the shape CVCV. This was divided by the number of CVCV words altogether, giving the overall expected occurrence rate of each word-initial consonant. Similar calculations were made for medial /t/; the number of instances in which word-medial /t/ occurred in words of the shape CVCV was divided over the total number of CVCV words, giving the overall expected medial /t/ occurrence rate. Then the number of observed cooccurrence between /t/ and each consonant divided by the total number of CVCV words was divided over the product of the two expected scores; an example with the consonant /k/ is shown in Figure 21:

$$\begin{array}{c}
 \text{Observed \{k, t\} Cooccurrence in CVCV words} \\
 \text{-----} \\
 \text{Total CVCV words} \\
 \text{-----} \\
 \frac{\text{Observed /k/-initial CVCV words}}{\text{-----}} \quad \times \quad \frac{\text{Observed /t/-medial CVCV words}}{\text{-----}} \\
 \text{Total CVCV words} \qquad \qquad \qquad \text{Total CVCV words}
 \end{array}$$

Figure 21. Calculation for O/E ratio of cooccurrence rate of initial /k/ and medial /t/.

The O/E ratios of all consonants' cooccurrence rates with /t/ are given in Figure 22 below. A score of zero (0) indicates that despite whatever cooccurrence rate we may have expected given the word list, there were no observed cooccurrences. A score of one (1) indicates that the observed rate of cooccurrence was equal to the expected rate. Scores above one (1) indicate that the observed cooccurrence rate was higher than what was expected given the word list.

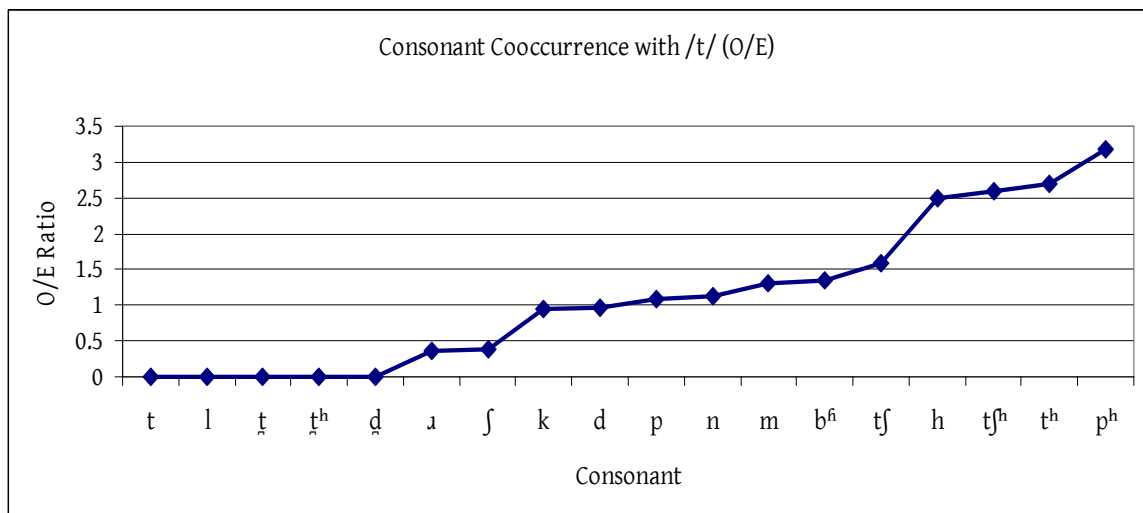


Figure 22. *Observed / Expected Cooccurrence Rates of selected Standard Kolkata Bengali phonemes with /t/, using data collected from Mallik et al. (1998).*³²

The phoneme /t/ never cooccurs with other /t/ within roots; the OCP effect is presumably very strong in maintaining identity avoidance.³³ This may be similar to the

³² Since Mallik *et al.* studied standard written Bengali, the transcriptions of a couple of phonemes listed in this graph are not exactly identical to the transcriptions used in this paper for Common East Bengali. Specifically, Mallik *et al.*'s /pʰ/ and /tʃʰ/ correspond to Common East Bengali /f/ and /s/, respectively.

avoidance of fixed segment /t/ in reduplicants of /t/-initial bases. The patterns in Figure 22 suggest that if similarity avoidance in Bengali fixed segment echo reduplication is regulated by the same OCP constraint as the apparent consonant cooccurrence restrictions are, bases starting with the consonants /t/, /l/, /t̪/, /t̪ʰ/, and /d̪/ will avoid reduplicants with fixed segment /t/. Bases starting with consonants such as /pʰ/ (Common East Bengali /f/), /tʰ/, /tʰ/ (Common East Bengali /s/), and /h/ would be predicted to allow more reduplicants with fixed segment /t/ than bases starting with other consonants. If it turns out that the Bengali grammar relates O/E cooccurrence rates between each consonant and /t/ to fixed segment /t/-use in echo reduplication, we can posit that the same constraints governing the consonant distribution in the lexicon are also regulating what reduplicants are acceptable given the initial consonant of the base.

2.5 FEATURE WEIGHTING

Shared features could also plausibly be a metric of similarity between consonants; both the Frisch *et al.* and Coetzee & Pater studies ultimately involve shared features, with Frisch *et al.*'s metric calculating similarity based on the number of natural classes shared between homorganic consonants, and Coetzee & Pater's metric calculating similarity based on numerous OCP constraints specific to particular features and

³³ Remember that the O/E calculation factors in the overall rare occurrence of /t/ in the language; even having considered this low occurrence, the {t,t} cooccurrence rate is still far lower than predicted.

feature combinations. Surely, consonants that share very few features will be considered highly dissimilar; however, counting the number of all features shared between consonants could be misleading. For example, if /t/ and /d/ share all features except for [voice], /t/ and /t^h/ share all features except for [spread glottis], and /t/ and /t̚/ share all features except for [distributed], does this mean that all three pairs are equally similar? If speakers do not seem to pattern with such a prediction, it might be more appropriate to consider certain features as having a greater effect or weight in similarity calculations than others. For example, if voiceless consonants in general seem to be considered more similar to /t/ than voiced consonants overall, then the feature [voice] would receive a stronger weight than, say, [lateral], which is largely irrelevant – it is only useful in distinguishing the behavior of one phoneme /l/, which would already be judged highly dissimilar from /t/ thanks to its values for the features [sonorant], [approximant], [articulatory continuant], [acoustic continuant], and so on.

If fixed segment choice in Common East Bengali echo reduplication is calculated by adding up the weights of all shared features between the base-initial consonant and the default fixed segment /t/, then we can examine the observed dissimilation patterns in echo fixed segment reduplication and extrapolate from these the weights of particular features, using maximum likelihood estimation³⁴ in the statistical computing environment R (R Development Core Team 2005). Specifically, the similarity of two consonants C_1 and C_2 would be calculated using Equation 1 given below.

³⁴ See Myung (2003) for a tutorial on maximum likelihood estimation.

$$sim(C_1, C_2) = \exp\left(3 \sum_{i=1}^{\# \text{ features}} w_i (1 - \delta_i(C_1, C_2))\right)$$

Equation 1. *Similarity as a decreasing function of feature disparity.*

In Equation 1, C_1 and C_2 are the two consonants being measured for similarity, w_i is the weight of the feature f_i , and $\delta_i(C_1, C_2)$ is 1 if and only if the two consonants share the same specification for feature f_i and 0 if they do not. The similarity score of a consonant and /t/ ($sim(C_1, t)$) can be plugged into an equation measuring the probability of fixed segment /t/-use given the base-initial consonant C_1 .

$$P = \left(\frac{m!}{n!(m-n)!}\right) (1 - sim(C_1, t))^n (sim(C_1, t))^{m-n}$$

Equation 2. *The probability using fixed segment /t/ in the reduplicant of a C_1 -initial base n times out of m trials is given by this binomial formula.*

In Equation 2, P stands for the probability that a base-initial consonant C_1 will undergo echo reduplication with fixed segment /t/ n times out of a total of m trials. Since similarity scores ($sim(C_1, C_2)$) range only from 0 to 1, the similarity score for the consonant pair $\{C_1, t\}$ can be subtracted from 1 to derive the dissimilarity score. If given the m and n values (the total number of reduplications for C_1 -initial bases and the total number of fixed segment /t/-reduplicants for C_1 -initial bases, respectively), a program such as R could calculate the best fit of similarity scores and feature weights.

The drawback to this model, in contrast with the other metrics mentioned in this study, is that feature weights as described here must be drawn from the observed patterns in the fixed-segment reduplication data itself. The Frisch *et al.* metric, the relativized OCP hierarchy, and the consonant cooccurrence correlation all propose that patterns in the lexicon can predict patterns in the reduplication, while feature weighting does not attempt to predict the patterns as much as it models the patterns. If feature weighting can accurately replicate the observed data, we are still left with the question of where the speaker acquires the feature weights of his or her language. Nevertheless, a successful modeling of the data could shed light on what sort of process occurs in the speaker's grammar when judging similarity, even if the source of these putative particular feature weights remains a mystery.

3 CURRENT STUDY

This current study is intended to extend the questions of the pilot survey to a wider subject pool in order to gather a more scientifically significant array of responses, increase the number and type of stimulus items in order to elicit reduplicants that can better illustrate their relationship to the similarity calculations mentioned earlier, more carefully control the effects of target consonants in non-target prosodic positions, and formalize the presentation and analysis.

3.1 METHODS

3.1.1 SUBJECTS

In total, 30 subjects participated in the current study; all participating subjects are native Bengali speakers born in Bangladesh or West Bengal, India, now residing in Southern California. All were educated in both Bengali and English while living in South Asia. Speakers were monetarily compensated for their participation. All subjects were interviewed at the beginning of the study to establish their birthplace, linguistic background, and dialectal familiarity.

3.1.2 STIMULI

The stimuli³⁵ were grouped into four lists of 60 stimuli each, yielding a total of 240 stimuli. The four lists are characterized by the source language and initial consonants of the words they include; each list has only one of the following four word types:

- 1) Native words in Standard Kolkata Bengali,
- 2) Native words in Common East Bengali,
- 3) English borrowings with Bengali pronunciation, and
- 4) Sanskrit and English borrowings with initial clusters.

3.1.2.1 STIMULUS SPECIFICATIONS

³⁵ See Appendices D, E, and F for tables of all stimulus items.

Stimuli were controlled for stress and structure; all stimulus items were disyllabic, and like most Bengali words, maintained primary stress on the initial syllable. All stimuli were nouns, infinitival verbs, or adjectives, as these are the three productive lexical categories that can undergo fixed segment echo reduplication. Most nouns were presented in nominative case, which is unmarked in Bengali, although there were some monosyllabic roots that were inflected for locative case in order to increase the number of syllables to two. For example, the English borrowing [fʌns] ‘France’ was inflected for locative case as [fʌnse] ‘in France’, creating a disyllabic stem. The echo reduplication of this stem would mean something like ‘in France and other such (European, first world, foreign) countries’. Locative declensions of nouns undergo fixed-segment reduplication fairly often. As there are no native Bengali words with initial clusters, commonly-used borrowings from Sanskrit and English had to be used in order to properly examine the effect of branching onsets in similarity avoidance. Many of these borrowings are so widely heard that Bengali speakers may not even be aware of their foreign etymological ancestry.

About half of the stimuli from each list were target words, *i.e.* words that started with a consonant predicted to be considered similar to the default fixed segment /t/. These consonants include /t/, /t^h/, /d/, /t̪/, /t̪^h/, and /tʃ/. For the cluster list, the target words started with /tɹ/, /stɹ/, /st/, /dɹ/, and /st̪/. The second half of each list included words that began with consonants and consonant clusters not predicted to be

considered similar to /t/, including /h/, /v/, /m/, /f/, /pɹ/, /pl/, /dʁ/, /tʰr/, and so on. No word included /t/, /tʰ/, /d/, /t̥/, /t̥ʰ/, or /tʃ/ in non-initial position.

3.1.2.2 RECORDING OF STIMULI

The stimuli were recorded in one sitting inside the sound booth at UCLA's Phonetics Laboratory, using a tabletop Telex M-540 microphone plugged into a Macintosh computer. An adult female native speaker produced each stimulus twice in with consistent falling intonation. The speaker produced two repetitions of ten words at a time; after each block of ten stimuli, the file was saved for later splicing. The files, recorded in .wav format on Amadeus II (Hairer, Martin, n.d.), were spliced into 240 individual stimulus files on a later date using WaveSurfer (Sjölander & Beskow 2004). Only the clearest and most easily-understandable repetition of each stimulus was saved.

3.1.2.3 RANDOMIZATION OF STIMULI

All stimuli were randomized in both the audio and written portions of the test for all speakers. Each speaker was presented with a different order of audio stimuli, although the first stimulus item in each list was kept constant. For example, all speakers presented with the English stimulus list heard [kɔfi] 'coffee' first, although all following stimuli were randomized. The first word in each list was kept constant, using extremely common words that were meant to cause as little confusion as possible, in

order to familiarize the subject with the task before proceeding to more difficult stimuli.

Microsoft Excel's random number generator was used to produce 35 different orderings of 59 stimuli (as the first stimulus of each set type was kept constant) for each of the four stimulus lists. After randomizing all 180 audio and written stimulus lists, each list was individually inspected to ensure that no two stimuli starting with the same consonant or consonant cluster occurred in sequence. Such sequences were reordered with surrounding stimuli to prevent any such instance.

3.1.2.4 PRESENTATION OF STIMULI

The experiment was held in Huntington Beach, California, at the home of a local member of the Bengali community. Each of the 30 speakers was asked in which language (*i.e.* English or Bengali) he or she would prefer to listen to directions (directions on all test materials were written in both English and Bengali), and in which dialect of Bengali (*i.e.* Standard Kolkata Bengali or Common East Bengali) he or she would prefer to read and listen to Bengali stimuli.³⁶ For all tasks, response time was not

³⁶ Out of 30 subjects, 17 chose Common East Bengali pronunciations and spellings for the purposes of the experiment, while the other 13 chose Standard Kolkata Bengali. Although their preferences for the pronunciation and spelling of stimuli varied, all subjects are native speakers of eastern Bengali dialects, and all use Common East Bengali in everyday communication with other Bangladeshi Americans. A cluster analysis of the data suggested no relationship between choice of stimulus dialect and rate of /t/-use in reduplicants.

measured; there was no time limit to complete any task. Subjects took approximately one hour each to complete the entire process.

3.1.2.4.1 *Audio Stimuli*

For all subjects, the audio elicitation (Part I) preceded the written judgment survey (Part II). As mentioned earlier, each speaker selected either Eastern or Western Bengali pronunciation for the Bengali audio stimuli, depending on his or her familiarity with the two dialects. The words in the two Bengali lists are equivalent, aside from the regular dialectical variation in pronunciation. Speakers were assigned one of the two remaining lists (*i.e.* the English borrowings, and the cluster-initial words) according to their subject identification number (1-30). The English loanword list was presented to the 15 even-numbered subjects, while the cluster-initial list was presented to the remaining 15 odd-numbered subjects.

Both investigator and subject were seated near the computer, with the monitor only visible to the investigator, and the audio speakers directed towards the subject. The investigator played the audio file of each stimulus item individually, allowing the subject to listen to the file several times if requested, before giving a final answer. Subjects were asked to repeat the stimulus in full reduplicated form; this was recorded simultaneously using an Olympus DS-300 digital voice recorder. Shorthand transcriptions of each response were also inputted into the Microsoft Word file

containing the links to the audio files. Subjects were seated in a position from which the computer screen was not visible to them.

3.1.2.4.2 *Written Stimuli*

For the written judgment survey (Part II), subjects were presented with the same stimuli they encountered in Part I, although now all words were found written in test format. In the written tests, multiple choice responses were listed to the right of each written stimulus. These responses were divided into five columns:

- 1) reduplicants with consonantal fixed segment /**t**/,
- 2) reduplicants with consonantal fixed segment /**f**/,
- 3) reduplicants with consonantal fixed segment /**m**/,
- 4) reduplicants with consonantal fixed segment /**p**/, and
- 5) reduplicants with vocalic fixed segment /**u**/.

The five columns were also randomized for each speaker using Microsoft Excel's random number generator. A sixth column, simply labeled [onno kit^hu] 'something else', was provided in case subjects felt that the best response was not given in the five previously-mentioned columns. Two relatively non-controversial examples of fixed-segment reduplication were given at the top of each form; these included the unexciting reduplication of [baʃa] 'house' as [baʃa taʃa], and the well-known irregular reduplication of [kapoɹ] 'clothing' as [kapoɹ tʃopoɹ]. Used in the second example response given to all subjects, the reduplication of [kapoɹ tʃopoɹ] clarified the use of the

blank sixth column, which was largely reserved for irregular reduplicants that would not otherwise appear in the five aforementioned columns.

3.2 RESULTS

3.2.1 REDUPLICANTS OF NATIVE WORDS

When presented with audio recordings of native Bengali words, subjects employed /t/ as the fixed segment in reduplicants of bases starting with /t/ far less often than in reduplicants of bases starting with other consonants. Although no two subjects produced the same exact answers, a similar gradient pattern of fixed segment /t/-use was seen across virtually all speakers.³⁷ Bases starting with consonants considered similar to /t/ (e.g. /t^h/, /d/, /t̪/, etc.) allowed far fewer reduplicants with fixed segment /t/ than bases starting with consonants considered relatively dissimilar from /t/ (e.g. /m/, /p/, /f/, etc.). The percentage of fixed segment /t/-use in reduplicants of bases identified by initial consonant is provided in Figure 23 below.

³⁷ A handful of speakers appeared to use fixed segments other than /t/ as their default choice; the data from these speakers was not excluded from the analysis, although doing so may increase the statistical significance of some of the pairwise comparisons of consonants in Figure 14.

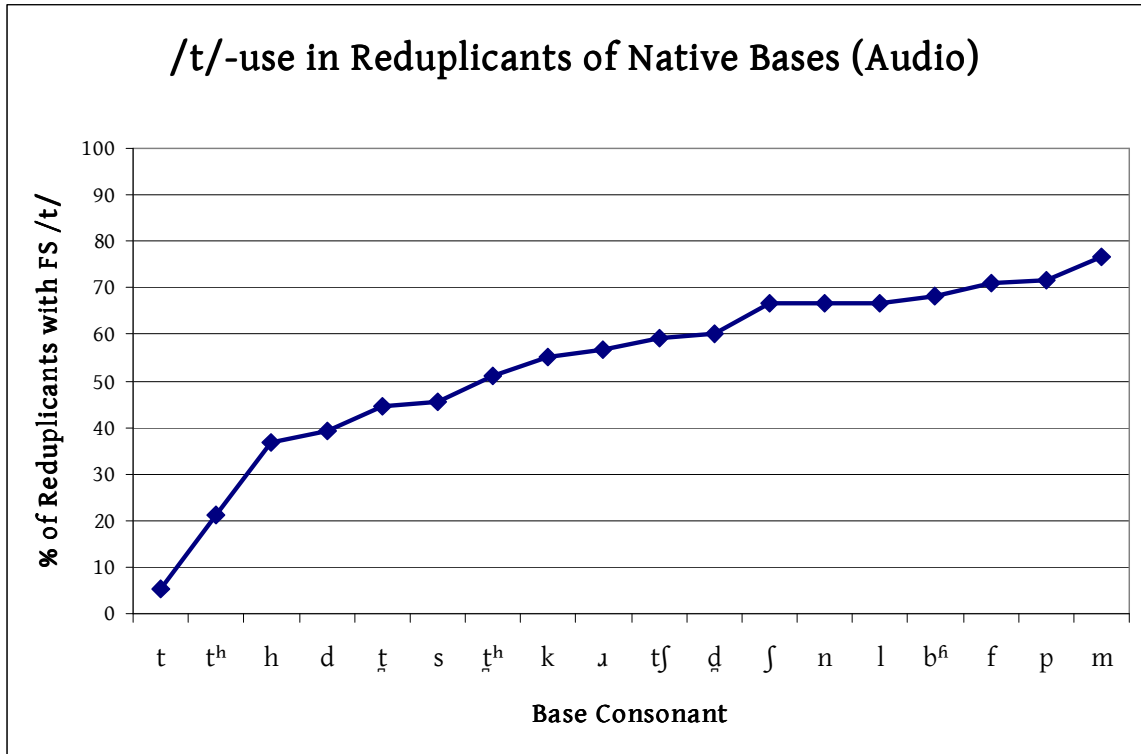


Figure 23. Percentage of fixed segment /t/-use in reduplicants of native words of varying initial consonants, when presented via audio recording format.

Use of fixed segment /t/ in reduplicants increases as the corresponding base-initial consonants become further dissimilar from /t/. Identity avoidance (avoidance of repeating the exact same string in succession) is largely obeyed, as the percentage of fixed segment /t/-use in /t/-initial bases is only just above 5.4%, while /b^h/-initial bases were followed by fixed segment /t/-initial reduplicants over 68.3% of the time. Similarity avoidance (avoidance of repeating a very similar string in succession) is also obeyed on a gradient scale, as six consonants (*i.e.* /t/, /t^h/, /h/, /d/, /t̥/, and /s/) appeared to trigger dissimilation from /t/ over 50% of the time.

The results of a repeated-measures Analysis of Variance (ANOVA) with one within-subjects factor (*i.e.* base-initial consonant) indicated that the initial consonant of a base had a significant effect on the acceptance of fixed segment /t/ in its reduplicant [$F(17,493) = 16.6$; $p < 0.01$]. However, pairwise *t*-tests³⁸ describing the behavior of bases with different initial consonants showed almost no significant distinctions between consonants that appear adjacent to one another in Figure 23. This lack of statistically significant distinctions is likely due to the highly gradient nature of similarity avoidance across the thirty speakers in this study.

The most important observations to note are that /t^h/-initial bases pattern more closely with /t/-initial bases with respect to fixed segment /t/-use than bases of any other initial consonant do. Bases starting with /d/ have lower fixed segment /t/-use than those starting with the dental consonants. Overall, it seems that the apical stops are the least likely to use fixed segment /t/, followed by laminal stops and affricates, with fricatives interspersed throughout. Looking at the graph in Figure 23, certain patterns seem to emerge: the voiceless unaspirated stop in each minor place of articulation (*i.e.* dental, alveolar, postalveolar, *etc*) uses fixed segment /t/ less often than the voiceless aspirated stop in that minor place of articulation (POA), which uses fixed segment /t/ less often than the voiced unaspirated stop in that minor POA, which

³⁸ The α value for the pairwise *t*-tests was set at 0.05 and adjusted for multiple comparisons using the Bonferroni method. See Appendix E for the results of these pairwise *t*-tests.

uses fixed segment /t/ less often than the fricatives of that minor POA. This is true for the apico-alveolars, the lamino-dentals, and the lamino-postalveolars.

It may be startling to notice in Figure 23 that /h/-initial words used fixed segment /t/ less often than /d/- or /t̥/-initial words; /h/ is the only non-coronal consonant to make use of fixed segment /t/ less than 50% of the time. The /h/-initial English borrowings also exhibited rather unexpected behavior with respect to the percentage of fixed segment /t/-use in their reduplicants. At this time, I can offer no explanation as to its peculiar behavior.

The fact that the bases whose reduplicants allowed the largest percentages of fixed segment /t/-use start with /m/, /p/, and /f/ should be of no surprise. As the backup fixed segments for default /t/ are most typically labial /m/, /p/, and /f/, it would be highly unlikely that a similarity-avoiding speaker would reduplicate an /m/-, /p/-, or /f/-initial base with any of these backup fixed segments.

The patterns of dissimilation suggest that certain features of initial consonants can trigger dissimilation from /t/. For example, the frequency of fixed segment /t/-use in reduplicants seems to be sensitive to the place ([+coronal] coronals vs. [-coronal] non-coronals), tongue distribution ([-distributed] apical vs. [+distributed] laminal consonants), sonorancy ([-sonorant] obstruents vs. [+sonorant] sonorants), and several other features and combinations of features of the base-initial consonant and the fixed

segment. When the base-initial consonant and fixed segment /t/ share that feature, dissimilation appears to be triggered. Base-initial consonants sharing the [+coronal] feature with fixed segment /t/ (i.e. coronals) are more likely to trigger dissimilation of the fixed segment than consonants that do not share this feature with fixed segment /t/ (i.e. dorsals and labials). Similarly, base-initial consonants that share the [-distributed], [-continuant], or [-sonorant] features of fixed segment /t/ are also more likely to trigger dissimilation of the fixed segment than those base-initial consonants that do not. As predicted by both the Frisch *et al.* and Coetzee & Pater models, shared features will trigger dissimilation.

When presented with the multiple-choice written survey in Part II of the study, speakers' acceptance of fixed segment /t/-use did not change drastically. The main difference seen is the overall greater acceptance of fixed segment /t/ in reduplicants of all sorts of bases, as speakers were given the option of choosing multiple reduplicants for the same base. Many speakers who produced reduplicants with backup fixed segments (e.g. /f/, /m/, etc.) in Part I of the study (audio elicitation) felt that both the dissimilated form (with the backup fixed segment) and the form with fixed segment /t/ were grammatical in Part II. Percentages of fixed segment /t/-use in Part II are graphed in Figure 24 below.

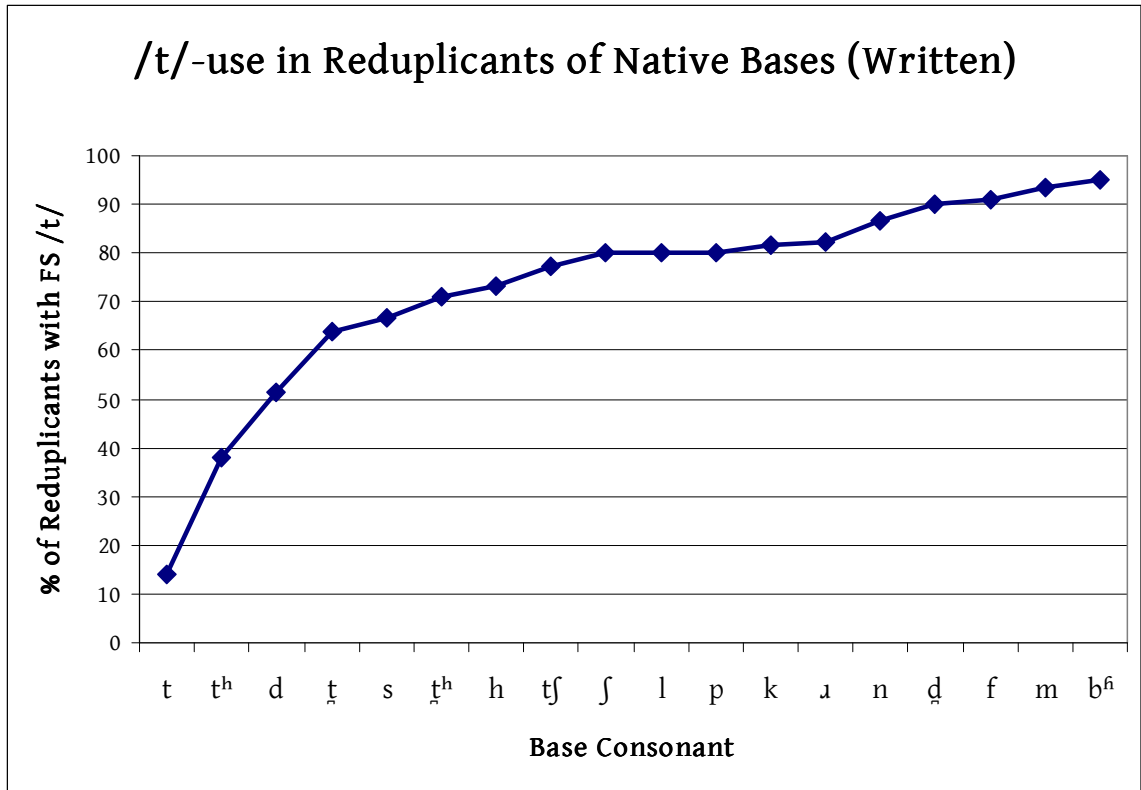


Figure 24. Percentage of fixed segment /t/-use in reduplicants of native words of varying initial consonants, when presented via written test format.

Once again, bases beginning with consonants similar to fixed segment /t/ (e.g. /t/, /t^h/, /d/, /ṭ/, etc.) allowed fewer reduplicants with fixed segment /t/ than bases beginning with consonants relatively dissimilar from fixed segment /t/ (e.g. /f/, /m/, /b^h/, etc.). Curiously, /h/ and /ɖ/ have moved further away from /t/ in Figure 24, while /p/ has moved closer. It is unclear as to exactly what factors are affecting the pattern, although one possibility is that having more options available has simply allowed certain bases to make use of more than one fixed segment, while the audio portion of the experiment only gave speakers the option of producing one reduplicant per base.

3.2.2 REDUPLICANTS OF ENGLISH BORROWINGS

Overall, reduplicants of English borrowings did not show as gradient a relationship between fixed segment /t/-use and the similarity of base-initial consonant to /t/ as the reduplicants of native Bengali words did. There does appear to be a trend in increased fixed segment /t/-use as the initial consonant of the base becomes increasingly dissimilar from /t/, as fixed segment /t/-use in reduplicants of /m/-, /p/-, and /f/-initial bases was around 90%, while fixed segment /t/-use in reduplicants of /t/- and /t^h/-initial bases remained below 50%, as shown in Figure 25 below. Still, the distinctions between the effects of most consonants are much more subtle.

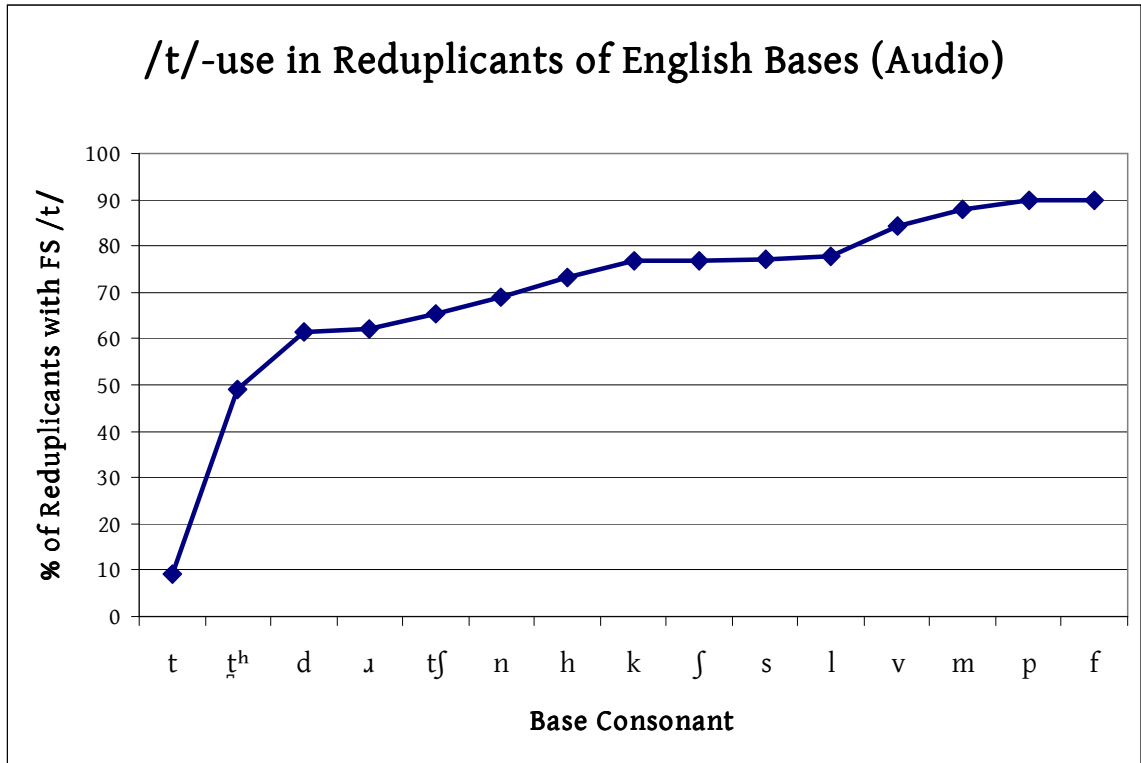


Figure 25. Percentage of fixed segment /t/-use in reduplicants of English borrowings of varying initial consonants, when presented via audio recording format.

The pattern of avoidance of fixed segment /t/ is almost binary in the results of the written task, as /t/-initial bases took fixed segment /t/ reduplicants 23.3% of the time, and all non-/t/-initial bases took fixed segment /t/ reduplicants over 70% of the time, as shown in Figure 26 below.

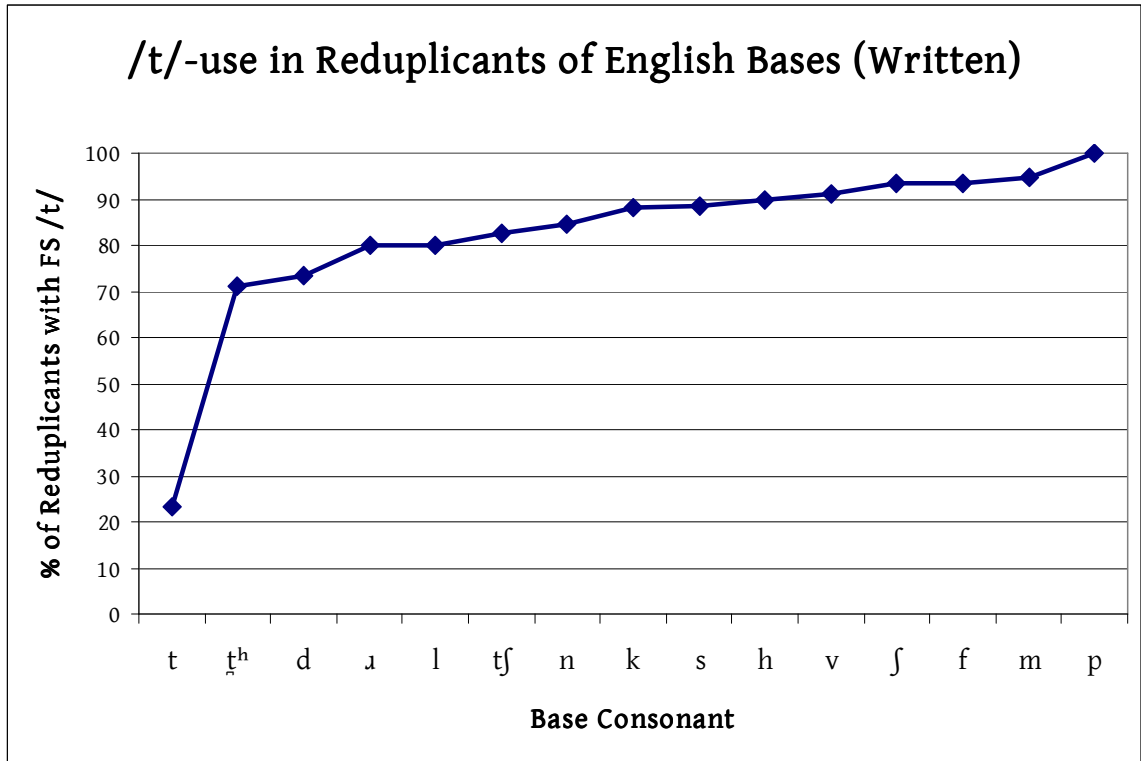


Figure 26. Percentage of fixed segment /t/-use in reduplicants of English borrowings of varying initial consonants, when presented via written test format.

Clearly, English bases accept far more reduplicants with fixed segment /t/ than Bengali bases; possibly as a consequence, the gradient nature of similarity avoidance in this reduplicative construction is exceedingly weak in the English loans. Considering the English bases are relatively common loanwords, and considering their phonological structure is presumably no more marked than the Bengali bases used in the study, it is unclear what aspect or aspects of the English loanwords is affecting the corresponding fixed segment /t/-usage patterns.

3.2.3 REDUPLICANTS OF SANSKRITIC AND ENGLISH BORROWINGS WITH INITIAL CLUSTERS

While indeed fascinating, the cluster-initial base data could not be included in this analysis due to time and space constraints.

3.3 ANALYSIS

Using the reduplicant data from the audio-elicited native Bengali stimuli, a direct comparison can be made between the predictions of each metric mentioned earlier and the observed Bengali pattern of similarity avoidance. The four metrics (the Shared Natural Classes metric, the Relativized OCP metric, the Consonant Cooccurrence metric, and the Feature Weight metric) are described below with a discussion of their effectiveness in modeling the Bengali data.

3.3.1 SHARED NATURAL CLASSES (SHARED NATURAL CLASSES) METRIC

The similarity metric based on shared natural classes, as described in detail in Frisch *et al.* (2004), predicts that if fixed segment /t/-use in Bengali echo reduplication is based on similarity of the fixed segment /t/ and the base-initial consonant, consonants that share more natural classes with /t/ will accept far fewer fixed segment /t/-reduplicants than those that share fewer natural classes with /t/.

Figure 27 graphs the predicted dissimilarity of selected Bengali consonants from /t/ (as per the Frisch *et al.* model of Shared Natural Classes similarity) versus the observed fixed segment /t/-use in reduplicants of native Bengali bases elicited via audio recording. Figure 27 presents Shared Natural Classes similarity as measured across the entire phoneme inventory of Common East Bengali.

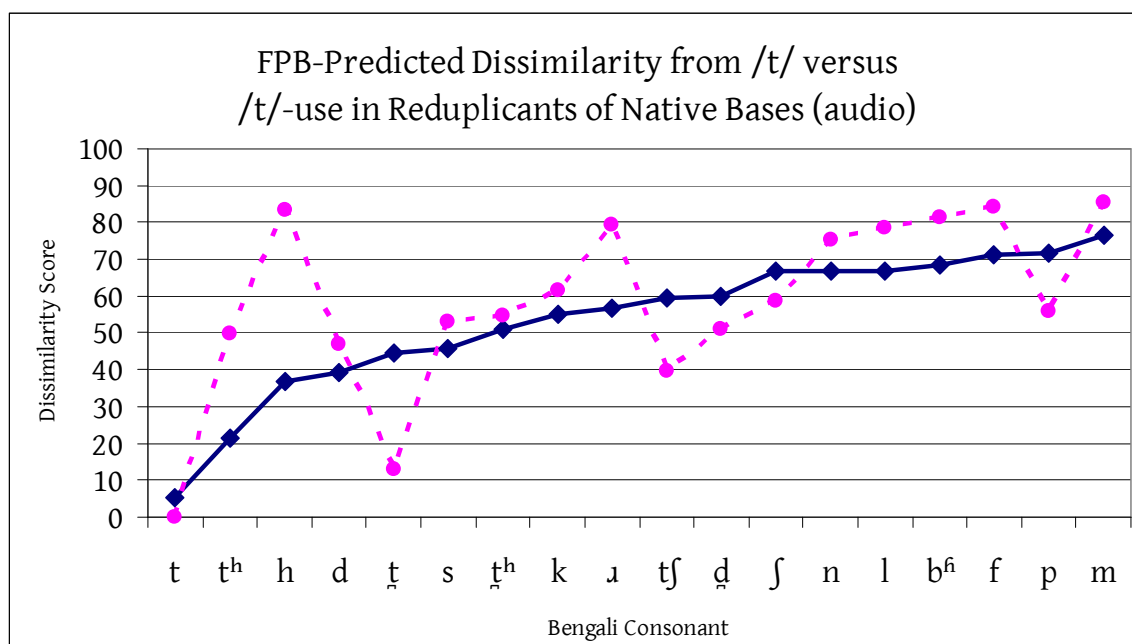


Figure 27. Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format (solid line), versus the predicted fixed segment /t/-use as per the Frisch *et al.* model of similarity based on Shared Natural Classes, measured across all Bengali phonemes (dotted line).

The overall trend of fixed segment /t/-use does loosely follow the Frisch *et al.* predictions; sonorant- and labial-initial bases accepted a high number of fixed segment /t/-reduplicants, while /t/-initial bases accepted the least. However, the predictions illustrated in Figure 27 are furthest from the observed fixed segment /t/-use when

dealing with the low fixed segment /t/-use with /h/-initial bases, the relative order of fixed segment /t/-use in /t^h/-, /d/-, /t̪/-, and /tʃ/-initial bases, the high fixed segment /t/-use with /p/-initial bases, and at many other points. A bivariate scattergram showing the correlation between the predictions of the Frisch *et al.* Shared Natural Classes metric and the observed data is given in Figure 28 below. Although the correlation shown in Figure 28 is statistically significant [$p < 0.01$], it is rather weak [$r^2 = .437$].

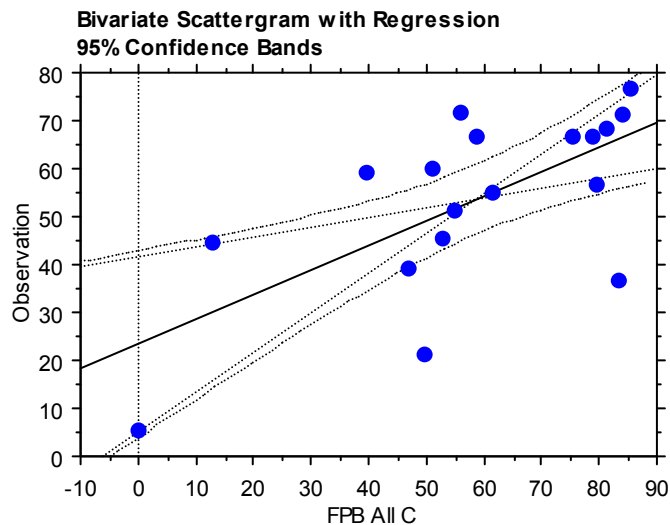


Figure 28. Fixed segment /t/-use predicted by Frisch *et al.* shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory. [$p < 0.01$, $r^2 = .437$]

Assuming that /h/-initial bases are simply inexplicably intolerant of fixed segment /t/-reduplicants for mysterious reasons, we can get a slightly more acceptable match between the predicted and observed patterns in Figure 29 below. A bivariate

scattergram depicting the correlation of the observed data with the Frisch *et al.* predictions minus /h/ is given in Figure 30.

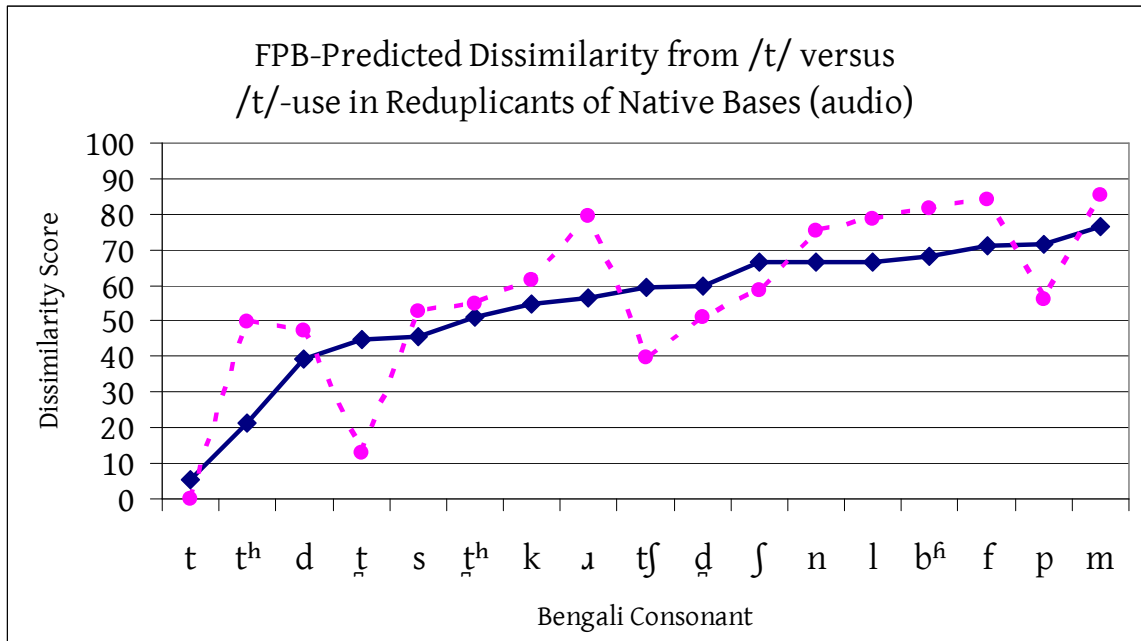


Figure 29. Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format (solid line), versus the predicted fixed segment /t/-use as per the Frisch *et al.* model of similarity based on Shared Natural Classes, measured across all Bengali phonemes (dotted line), excluding the phoneme /h/.

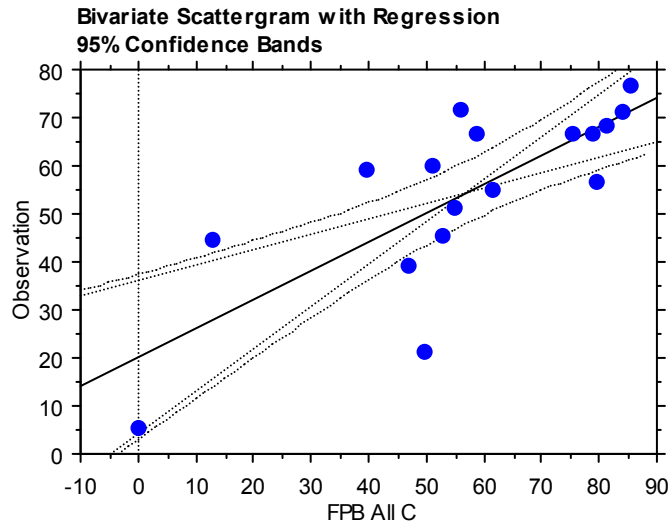


Figure 30. Fixed segment /t/-use predicted by Frisch *et al.* shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed. [$p < 0.01$, $r^2 = .584$]

This correlation is only marginally stronger [$p < 0.01$, $r^2 = .584$]; removing /h/ of course does not remedy the poor predictions with respect to the coronal-initial bases. Considering Frisch *et al.* only focused their Shared Natural Classes calculations within each major place, narrowing down the similarity measurements to the coronal subset of Bengali phonemes may improve the match between the Frisch *et al.* prediction and the observed patterns. Figure 31 illustrates how the predictions better approximate the data in this restricted inventory.

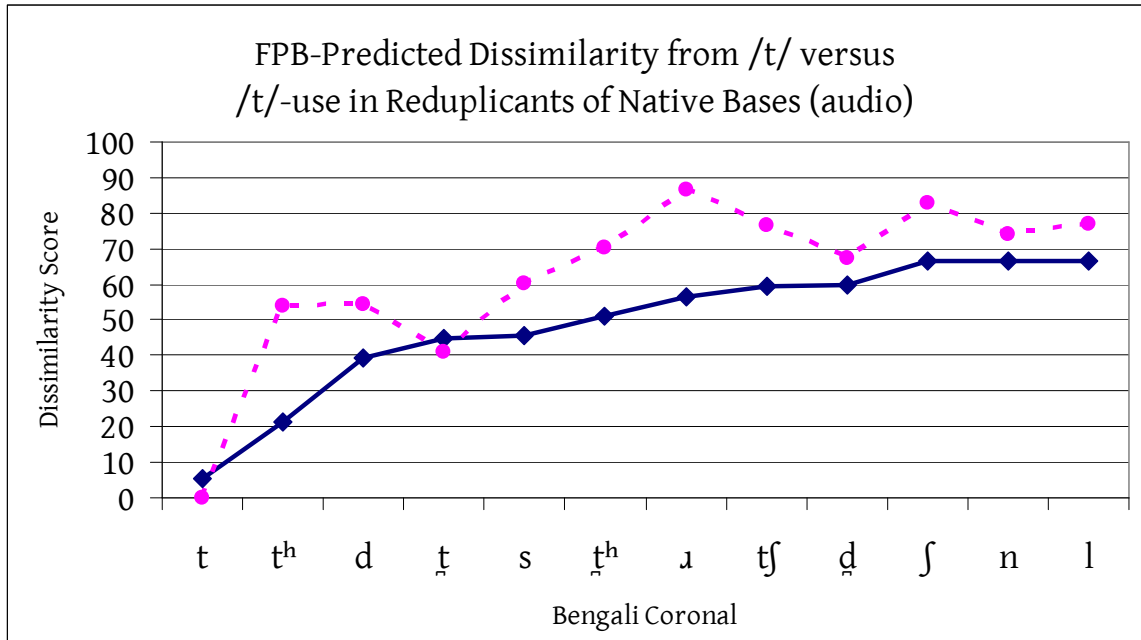


Figure 31. Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format (solid line), versus the predicted fixed segment /t/-use as per the Frisch et al. model of similarity based on Shared Natural Classes, measured across Bengali coronals only (dotted line).

Unfortunately, the relative order of /t^h/-, /d/-, and /ṭ/-initial bases is not resolved; still, the correlation is now stronger [$p < 0.01$, $r^2 = .775$]. The bivariate scattergram for the correlation is shown in Figure 32. Non-coronal consonants are not calculated for similarity with /t/, and are thus predicted to freely cooccur with fixed segment /t/.

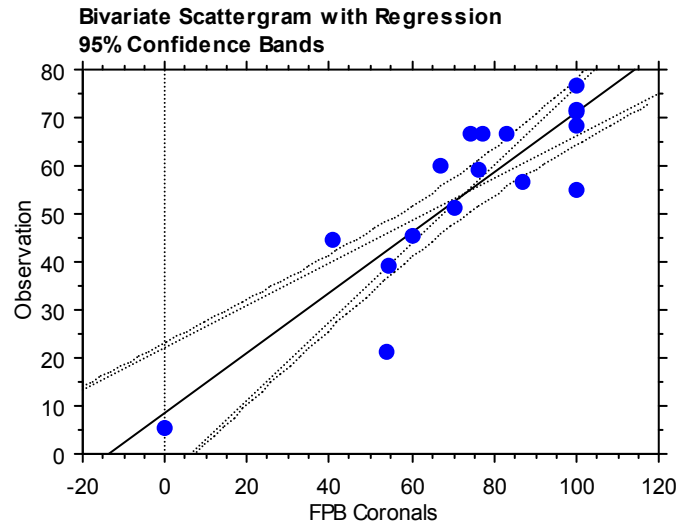


Figure 32. Fixed segment /t/-use predicted by Frisch *et al.* shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the coronal consonants only. [$p < 0.01$, $r^2 = .775$]

The strictest interpretation of Frisch *et al.*'s model further divides up coronal obstruents from sonorants. Given that the coronal obstruents have been the most problematic in the predictions portrayed in Figure 27, Figure 29, and Figure 31, constraining the inventory to only coronal obstruents as shown in Figure 33 may produce better results. The bivariate scattergram illustrating the correlation is given in Figure 34.

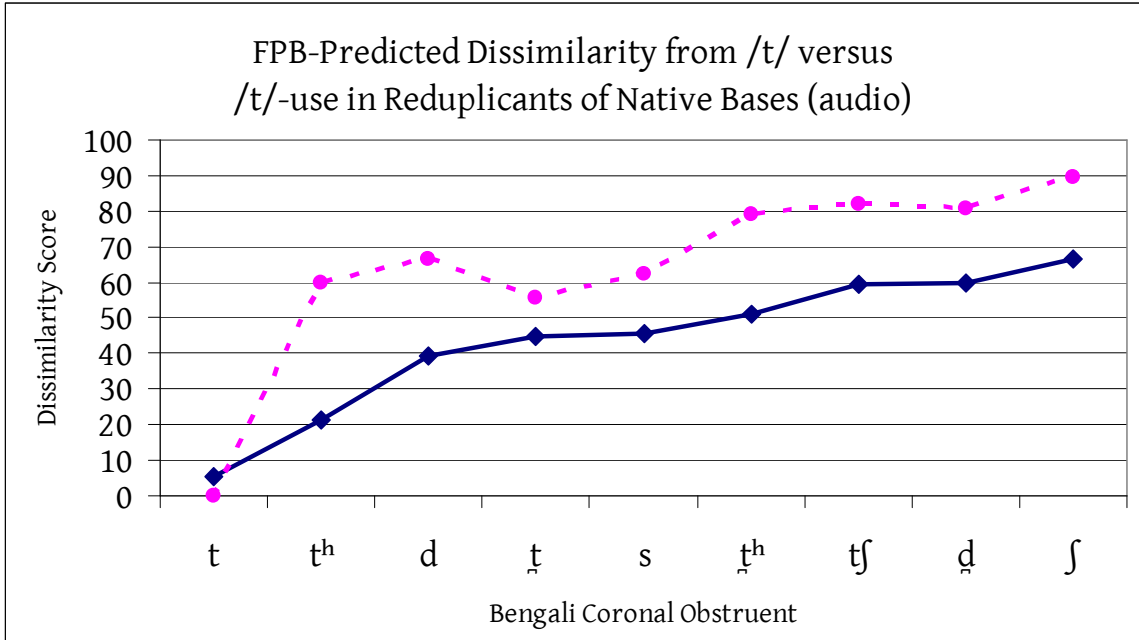


Figure 33. Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format (solid line), versus the predicted fixed segment /t/-use as per the Frisch et al. model of similarity based on Shared Natural Classes, measured across Bengali coronal obstruents only (dotted line).

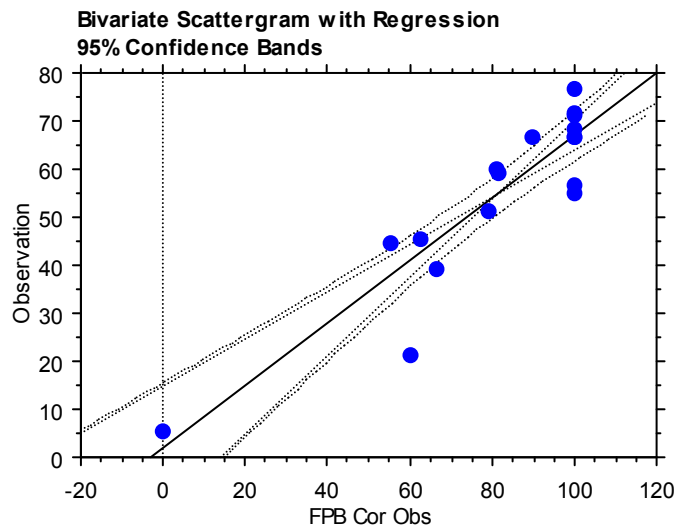


Figure 34. Fixed segment /t/-use predicted by Frisch et al. shared natural classes metric graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the coronal obstruents only. [$p < 0.01$, $r^2 = .831$]

Although Figure 33 is the closest match thus far using the Frisch *et al.* Shared Natural Classes metric of similarity to predict fixed segment /t/-use [$p < 0.01$, $r^2 = .831$], most of the apico-alveolar distinctions cannot be accounted for using this metric. For example, while fixed segment /t/-use is greater for /d/-initial bases than for /t^h/-initial bases, and greater for /t̪/-initial bases than for /t^h/-initial bases, the Frisch *et al.* metric predicts that /t^h/-, /d/, and /t̪/-initial bases should accept about the same number of fixed segment /t/-reduplicants, with /t̪/-initial bases accepting somewhat fewer than the other two, in contrast to the observed pattern. Considering these phonemes showed some of the most important distinctions in the dissimilation pattern (see Figure 23), the inability of the Frisch *et al.* metric (restricted to coronal obstruents) to model these phenomena detracts from its otherwise close prediction for /t̪^h/-, /tʃ/-, /d̪/-, and /ʃ/-initial bases.

Measuring similarity of consonants using shared natural classes is an attractive theory in its invariable application across languages, allowing for a universal metric of consonant similarity. However, this metric fails in languages where certain places of articulation have disproportionately large inventories of consonants, with subcategories of minor place, as with the case of lamino-dentals, lamino-alveolars, and apico-alveolars in Bengali. While Bengali speakers seem to be extremely attentive to these minor place distinctions, the shared natural classes metric overemphasizes the role of voicing and aspiration in measuring similarity.

3.3.2 RELATIVIZED OCP CONSTRAINTS

Applying the Coetzee & Pater model of relativized OCP constraints to determine consonant similarity to Bengali fixed segment reduplication, a hierarchy of more specific constraints against the cooccurrence of /t/-like segments ranked over a more general constraint against the cooccurrence of all coronals can be proposed. One possible hierarchy, first shown in Table 1, is repeated in

Figure 35 below, with nonsense word candidates rated for relative acceptance predictions. It is important to keep in mind that these constraints are not strictly-ranked in the traditional manner, but are arranged in the order of how many exceptions are allowed; more effective constraints (*i.e.* those that allow the fewest exceptions to the generalization) are placed above less effective ones (*i.e.* those that allow many exceptions).

	OCP-COR (αs.g., αvoi, αdist, αd.l., αant, αson, αnas, αlat)	OCP-COR (αvoi, αdist, αd.l., αant, αson, αnas, αlat)	OCP-COR (αdist, αd.l., αant, αson, αnas, αlat)	OCP-COR (αd.l., αant, αson, αnas, αlat)	OCP-COR (αant, αson, αnas, αlat)	OCP-COR (αson, αnas, αlat)	OCP-COR (αnas, αlat)	OCP-COR (αlat)	OCP-COR
tana-tana	*!	*	*	*	*	*	*	*	*
t ^h ana-tana		*!	*	*	*	*	*	*	*
dana-tana			*!	*	*	*	*	*	*
ṭana-tana				*!	*	*	*	*	*
sana-tana					*!	*	*	*	*
t ^h ana-tana				*!	*	*	*	*	*
ɰana-tana						*!	*	*	*
tʃana-tana						*!	*	*	*
ɖana-tana				*!	*	*	*	*	*
ʃana-tana						*!	*	*	*
nana-tana								*!	*
lana-tana									*!
hana-tana									
kana-tana									
pana-tana									
fana-tana									
b ^h ana-tana									
mana-tana									

Figure 35. Nonsense candidates subjected to violations of relativized OCP constraints arranged in order of allowing fewest to most exceptions.

Calculating the strength of the highest-ranked constraint each consonant violates when cooccurring with fixed segment /t/ as discussed in Section 2.3 can produce a rough similarity score for the pair (e.g. {t,t} = 100, {p,t} = 0); subtracting this score from 100 can give the dissimilarity score (e.g. {t,t} = 0, {p,t} = 100) and thus the predicted pattern of fixed segment /t/-use. This prediction is graphed against the observed pattern in Figure 36 below.

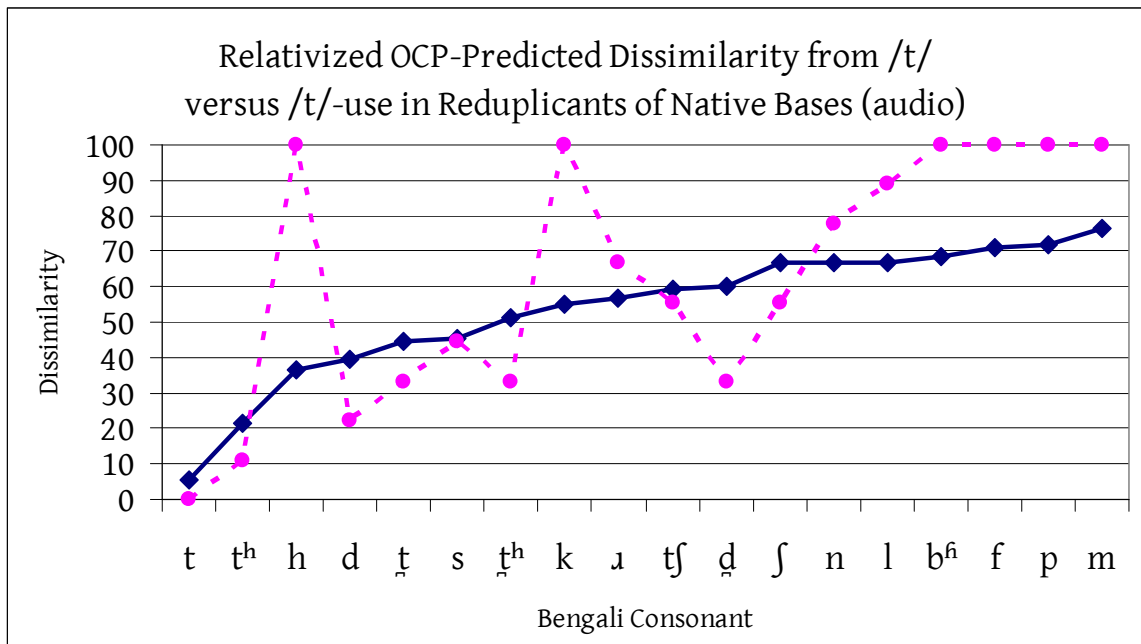


Figure 36. Relativized OCP-predicted dissimilation pattern (dotted line) graphed against the observed pattern (solid line). Predicted dissimilarity is calculated as the percentage of constraints **not** violated by the cooccurrence of the consonant with fixed segment /t/.

Similar to the predictions of the Frisch *et al.* Shared Natural Classes metric, the relativized OCP cannot account for /h/- and /k/-initial bases. However, the relativized OCP predictions significantly match the overall pattern observed in dissimilation [$p < 0.01$, $r^2 = .717$]. The correlation of the relativized OCP predictions and the observed pattern is found in the bivariate scattergram given in Figure 37.

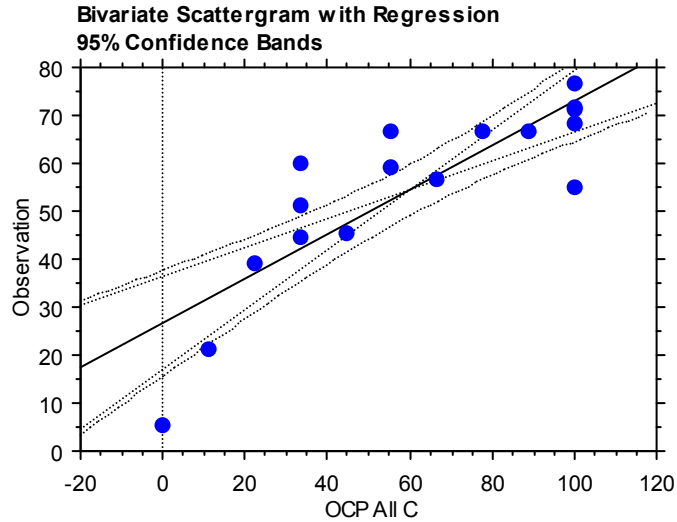


Figure 37. Fixed segment /t/-use predicted by a relativized OCP constraint hierarchy graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed. [$p < 0.01$, $r^2 = .717$]

Because the relativized OCP constraint metric, like the Frisch *et al.* Shared Natural Classes metric, can only measure similarity within each major place of articulation, Figure 38 measures similarity only among the coronal consonants; the scattergram for this correlation is the same as that shown in Figure 37, as the predictions for non-coronal-initial bases do not change when they are removed from the calculations (they would still be considered totally dissimilar to /t/ and thus would be assumed to freely cooccur with fixed segment /t/) [$p < 0.01$, $r^2 = .717$].

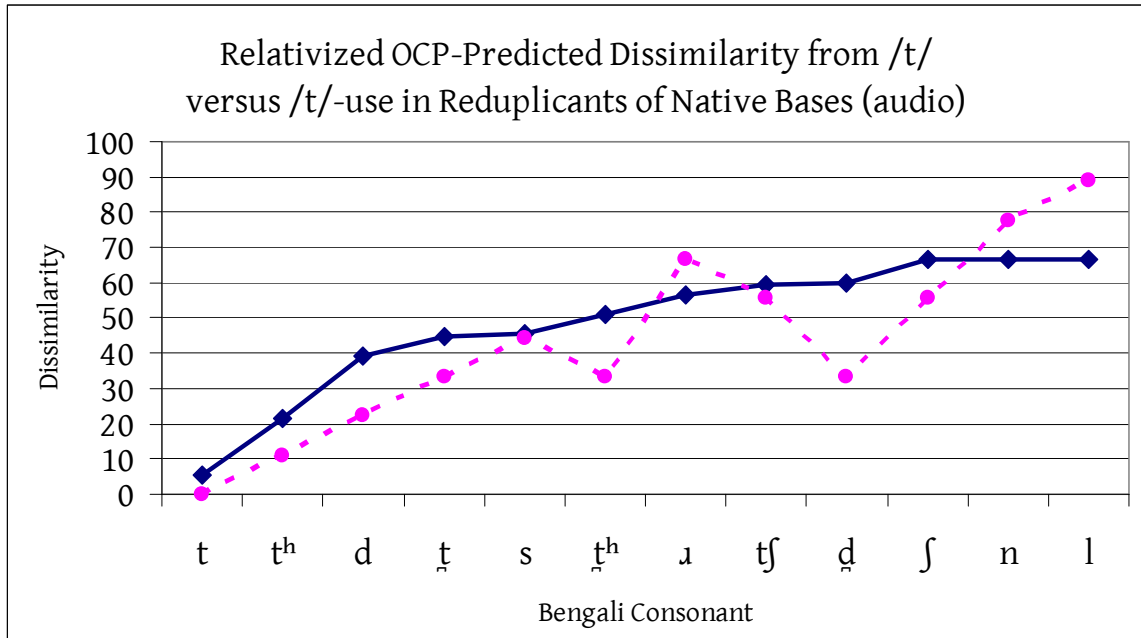


Figure 38. Relativized OCP-predicted dissimilation pattern (dotted line) graphed against the observed pattern (solid line), measuring coronals only. Predicted dissimilarity is calculated as the percentage of constraints **not** violated by the cooccurrence of the consonant with fixed segment /t/.

The match between the predictions of a similarity metric based on number of relativized OCP constraints violated and the observed dissimilation data in Common East Bengali fixed segment reduplication is actually very close when it comes to the cooccurrence of apical consonants (i.e. /t/, /t^h/, /d/, /s/, /n/, and /l/) with fixed segment /t/; voicing, aspiration, continuancy, and sonorancy distinctions are beautifully modeled thanks to the feature configurations of the more specific OCP constraints. The overall correlation of the prediction with the data is close [$p < 0.01$, $r^2 = .717$], although the cooccurrence predictions for the laminal consonants (i.e. /ṭ/, /ṭ^h/, /ḍ/, /ɹ/, /tʃ/, and /ʃ/) with fixed segment /t/ are unfortunately all problematic; all lamino-dental consonants are judged equally similar by the relativized OCP constraint

hierarchy, although their observed behavior shows that they are distinct – voiceless aspirated /tʰ/-initial bases, for example, allow fewer fixed segment /t/-reduplicants than do /d/-initial bases.

The ability to combine features into each OCP constraint with more features specified in higher-ranked constraints is both a benefit and a shortcoming of the relativized OCP metric of similarity. Although this allows for language-specific combinations of features, it does not break down the idea of major place of articulation into smaller places, which appears to be a salient aspect for measuring similarity in Bengali, with apico-alveolars, lamino-alveolars, and lamino-dentals all contrasting within the coronal major place of articulation.

3.3.3 CORRELATION WITH CONSONANT COOCCURRENCE

Although Frisch *et al.* (2004) and Coetzee & Pater (2005) found strong correlations between consonant similarity and consonant cooccurrence within roots in different languages, this correlation does not seem to hold true for Bengali. While Bengali roots (which include native, borrowed, and re-borrowed Sanskrit vocabulary) seem to illustrate a restriction against the cooccurrence of the identical pair {t, t} within a CVCV root and a preference for the cooccurrence of rather dissimilar {f, t}, the pattern between these extremes does not accurately model the data seen in Bengali fixed segment echo reduplication. For example, while the cooccurrence of {tʰ, t} is

overrepresented in Bengali roots of the shape CVCV, the reduplication of /t^h/-initial bases with fixed segment /t/ is extremely low. The predictions of a metric based on consonant cooccurrence (the O/E value of the pair, multiplied by 30 to scale predictions up) are shown with a dotted line below in Figure 29, with the observed fixed segment /t/-usage graphed in a solid line.

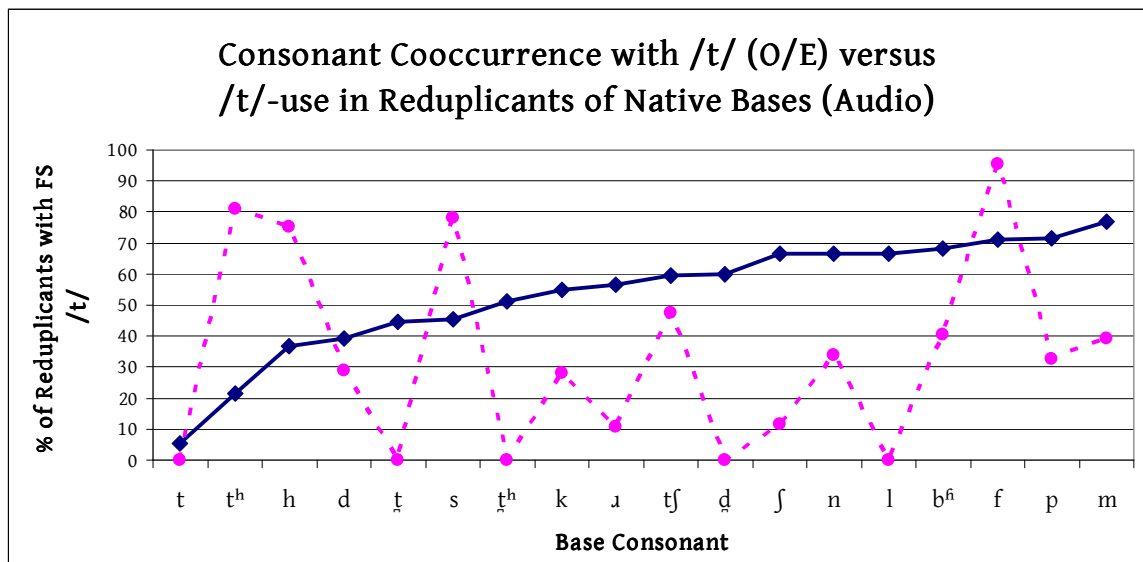


Figure 39. Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format (solid line), versus the predicted fixed segment /t/-use as determined by consonant cooccurrence within CVCV roots (dotted line).

As given in the bivariate scattergram in Figure 40, the correlation of root-internal consonant cooccurrence and echo reduplication cooccurrence with fixed segment /t/ is atrocious [$p > 0.81$, $r^2 = .004$]. There appears to be absolutely no relationship between cooccurrence restrictions within CVCV roots and similarity avoidance in Common East Bengali fixed segment reduplication.

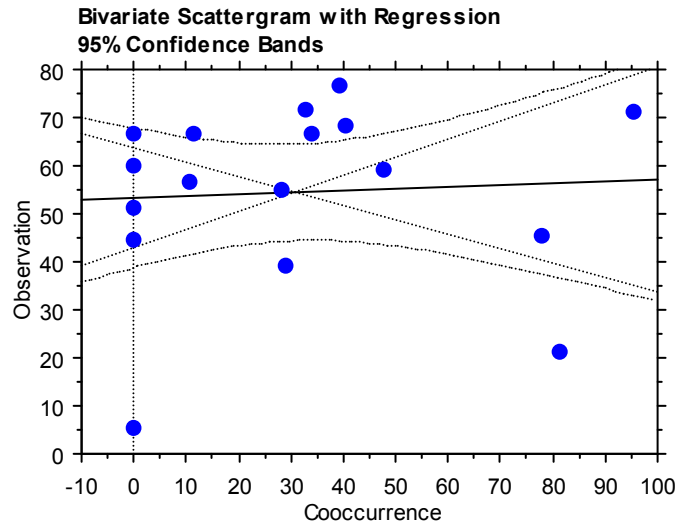


Figure 40. Fixed segment /t/-use predicted by root-internal consonant cooccurrence graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed. [$p > 0.81$, $r^2 = .004$]

Clearly, the hypothesis that the same metric of similarity can determine both consonant cooccurrence within roots and fixed segment /t/-use in echo reduplication is not supported. What is of interest here is the fact that consonant cooccurrence within roots is definitely subject to some sort of OCP restrictions as seen in Figure 22, with dental consonants and /l/ found to never cooccur with /t/, and aspirated /p^h/, /t^h/, and /tʰ/ found to cooccur with /t/ far more often than would be predicted by chance. Given the cooccurrence data, there seems to be an OCP restriction against the cooccurrence of dental consonants and /l/ with alveolar /t/, although the effects of that restriction are apparently not seen in fixed segment /t/-use in reduplicants of /l/- or dental-initial bases. Assuming that the OCP is responsible both for cooccurrence

restrictions within roots (in the lexicon) and for dissimilation patterns with fixed segment /t/ (in the grammar), it is astonishing to see that the particular OCP restrictions in the lexicon have little or nothing to do with the particular OCP restrictions at work in the productive grammar.

3.3.4 FEATURE WEIGHTING

If they are not using the Shared Natural Classes metric, a relativized OCP constraint hierarchy, or OCP constraints from the lexicon, it may be that Central East Bengali speakers are measuring the similarity of two consonants based on the number of shared features, with certain features weighted more heavily than others. While features such as [voice] and [spread glottis] might be weighted more heavily due to their important role in distinguishing the most similar consonants to /t/ (as seen in Figure 23), features such as [labial] and [lateral] might be weighted less heavily. Inputting the observed patterns of dissimilation in the reduplicants of native Bengali words presented in audio format, along with the feature specifications³⁹ of the Bengali inventory into the software program R (R Development Core Team 2005), the program attempted to model the data by assigning different constraints different weights, using the equations of similarity and fixed segment /t/-use probability given in Equation 1 and Equation 2, respectively. After the program had completed all trials in an attempt to calculate the similarity scores and simultaneously adjust the feature weights

³⁹ Feature specifications were converted to the values 0 and 1, with 0 signifying that the feature specification was the same as that of /t/, and 1 signifying that it was opposite that of /t/.

accordingly, it was found that only four of the seventeen relevant features⁴⁰ were weighted more heavily than others (which maintained the default weight of 0.100). While features such as [lateral] and [labial] were found to carry a weight of only 0.100, the feature [distributed] was found to carry four times as much weight, with a w-value of 0.400.

Table 3 below includes the four features found to carry more weight than the default 0.100 assigned to all other features. Note that these four features are the most important in distinguishing the consonants judged most similar to fixed segment /t/ (e.g. /t/, /t^h/, /d/, /t̚/, /s/, etc.) in Figure 23.

Heavily-Weighted Feature	Weight
[voice]	0.554
[distributed]	0.400
[strident]	0.249
[spread glottis]	0.198

Table 3. Feature weights predicted by the feature-weighting metric to be greater than default 0.1.

Because they often are the only feature distinguishing certain consonants from /t/, it is understandable that these features would be weighted more heavily than other features. Converting these features weights into consonant similarity scores (using Equation 2), where larger sums of shared feature weights (remember all features other than those mentioned in

⁴⁰ All features are listed in Appendix C.

Table 3 are still counted, but just at the default value of 0.1) correspond directly to low fixed segment /t/-use, we find the pattern shown in Figure 41, where predicted fixed segment /t/-use given the feature weights mentioned in Table 3 is graphed against the observed fixed segment /t/-use in reduplicants of native Bengali bases presented in audio format.

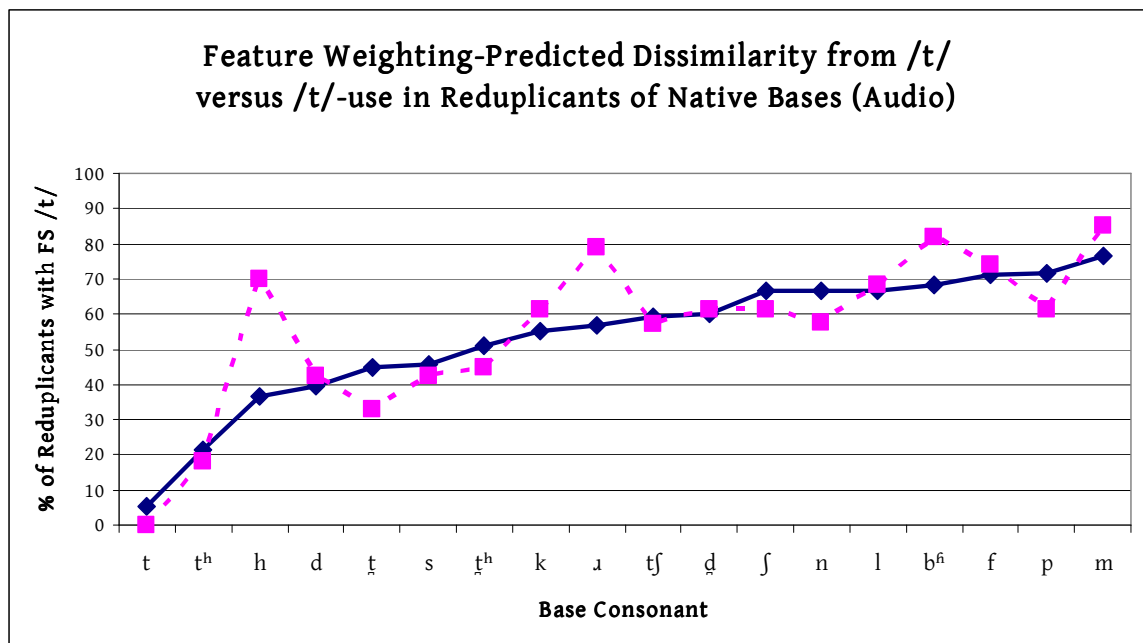


Figure 41. Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format (solid line), versus the predicted fixed segment /t/-use as determined by feature weighting as described in Table 3 (dotted line). Similarity measured across the entire phoneme inventory.

A bivariate scattergram illustrating the correlation between the observed data and the predictions of measuring similarity across the entire phoneme inventory based on the feature weights listed in

Table 3 is provided in Figure 42 below. The correlation is not only highly significant but also relatively strong [$p < 0.01$, $r^2 = .729$].

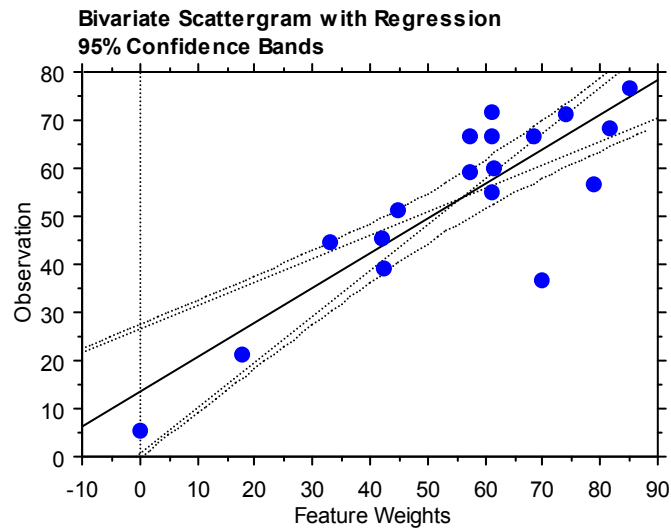


Figure 42. *Fixed segment /t/-use predicted by feature weighting graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format. Similarity measured across the entire phoneme inventory. [$p < 0.01$, $r^2 = .729$].*

As with the other metrics investigated in this study, removing the problematic /h/-initial bases from the data for exploratory purposes helps tighten the correlation between the predictions of the metric and the observed data, as shown in Figure 43. As shown in the bivariate scattergram in Figure 44, the correlation of the feature weighting metric with the observed data is extremely close [$p < 0.01$, $r^2 = .855$].

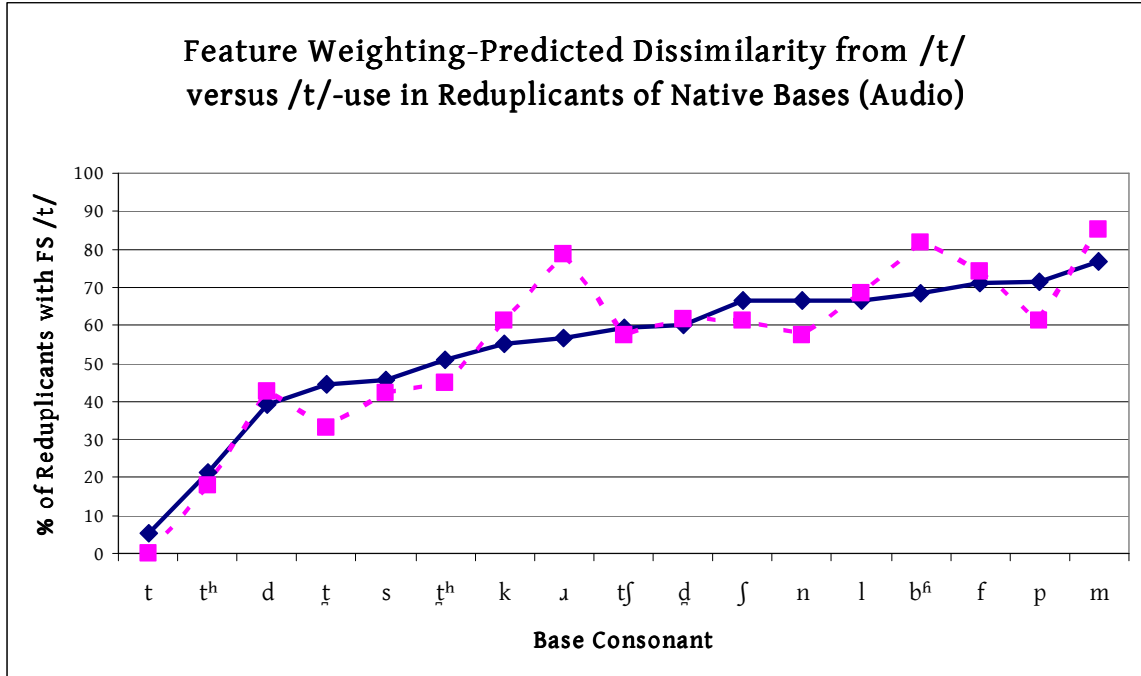


Figure 43. Observed fixed segment /t/-use in reduplicants of native words presented in audio recording format (solid line), versus the predicted fixed segment /t/-use as determined by feature weighting as described in Table 3 (dotted line). The data point for /h/-initial bases is removed.

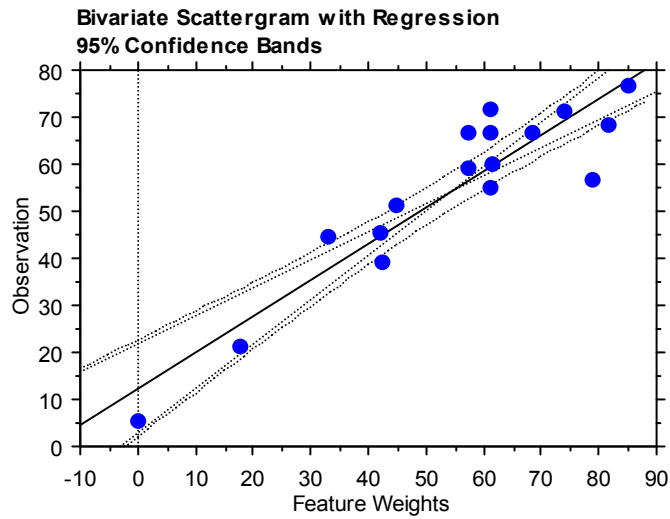


Figure 44. Fixed segment /t/-use predicted by feature weighting graphed against the observed fixed segment /t/-use in reduplicants of native Bengali words presented in audio format.

Similarity measured across the entire phoneme inventory with the data point for /h/-initial bases removed. [$p < 0.01$, $r^2 = .855$].

The correlation between the feature weighting metric and the observed data is not only exceptionally tight, but also attentive to the distinctions between most of the coronal consonants. Other metrics, such as the Frisch *et al.* shared natural classes metric, could not accurately predict the behavior of dentals and alveolars unless the domain of similarity calculations was restricted to coronal obstruents only. Unfortunately, restricting the domain to just the coronal obstruents may keep the Shared Natural Classes correlation strong within coronal obstruents at the expense of the close correlation with the noncoronals, which are all considered to be equally dissimilar from /t/ in the strictest interpretation of the Shared Natural Classes metric. Also, unlike the relativized OCP hierarchy, the weighted features model does not treat dentals all equally; most important distinctions are carefully reflected in this metric.

Whether describing similarity across the entire phoneme inventory or just the coronal obstruent subset, the Frisch *et al.* Shared Natural Classes metric could not predict that Bengali speakers would consider /t^h/ to be the most similar non-identical segment to /t/. The weighted features metric, however, can beautifully replicate this observation. Unfortunately, with the limited number of repetitions run in R to determine the appropriate feature weights, the weighted feature model could not predict that /d/ would be the next most similar consonant to /t/ after /t^h/. Nonetheless, the high value for the coefficient of correlation between the weighted

feature predictions and the observed data along with the ability to predict the close similarity between /t/ and /t^h/, and the ability to distinguish the dental consonants from one another all suggest that the feature weighting model is a good basis for describing the patterns in Common East Bengali echo fixed segment reduplication.

4 CONCLUSIONS AND FURTHER INVESTIGATION

Speakers of Common East Bengali apply both identity avoidance and similarity avoidance in fixed-segment echo reduplication. To avoid total homophony between a /t/-initial base and its fixed-segment-initial reduplicant (identity avoidance), speakers will use backup fixed segments such as /f/ or /m/ instead of the default fixed segment /t/. To avoid near-homophony between a base and its fixed-segment-initial reduplicant (similarity avoidance), speakers will use backup fixed segments even when the base begins with consonant not identical to but judged relatively similar to /t/. The more dissimilar a base-initial consonant is judged to be from /t/, the more likely it is to allow a fixed segment /t/-initial reduplicant. Although the gradient nature of similarity avoidance in Bengali fixed-segment echo reduplication is clear, the metric used to actually calculate how similar a consonant is to the fixed segment /t/ is less so.

The Frisch *et al.* metric of similarity based on shared natural classes maintains that homorganic segments are considered more similar if they both belong to more of the same natural classes than other segments do. This predicts that bases starting with

coronal consonants that share more natural classes with /t/ will use fixed segment /t/ less often than bases starting with coronal consonants that share fewer natural classes with /t/, which will use fixed segment /t/ less often than bases starting with non-coronal consonants. Although the overall observed pattern of dissimilation can be predicted by the shared natural classes metric, it seems to be too rigid to deal with the language-specific features of Bengali when it comes to homorganic consonants. The Shared Natural Classes metric accurately predicts homorganic consonant cooccurrence phenomena within Arabic roots, accurately describing how coronal obstruents rarely cooccur with other coronal obstruents while they can far more freely cooccur with coronal sonorants, how the occurrence of oral gutturals (*i.e.* uvular fricatives) is underrepresented in roots that contain either dorsal consonants (velar and uvular stops) or gutturals of any kind (glottal, pharyngeal, or oral), and many other cooccurrence phenomena. However, the intricate distinctions of voicing, aspiration, and laminal versus apical tongue orientation in Bengali coronals are not predicted by this theory. It appears that the Shared Natural Classes metric of similarity as presented in Frisch *et al.* (2004) is better suited to the natural classes that best describe Arabic phonemic distinctions than to those that describe the phonemic distinctions of Bengali.

The Coetzee & Pater metric of similarity based on relativized OCP constraints allows for greater freedom in arranging features shared by a pair of consonants, and might thus be more easily tailored to describe language-specific phonemic distinctions. Positing relativized OCP constraints (with the more specific constraints ranked above

the more general constraints) with lexically-specific faithfulness constraints protecting certain forms accurately describes the consonant cooccurrence patterns seen in Muna roots. The particular voicing and prenasalization specifications of Muna consonants can be better encoded in these constraints, allowing the OCP to be sensitive to those distinctions that carry the greatest weight in phonemic contrasts in Muna. Applying this theory to the Bengali inventory could encode distinctions in tongue orientation, voicing, and aspiration in the relativized OCP, but requiring that each successive OCP constraint be more general than the one above it forces certain feature combinations to be overlooked, ignoring distinctions in aspiration and voicing within a particular tongue orientation, for example. Once again, the particular combinations of features interacting with both major (*e.g.* coronal) and minor (*e.g.* alveolar) place of articulation lessen the applicability of the relativized OCP metric to Bengali echo fixed segment reduplication.

Comparing cooccurrence of a base-initial consonant with reduplicant-initial fixed segment /t/ to the cooccurrence of that consonant with /t/ within roots in the Bengali lexicon is equally unfruitful, as the OCP effects seen in the lexicon do not line up with fixed segment /t/-use in echo reduplication. While there are OCP effects in the lexicon, with certain consonants occurring very rarely, if at all, with /t/ in roots, these effects do not seem to be playing a significant role in the productive grammar, insofar as echo reduplication is concerned. This observation has interesting implications for both the Coetzee & Pater and Frisch *et al.* studies. Both of these studies formalize consonant

cooccurrence in the lexicon as an aspect of the synchronic grammar, suggesting that speakers have an implicit awareness of consonant similarity as based on the patterns in their vocabulary. Frisch *et al.* hypothesize that this similarity can be measured by calculating the number of natural classes shared between consonants, and that this similarity is avoided within Arabic roots. Coetzee & Pater instead posit relativized OCP constraints interlaced with lexically-specific faithfulness constraints in order to describe the effects of similarity avoidance in Muna roots. However, both theories would be unable to describe both (a) the OCP effects in lexical consonant cooccurrence and (b) the OCP effects in fixed segment /t/-use in echo reduplication. While the cooccurrence patterns of certain Bengali consonants are clearly over- or underrepresented, these phenomena do not correspond to the over- or underrepresentation of fixed segment /t/ use in echo reduplicants of words starting with those consonants.

This observed distinction between patterns in the lexicon and patterns in reduplication phenomena suggests that the same language can have two sets of active OCP constraints, with one set maintaining similarity avoidance by preventing roots with certain consonant combinations from entering the lexicon and another set actively promoting use of particular fixed segments in the grammar in order to create the most dissimilarity between a base and its echo reduplicant. Another analysis would posit that cooccurrence restrictions in the lexicon are shaped by generations of different OCP constraints that are no longer synchronically active, while similarity

avoidance in the productive aspects of grammar (*e.g.* in echo reduplication) obey a synchronically active OCP constraint (or set of constraints) that has little to no effect on the lexicon.

Assigning a particular weight to each feature used in the phoneme inventory seems to be the most effective metric in describing the data observed in this study. Features that play a significant role in distinguishing consonants from one another in a particular language may be weighted more heavily to reflect their relative importance. Other features, especially those that are only used to further distinguish consonants that already do not share very many features, would be assigned lower weights. While all languages presumably have basically the same feature set, this metric makes allowances for language-specific distinctions in the phoneme inventory. For example, a language like Bengali with a four-way laryngeal distinction in stops (*e.g.* voiceless unaspirated, voiceless aspirated, voiced unaspirated, voiced breathy aspirated) would presumably weight the features [spread glottis] and [voice] more heavily than a language like Hawaiian where aspiration and voicing are not contrastive. This balance of universality and language-specificity is, I imagine, a goal of both the Frisch *et al.* study and the Coetzee & Pater study, as both the Shared Natural Classes metric and the relativized OCP hierarchy involve combining universal components (*e.g.* features, constraints) with language-specific features (*e.g.* the phoneme inventory, particular combinations of features). The feature weighting metric allows for one universal component (*i.e.* features) and one language-specific feature (*i.e.* weights) to determine

consonant similarity, maintaining a balance between application to Bengali and the applicability elsewhere.

If speakers do measure similarity by summing the values of weighted features, where do they acquire these weights? Although this study does not attempt to answer this question empirically, some possible sources of feature weights are discussed here merely as speculations and hypotheses for further investigation. The simplest, although probably least likely, hypothesis is that (1) feature weights are universal, as are the features they associate with. Speakers of all languages would be predicted to judge the similarity of consonants in exactly the same way, with the phoneme inventory as the only language-specific component to the calculations. Another hypothesis could state that (2) weights are based on language-specific perceptual similarity. This is a testable hypothesis, as speakers of a language could be asked to participate in a study in which they would be asked to discern which consonants are more similar to each other than to other consonants. This hypothesis incorporates a language-specific component that would presumably be shaped by the phoneme inventory of the language. A language with eight coronal stops, such as Bengali, could be presumed to exaggerate the perceptual distinctions within the coronal place of articulation, thus requiring that the feature set as we know it would have to be adjusted for this perceptual stretching. Another interpretation of this hypothesis would be that (3) feature weights are basically universal, but can be shaped by lexical frequency or consonant cooccurrence in roots. This hypothesis is attractive for the balance between

universality and language-specificity, but the consonant cooccurrence data seen in this study suggest that the lexicon has little or nothing to do with OCP effects in echo fixed segment reduplication. Another hypothesis could be that (4) weights are incorporated into the grammar as the speaker acquires the construction in which it will be used. In the case of Bengali, Common East Bengali speakers would acquire the feature weights of their language as they acquire the echo fixed segment reduplication construction, possibly adjusting weights as they are exposed to more examples of the reduplication. This is not an unlikely hypothesis, although it does potentially limit the applicability of the feature weights to reapplication in the very construction the weights were acquired from. Of course, there is always the possibility that (5) speakers acquire these feature weights from some other component of the grammar, and then apply the similarity calculations drawn from those weights to fixed-segment reduplication; unfortunately, as far as I know and as far as the literature describes, no other productive construction in Bengali employs consonant similarity to the extent that fixed segment echo reduplication has been seen to do.

In addition to exploring the source of these feature weights, further investigation into the differences between reduplication patterns of native and borrowed words, into the reduplicative behavior of cluster-initial words, into the consonants cooccurrence restrictions in a much wider range of roots in the lexicon, and into the possible existence of sub-patterns in the data set would be tremendously beneficial in better understanding the metric or metrics of similarity used in the Bengali language. Also, as

this study attempts to very crudely compare existing theories of similarity, the issue of free parameters deserves some mentioning. Using the feature weighting metric, which has been shown to more accurately model the Common East Bengali data, one must consider that for each feature in the grammar, there would be one free parameter, *i.e.* its assigned weight. Similarly, as each relativized OCP constraint in the Coetzee & Pater analysis is posited based on the patterns of the lexicon of the language, the relativized OCP metric involves $n-1$ free parameters, where n is equal to the number of constraints posited. These models stand in contrast with the Frisch, *et al.* shared natural classes metric, which in its basic form requires no free parameters. The metric derives similarity scores simply based on the feature specifications of the phonemes in a language, which is surely an advantage of this metric over those that require additional information.

It appears that while Common East Bengali speakers indeed measure the similarity of consonants in fixed-segment echo reduplication, they do not base these calculations on the raw number of natural classes shared across each consonant pair, the severity of relativized OCP violations accrued by each consonant pair, or the root-internal cooccurrence restrictions on each consonant pair. The data observed in this study suggests that speakers may be calculating the similarity of consonants by summing together the weights of the features they share, where certain features would be more heavily weighted in particular languages. Comparing the observed dissimilation patterns in Common East Bengali fixed-segment echo reduplication to the predictions

of these different similarity metrics calls into question both the universality of any measure of similarity, and the tacit assumption that OCP effects in the lexicon should pattern with productive dissimilation patterns in the grammar.

5 APPENDICES

5.1 APPENDIX A: TOTAL REDUPLICATION TYPES IN BENGALI

Total Reduplication is used in Bengali to derive constructions that refer to ongoing actions, attenuation⁴¹, adjectives, distributive/plurality (of adjectives, locatives, or pronominals), plural numerals, manner adverbs, frequency/temporal adverbs, intentional adverbs, reciprocal adverbs, sequences, adjectival intensification, habitual actions, and habitual behavior, as shown in the following examples.⁴² The following is not meant to be an exhaustive list of Bengali reduplication types; many more are not mentioned simply for sake of brevity.

Ongoing Actions

- | | | |
|--------|------------------------------------|--|
| (32) | tʃillajte | ‘to yell’ |
| | tʃuj-ki tʃillajte ajsos? | ‘Have you come to yell?’ |
| | tʃuj-ki tʃillajte tʃillajte ajsos? | ‘Have you come yelling (all the way)?’ |

Attenuation

⁴¹ This label, along with many other labels for reduplicative constructions described in this study, is taken from Moravcsik (1978).

⁴² The examples in these appendices, other than those explicitly described otherwise, were constructed with the help of my primary consultant, Farida Amin Khan.

- (33) dzɔɪ
 hæɪ dzɔɪ lagse.
 hæɪ dzɔɪ dzɔɪ lagt̪ese.
 ‘fever’
 ‘She has a fever.’
 ‘She’s feeling a little feverish.’

Adjectives

- (34) haʃi⁴³
 haʃi haʃi muk^h
 ‘smile’
 ‘smiling face’

Distributive Plurality (Adjectives)

- (35) b^hətka
 ami b^hətka dzuʈa piŋdi na.
 ami b^hətka b^hətka dzuʈa piŋdi na.
 ‘bulky’
 ‘I don’t wear bulky shoes.’
 ‘I don’t wear (any sort of) bulky shoes.’

Distributive Plurality (Locatives)

- (36) pəðe
 pəðe pəðe biɸə.
 ‘at a step’, ‘at a position’
 ‘There are problems at every step.’

Distributive Plurality (Pronouns)

- (37) ke:
 ke: pəɸlo?
 ke: ke: pəɸlo?
 ‘who’
 ‘Who farted?’
 ‘Who (all) farted?’ (requesting a list)

Plural Numerals

⁴³ Example drawn from Bykova (1981:105).

- | | |
|---|--|
| (38) lak ^h
æk lak ^h manuʃ
lak ^h lak ^h manuʃ | ‘hundred thousand’
‘one hundred thousand people’
‘hundreds of thousands of people’ |
|---|--|

Manner Adverbs

- | | |
|--|--|
| (39) kan
amaɹ kane
amaɹ kane kane kəɔ. | ‘ear’
‘in my ear’
‘Whisper it to me.’ (<i>lit.</i> ‘Talk in my ear ear.’) |
|--|--|

Frequency/Temporal Adverbs

- | | |
|---|---|
| (40) ʃaɽ ^{he}
ʃaɽ ^{he} ʃaɽ ^{he} | ‘(along) with’
‘simultaneously’, ‘immediately’ |
|---|---|

Intentional Adverbs

- | | |
|--|--|
| (41) aɽte
aɽte haɽtesilen.
aɽte aɽte haɽtesilen. | ‘slowly’
‘You were walking slowly.’
‘You were (purposely) walking slowly.’ |
|--|--|

Reciprocal Adverbs

- | | |
|--|---|
| (42) alaɽa
alaɽa kojɹa ɹak ^h o.
alaɽa alaɽa kojɹa ɹak ^h o. | ‘separate, apart’
‘Set it apart.’
‘Separate them from one another.’ |
|--|---|

Sequences

- (43) $\underset{\cdot}{d}in$ 'day'
 $\underset{\cdot}{t}uj$ $\underset{\cdot}{s}ukaj\underset{\cdot}{t}esos.$ 'You are getting thinner.'
 $\underset{\cdot}{t}uj$ $\underset{\cdot}{d}in$ $\underset{\cdot}{d}in$ $\underset{\cdot}{s}ukaj\underset{\cdot}{t}esos.$ 'You are getting thinner day by day.'

Adjectival Intensification

- (44) $d\underset{\cdot}{z}^{\text{h}}ap\underset{\cdot}{s}a^{44}$ 'hazy', 'dim'
 $d\underset{\cdot}{z}^{\text{h}}ap\underset{\cdot}{s}a$ $d\underset{\cdot}{z}^{\text{h}}ap\underset{\cdot}{s}a$ $d\underset{\cdot}{z}osna$ 'very dim moonlight'

Habitual Actions

- (45) $\underset{\cdot}{s}uja$ 'having lain down, while lying down'
 $\underset{\cdot}{s}uja$ $tivi$ $\underset{\cdot}{d}æk^{\text{h}}e.$ 'She's watching TV lying down.'
 $\underset{\cdot}{s}uja$ $\underset{\cdot}{s}uja$ $tivi$ $\underset{\cdot}{d}æk^{\text{h}}e.$ 'She (regularly) watches TV lying down.'

Habitual Behavior

- (46) $k^{\text{h}}it$ 'grit'
 $k^{\text{h}}it$ $k^{\text{h}}it$ 'overzealous fault-finding'

*Continuous Noise or Action (Onomatopoeic)*⁴⁵

- (47) $b^{\text{h}}æ:$ 'wah' (imitative of loud wailing)
 $b^{\text{h}}æ:$ $koj\underset{\cdot}{a}$ $kand\underset{\cdot}{lo}.$ 'He let out a wail.'

⁴⁴ Example drawn from Bykova (1981:104).

⁴⁵ A subset of this reduplicative construction includes reduplications for which there is no nonreduplicated stem, as in [$\underset{\cdot}{d}^{\text{h}}u:$ $\underset{\cdot}{d}^{\text{h}}u:$] 'desolate' (imitative of a barren wasteland). There is no bare stem * $[\underset{\cdot}{d}^{\text{h}}u:]$. Such reduplications usually denote qualities that do not involve sound, but denote permanent or behavioral qualities.

5.3 APPENDIX C: FEATURE SPECIFICATIONS USED

Common East Bengali Phonemes

Bengali Consonant	consonantal	sonorant	continuant (artic.)	continuant (acou.)	delayed release	approximant	nasal	voice	spread glottis	LABIAL	labiodental	CORONAL	anterior	distributed	strident	lateral	DORSAL
k	+	-	-	-	-	-	-	-	-	-	-	-	0	0	0	-	+
k ^h	+	-	-	-	-	-	-	-	+	-	-	-	0	0	0	-	+
g	+	-	-	-	-	-	-	+	-	-	-	-	0	0	0	-	+
g ^h	+	-	-	-	-	-	-	+	+	-	-	-	0	0	0	-	+
ŋ	+	+	-	+	-	-	+	+	-	-	-	-	0	0	0	-	+
tʃ	+	-	-	-	+	-	-	-	-	-	-	+	-	+	+	-	-
tʃ ^h	+	-	-	-	+	-	-	-	+	-	-	+	-	+	+	-	-
dʒ	+	-	-	-	+	-	-	+	-	-	-	+	-	+	+	-	-
dʒ ^h	+	-	-	-	+	-	-	+	+	-	-	+	-	+	+	-	-
t	+	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-
t ^h	+	-	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-
d	+	-	-	-	-	-	-	+	-	-	-	+	+	-	-	-	-
d ^h	+	-	-	-	-	-	-	+	+	-	-	+	+	-	-	-	-
ṭ	+	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-
ṭ ^h	+	-	-	-	-	-	-	-	+	-	-	+	+	+	-	-	-
ḍ	+	-	-	-	-	-	-	+	-	-	-	+	+	+	-	-	-
ḍ ^h	+	-	-	-	-	-	-	+	+	-	-	+	+	+	-	-	-
n	+	+	-	+	-	-	+	+	-	-	-	+	+	-	-	-	-
p	+	-	-	-	-	-	-	-	-	+	-	-	0	0	0	-	-
f	+	-	+	+	+	-	-	-	-	+	+	-	0	0	0	-	-
b	+	-	-	-	-	-	-	+	-	+	-	-	0	0	0	-	-
b ^h	+	-	-	-	-	-	-	+	+	+	-	-	0	0	0	-	-
m	+	+	-	+	-	-	+	+	-	+	-	-	0	0	0	-	-
ɹ	-	+	+	+	+	+	-	+	-	-	-	+	-	+	-	-	-
l	+	+	+	+	+	+	-	+	-	-	-	+	+	-	-	+	-
ʃ	+	-	+	+	+	-	-	-	-	-	-	+	-	+	+	-	-
s	+	-	+	+	+	-	-	-	-	-	-	+	+	-	+	-	-
h	-	-	+	+	+	-	-	-	+	-	-	-	0	0	0	-	-
z	+	-	+	+	+	-	-	+	-	-	-	+	+	-	+	-	-

Additional English Phonemes

English Consonant	consonantal	sonorant	continuant (artic.)	continuant (acou.)	delayed release	approximant	nasal	voice	spread glottis	LABIAL	labiodental	CORONAL	anterior	distributed	strident	lateral	DORSAL
θ	+	-	+	+	+	-	-	-	-	-	-	+	+	+	-	-	-
ð	+	-	+	+	+	-	-	+	-	-	-	+	+	+	-	-	-
ʒ	+	-	+	+	+	-	-	+	-	-	-	+	-	+	+	-	-
v	+	-	+	+	+	-	-	+	-	+	+	-	0	0	0	-	-
j	-	+	+	+	+	+	-	+	-	-	-	-	0	0	0	-	+
w	-	+	+	+	+	+	-	+	-	+	-	-	0	0	0	-	+
m	-	-	+	+	+	-	-	+	-	+	-	-	0	0	0	-	+

5.4 APPENDIX B: STIMULUS SETS I AND II (NATIVE STANDARD KOLKATA BENGALI AND COMMON EAST BENGALI WORDS)

Initial	Standard Kolkata Bengali in Orthography	Standard Kolkata Bengali in Transcription	Common East Bengali in Orthography	Common East Bengali in Transcription	Gloss
t	টাকা	taka	টাকা	taka	'money'
t	টুপি	tupi	টুপি	tupi	'hat'
t	টিকে	tike	টিকা	tika	'vaccine'
t	টিকলি	tikli	টিকলি	tikli	type of jewelry
t	টুকরো	tukɔ	টুকরা	tukɔa	'piece'
t	টেরা	tæɔa	ট্যাগরা	tægɔa	'cross-eyed'
t	টিপে	tipe	টিইপ্যা	tippa	'h. pressed' ⁵⁰
t	টেনে	tene	টাইন্যা	tajnna	'h. pulled'

⁵⁰ In this table, "h." stands for "having", so "h. pressed" is to be understood as the completive participle "having pressed".

t ^h	ঠোঙা	t ^h oŋa	ঠোঙ্গা	t ^h oŋga	'bag'
t ^h	ঠোকা	t ^h oka	ঠোকা	t ^h oka	'knock'
t ^h	ঠেকা	t ^h æka	ঠেকা	t ^h æka	'obstruction'
t ^h	ঠেলে	t ^h eɽe	ঠেইল্যা	t ^h ejlla	'h. pushed'
t ^h	ঠেসে	t ^h eʃe	ঠাইস্যা	t ^h ajʃʃa	'h. stuffed'
d	ডিবে	dibe	ডিব্যা	dibba	'container'
d	ডানা	dana	ডানা	dana	'fin'
d	ডলে	dole	ডইল্যা	dojlla	'h. rubbed'
d	ডুবে	dube	ডুইব্যা	dujbba	'h. drowned'
d	ডেকে	deke	ডাইক্যা	dajkka	'h. called'
t̪	তালা	t̪ala	তালা	t̪ala	'lock'
t̪	তুলো	t̪ulo	তুলা	t̪ula	'cotton'
t̪	তালি	t̪ali	তালি	t̪ali	'applause'
t̪	তাজা	t̪adʒa	তাজা	t̪aza	'fresh'
t̪	তুলে	t̪ule	তুইল্যা	t̪ujlla	'h. thrown'
tʃ	চুলো	tʃulo	চুলা	tʃula	'oven', 'stove'
tʃ	চাকরি	tʃakari	চাকরি	tʃakari	'job'
tʃ	চুড়ি	tʃuri	চুড়ি	tʃuri	'bracelet'
tʃ	চেপে	tʃepe	চাইপ্যা	tʃajppa	'h. pressed'
tʃ	চুষে	tʃuʃe	চুষিয়া	tʃuʃʃa	'h. sucked'
t̪ ^h	থলে	t̪ ^h ole	থইল্যা	t̪ ^h ojlla	'satchel'
t̪ ^h	থাপ্পর	t̪ ^h appaɽ	থাপ্পর	t̪ ^h appaɽ	'slap'
t̪ ^h	থেমে	t̪ ^h eme	থাইম্যা	t̪ ^h ajmma	'h. stopped'
ɖ	দুঃখ	ɖukk ^h o	দুঃখ	ɖukk ^h o	'sadness'
ɖ	দেৱী	ɖe:ɽi	দেৱী	ɖe:ɽi	'late'
ɖ	দিয়ে	ɖie	দিয়া	ɖia	'h. given'
tʃ ^h /s	ছিলকে	tʃ ^h ilke	ছিলকা	silka	'peel'
tʃ ^h /s	ছেড়ে	tʃ ^h ere	ছাইড়া	sajɽa	'h. let go'
tʃ ^h /s	ছিলে	tʃ ^h ile	ছিইল্যা	silla	'h. peeled'
m	মুড়ি	muri	মুড়ি	mu:ɽi	'puffed rice'
m	মুছে	mutʃ ^h e	মুইছ্যা	mujttʃ ^h a	'h. wiped'
m	মেরে	me:ɽe	মাইরা	majɽa	'h. hit'

f	ফেনা	fæna	ফেনা	fæna	'foam'
f	ফাঁকি	făki	ফাঁকি	faki	'fraud'
f	ফুলে	fule	ফুইল্যা	fujlla	'h. swollen'
ɟ	রুমাল	ɟumal	রুমাল	ɟumal	'handkerchief'
ɟ	রেগে	ɟege	রাইগ্যা	ɟajgga	'h. gotten angry'
ɟ	রুঁধে	ɟeɖ ^h e	রাইন্ধ্যা	ɟajɖɔ	'h. cooked'
ʃ	শুনে	ʃune	শুইন্যা	ʃujɳna	'h. heard'
ʃ	শাড়ী	ʃari	শাড়ী	ʃari	'sari'
n	নিয়ম	niom	নিয়ম	niom	'custom'
n	নেমে	neme	নাইম্যা	najmma	'h. gone down'
l	লোহা	loha	লোহা	loha	'iron', 'nail'
l	লেগে	lege	লাইগ্যা	lajgga	'h. struck'
k	কাশি	kăʃi	কাশি	kăʃi	'cough'
k	কেঁদে	kêɖe	কাইন্ধ্যা	kajɖɔ	'h. cried'
p	পানি	pani	পানি	pani	'water'
p	পড়ে	pore	পইড়া	pojɔ	'h. read'
b ^h	ভালো	b ^h alo	ভালো	b ^h alo	'good'
b ^h	ভিজে	b ^h idʒe	ভিইজ্যা	b ^h iddʒa	'h. gotten wet'
h	হেসে	heʃe	হাইস্যা	hajʃʃa	'h. laughed'
h	হেরে	heɾe	হাইরা	hajɔ	'h. lost'

5.5 APPENDIX C: STIMULUS SET III (ENGLISH LOANWORDS)

<i>Bengali Initial</i>	<i>English Initial</i>	<i>Bengali in Orthography</i>	<i>Bengali in Transcription</i>	<i>Gloss</i>
t	t	টাইপিং	tajpiŋ	'typing'
t	t	টিভি	tivi	'TV'
t	t	টানেল	tanel	'tunnel'
t	t	ট্যাক্সি	tæksi	'taxi'
t	t	টেবিল	tebil	'table'
t	t	টেনশন	tenʃon	'tension'

t	t	টার্কি	taʊki	'turkey'
t	t	টাইমিং	tajmiŋ	'timing'
d	d	ডিফারেন্স	difɛns	'difference'
d	d	ডান্সিং	dansiŋ	'dancing'
d	d	ডিজাইন	dizajn	'design'
d	d	ডিগ্রী	digɹi	'degree'
d	d	ডীলার	dilaɹ	'dealer'
tʃ	tʃ	চিমনী	tʃimni	'chimney'
tʃ	tʃ	চিকেন	tʃiken	'chicken'
tʃ	tʃ	চেরী	tʃeɹi	'cherry'
tʃ	tʃ	চ্যানেল	tʃænel	'channel'
tʃ	tʃ	চাইনীজ	tʃajniːz	'Chinese'
t̪ʰ	θ	থার্মাল	t̪ʰaʊmal	'thermal'
t̪ʰ	θ	থার্মস	t̪ʰaʊmos	'thermos'
t̪ʰ	θ	থীসিস	t̪ʰisis	'thesis'
s	s	সার্ভিস	sɑʊvis	'service'
s	s	সকার	səkɑɹ	'soccer'
s	s	সিরাপ	sɪɹap	'syrup'
s	s	সোফা	sofa	'sofa'
s	s	সিম্পাল	sɪmpal	'simple'
s	s	সাইনে	sajne	'on the sign'
s	s	সামার	samaɹ	'summer'
f	f	ফিগার	figɑɹ	'figure'
f	f	ফিনিশ	finiʃ	'finish'
m	m	মেশিন	meʃin	'machine'
m	m	ম্যাজিক	mædʒik	'magic'
m	m	ম্যানেজ	mænedʒ	'manage'
m	m	মেয়র	meɹɑɹ	'mayor'
m	m	মেম্বর	membaɹ	'member'
ɹ	ɹ	রিসীভ	ɹisiv	'receive'
ɹ	ɹ	রীজন	ɹizən	'reason'
ɹ	ɹ	রিদাম	ɹiðəm	'rhythm'

ɹ	ɹ	রীফিল	ɹifil	'refill'
ɹ	ɹ	রিফোর্ম	ɹifɔɹm	'reform'
ɹ	ɹ	রিল্যাক্স	ɹilæks	'relax'
n	n	নভেল	nɔvel	'novel'
n	n	নার্সিং	narsɪŋ	'nursing'
n	n	নর্মাল	nɔɹmal	'normal'
ʃ	ʃ	শাওয়ার	ʃaɔɹɹ	'shower'
ʃ	ʃ	শ্যাম্পু	ʃæmpu	'shampoo'
l	l	লয়ার	lɔɹɹ	'lawyer'
l	l	লাইসেন্স	lajsens	'license'
l	l	লিভার	livaɹ	'liver'
k	k	ক্যান্সার	kænsaɹ	'cancer'
k	k	ক্যাম্পিং	kæmpɪŋ	'camping'
k	k	ক্যান্সেল	kænsel	'cancel'
k	k	কফি	kɔfi	'coffee'
p	p	পার্কিং	paɹkiŋ	'parking'
p	p	প্যালেস	pæles	'palace'
v	v	ভাইরাস	vajɹas	'virus'
v	v	ভ্যালী	væli	'valley'
v	v	ভিসা	visa	'visa'
h	h	হাইজ্যাক	hajdzæk	'hijack'
h	h	হ্যামার	hæmaɹ	'hammer'

5.6 APPENDIX D: STIMULUS SET IV (CLUSTER-INITIAL LOANWORDS)

Bengali Initial	English/ Sanskrit Initial	Bengali in Orthography	Bengali in Transcription	Gloss
tɹ	tɹ	ট্রামে	tɹæme	'by tram'
tɹ	tɹ	ট্রাকে	tɹake	'by truck'
tɹ	tɹ	ট্রেনে	tɹene	'by train'
tɹ	tɹ	ট্রান্কে	tɹaŋke	'in trunks'
tɹ	tɹ	ট্রিগার	tɹigaɹ	'trigger'

sm	sm	সুদলী	smudli	'smoothly'
fɹ	fɹ	ফ্রান্সে	fɹanse	'in France'
fɹ	fɹ	ফ্রীজার	fɹizaɹ	'freezer'
fɹ	fɹ	ফ্রিজবী	fɹizbi	'frisbee'
sɹ	ʃɹ	শ্রাবণ	sɹabon	<i>Bengali month</i>
sɹ	ʃɹ	শ্রমিক	sɹomik	'worker'
sɹ	ʃɹ	শ্রদ্ধা	sɹoɖɖ ^h a	'respect'
sɹ	ʃɹ	শ্রেণী	sɹeni	'class'
sn	sn	স্নেহ	sneho	'affection'
sn	sn	স্নিগ্ধ	snigd ^h o	'cool'
sn	sn	স্নোম্যান	snomæn	'snowman'
sn	sn	স্নীকার	snikaɹ	'sneaker'
sk	sk	স্কুলে	skule	'at school'
sk	sk	স্কেজুল	skedzɹl	'schedule'
kɹ	k	কৃষি	kɹiʃi	'agriculture'
kɹ	k	কৃমি	kɹimi	'worm'
kɹ	kɹ	ক্রীড়া	kɹira	'sport'
kɹ	kɹ	ক্র্যাকার	kɹækɹɹ	'cracker'
kɹ	kɹ	ক্রিপাল	kɹipal	'cripple'
sp	sp	স্পর্শ	spɹɹʃo	'touch'
sp	sp	স্পেশাল	speʃal	'special'
sp	sp	স্পীকার	spikaɹ	'speaker'
sp	sp	স্প্যানিশ	spæniʃ	'Spanish'

5.7 APPENDIX E: PAIRWISE T-TEST RESULTS

The results of pairwise *t*-tests to show statistical distinctions between the likelihood of certain base-initial consonants using fixed segment /t/ significantly more

often than other base-initial consonants are shown below. The α value for the pairwise t -tests was set at 0.05 and adjusted for multiple comparisons using the Bonferroni method. In the interest of space, individual p -values were not included. “S” indicates that the p -value was below 0.05 (significant), while “N” indicates that the p -value was above 0.05 (not significant).

	b ^h	tʃ	d	f	h	k	l	m	n	p	ṭ	ṭ ^h	ɹ	s	ʃ	t	t ^h	
tʃ	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
d	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
f	N	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
h	N	N	N	S	-	-	-	-	-	-	-	-	-	-	-	-	-	
k	N	N	N	N	N	-	-	-	-	-	-	-	-	-	-	-	-	
l	N	N	N	N	N	N	-	-	-	-	-	-	-	-	-	-	-	
m	N	N	N	N	S	N	N	-	-	-	-	-	-	-	-	-	-	
n	N	N	N	N	N	N	N	N	-	-	-	-	-	-	-	-	-	
p	N	N	N	N	S	N	N	N	N	-	-	-	-	-	-	-	-	
ṭ	N	N	N	N	N	N	N	N	N	N	-	-	-	-	-	-	-	
ṭ ^h	N	N	N	N	N	N	N	N	N	N	N	-	-	-	-	-	-	
ɹ	N	N	N	N	N	N	N	N	N	N	N	N	-	-	-	-	-	
s	N	N	N	N	N	N	N	N	N	N	N	N	N	-	-	-	-	
ʃ	N	N	N	N	N	N	N	N	N	N	N	N	N	N	-	-	-	
t	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	-	
t ^h	S	S	N	S	N	S	S	S	S	S	N	N	S	N	S	N	-	
ɖ	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S

6 REFERENCES

- আলম, মাহবুবুল (২০০০ ইং) Alam, Mahbubul (2000). *ভাষা সৌরভ: ব্যাকরণ ও রচনা - ষষ্ঠ সংস্করণ* Bhasha Shourôbh: Bêkorôn o Rôcona - Shôshtho Shôngskorôn (The Fragrance of Language: Grammar and Composition - Sixth Edition). Dhaka, Bangladesh: Muhammad Ayyub Ali Publishers.
- van den Berg, René (1989). *A grammar of the Muna language*. Dordrecht, Netherlands: Foris.
- Bhaskararao, Peri (1977). *Reduplication and onomatopoeia in Telugu*. Poona, India: Centre of Advanced Study in Linguistics.
- Bykova, E. M. (1981). *The Bengali Language*. Moscow, USSR [now Russian Federation]: Nauka Publishing House.
- Chatterji, Suniti Kumar (1970). *The Origin and Development of the Bengali Language: Part I: Introduction, Phonology*. London, United Kingdom: George Allen & Unwin, Ltd.
- চৌধুরী, শহীদ মুনীর, শহীদ মফাজ্জল হায়দার চৌধুরী, মরহুম ইব্রাহীম খলিল, ডঃ কাজী দীন মুহম্মদ, শিবপ্রশন্ন লাহিড়ী (১৯৯৭ ইং) Choudhuri, Shohid Munir, Shohid Mofazzol Haedar Choudhuri, the late Ibrahim Kholil, Dr. Kazi Din Muhammod, Shiboproshonno Lahiri (1997). *বাংলা ভাষার ব্যাকরণ: দ্বিতীয় সংস্করণ* Bangla Bhashar Bêkoron: Ditio Shôngskôrôn (The Grammar of the Bengali Language: Second Edition). Motijhil, Dhaka, Bangladesh: National Education System and Textbook Board.
- Coetzee, Andries & Joe Pater (2005). Gradient phonotactics in Muna and Optimality Theory. University of Michigan & University of Massachusetts, Amherst.
- Dev, Ashutosh (1999). *Students' Favourite Dictionary: Bengali-English*. Kolkata, India: A.C. Mazumder Dev Sahitya Kutir Publishers Private Ltd.
- Esposito, Christina M., Sameer ud Dowla Khan, and Alex Hurst (2005). Breathly Nasals and /Nh/ Clusters in Bengali, Hindi, and Marathi. *UCLA Working Papers in Phonetics*, no. 104.
- Fitzpatrick, Justin & Andrew Nevins (2002). Phonological Occurrences: Relations and Copying. Accepted for publication in the Proceedings of the 2nd North American Phonology Conference, Montreal, April 2002.

- Fitzpatrick-Cole, Jennifer (1994).** *The Prosodic Domain Hierarchy in Reduplication*. Unpublished Ph.D. dissertation, Stanford University.
- Fitzpatrick-Cole, Jennifer (1996).** Reduplication meets the phonological phrase in Bengali. *The Linguistic Review* 13: 305–356.
- Frisch, Stefan, Janet Pierrehumbert, & Michael Broe (2004).** Similarity Avoidance and the OCP. *Natural Language and Linguistic Theory* 22.179-228.
- Gordon, Raymond G., Jr. (ed.) (2005).** *Ethnologue: Languages of the World*, Fifteenth edition. Dallas, Tex.: SIL International. Online version: www.ethnologue.com.
- Greenberg, Joseph H. (1950).** The patterning of root morphemes in Semitic. *Word* 6.162-181.
- Hai, Muhammad Abdul (1960).** *A Phonetic and Phonological Study of Nasals and Nasalization in Bengali*. Dacca, East Pakistan: Dacca University Press [now Dhaka, Bangladesh: Dhaka University Press].
- Hairer, Martin (n.d.).** Amadeus II, version 3.7.1. Software program.
- Halder, Gopal (1986).** *A Comparative Grammar of East Bengali Dialects*. Calcutta [now Kolkata], India: Puthipatra.
- Jha, Chandra Kant, Kamalesh Sadanand, and K.G. Vijaykrishnan (1997).** “Some Salient Properties of Echo Word Formation in Indian Languages,” in M. Hariprasad, H. Nagarajan, P. Madhavan and K.G. Vijaykrishnan (eds). *Phases and Interfaces of Morphology*. Hyderabad, India: CIEFL Publications.
- Keane, Elinor (2001).** Echo Words in Tamil. Doctoral Thesis, Merton College, Oxford.
- Khan, Sameer ud Dowla (2005).** Morphological Constraints on Reduplicant Shapes. Handout presented at UCLA Phonology Seminar, 1 February 2005. Ms.
- Khan, Sameer ud Dowla (2004).** Antifaithfulness and Phonotactics in Bengali t-Reduplication. Term paper for UCLA Spring 2004 course on Phonology, 15 June 2004. Ms.
- Koul, Omkar Nath (n.d.).** Reduplication in Kashmiri. Indian Institute of Language Studies.
- Leben, Will (1973).** Suprasegmental Phonology. Doctoral dissertation, Massachusetts Institute of Technology.

Lidz, Jeffrey (2001). Echo Reduplication in Kannada and the Theory of Word Formation. *The Linguistic Review* 18: 375-394.

Mallik, Bhakti Prasad, Nikhilesh Bhattacharya, Subhas Chandra Kundu, and Mina Dawn (1998). *The Phonemic and Morphemic Frequencies of the Bengali Language*. Calcutta [now Kolkata], India: The Asiatic Society.

Malten, Thomas (1989). *Reduplizierte Verbstämme im Tamil*. Stuttgart, Germany: Steiner Verlag Wiesbaden.

McCarthy, John J. (1986). OCP Effects: Gemination and Antigemination. *Linguistic Inquiry* 17. 207-264.

McCarthy, John J. (1988). Feature geometry and dependency: A review. *Phonetica* 45. 84-108.

McCarthy, John J. (1994). The phonetics and phonology of Semitic pharyngeals. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology 3*, ed. Patricia Keating, 191-233. Cambridge: Cambridge University Press.

McCarthy, John J. & Alan Prince (1995). Faithfulness and Reduplicative Identity. In Jill Beckman, Suzanne Urbanczyk and Laura Walsh Dickey (eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*. Amherst, USA.

Moravcsik, Edith A. (1978). Reduplicative constructions. In Joseph Greenberg (ed.), *Universals of Human Language*, volume 3: Word Structure (pp. 297- 334). Stanford, USA: Stanford University Press.

Myung, In Jae (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* 47, pp 90-100.

Nevins, Andrew & Bert Vaux (2003). Metalinguistic Shmetalinguistic: the Phonology of shm-Reduplication. Massachusetts Institute of Technology and Harvard University.

Nevins, Andrew & Michael Wagner (2001). Backcopying, Shmackcopying: Correspondence Relations and Dissimilation in Echo Reduplication. Presented at the LOT Summer School, Nijmegen. Ms.

Padgett, Jaye (1995). *Stricture in Feature Geometry*. Stanford: Center for the Study of Language and Information - Dissertations in Linguistics.

Pierrehumbert, Janet B. (1993). Dissimilarity in the Arabic Verbal Roots. In *Proceedings of the North East Linguistics Society* 23, pp. 367-381.

Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science. Cambridge, USA: Massachusetts Institute of Technology Press.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.

Ramaswami, N. (1999). *Common Linguistic Features in Indian Languages: Phonetics*. Mysore, India: Central Institute of Indian Languages.

Ray, Punya Sloka, Muhammad Abdul Hai, & Lila Ray (1966). *Bengali Language Handbook*. Washington, United States: Center for Applied Linguistics.

Sjölander, Kåre & Jonas Beskow (2004). WaveSurfer 1.6.2. Software freely available.

Steriade, Donca (1995). Positional Neutralization. UCLA Department of Linguistics. Ms.

Trivedi, G.M. (1990). "Echo Formation," in Shree Krisha (ed.) *Linguistic Traits Across Language Boundaries*. Anthropological Survey of India. Calcutta [now Kolkata], India: Ministry of Human Resource Development.

Tunga, Sudhansu S. (1995). *Bengali and Other Related Dialects of South Assam*. New Delhi, India: Mittal Publications.

Wood, Alan (2005). Test for Unicode support in Web browsers: Bengali. Alan Wood's Unicode Resources, last updated 21 April 2005.
www.alanwood.net/unicode/bengali.html

Yip, Moira (1998). Identity Avoidance in Phonology and Morphology. S. LaPointe, D. Brentari, and P. Farrell, eds. *Morphology and its Relation to Phonology and Syntax*. Stanford, USA: CSLI Publications, pp 216-246.

Zuraw, Kie (1998). UCLA FeaturePad 1.0. Copyright Regents of the University of California. Software freely available.
www.linguistics.ucla.edu/people/hayes/120a/FeaturePad.htm

Zuraw, Kie (n.d.). Similar.exe. Software program.