What is India speaking: The “Hinglish” invasion

RANA D. PARSHAD,* VINEETA CHAND †, NEHA SINHA, ‡ & NITU KUMARI§

Abstract

While language competition models of diachronic language shift are increasingly sophisticated, drawing on sociolinguistic components like variable language prestige, distance from language centers and intermediate bilingual transitional populations, in one significant way they fall short. They fail to consider contact-based outcomes resulting in mixed language practices, e.g. outcome scenarios such as creoles or unmarked code switching as an emergent communicative norm. On these lines something very interesting is uncovered in India, where traditionally there have been monolingual Hindi speakers and Hindi/English bilinguals, but virtually no monolingual English speakers. While the Indian census data reports a sharp increase in the proportion of Hindi/English bilinguals, we argue that the number of Hindi/English bilinguals in India is inaccurate, given a new class of urban individuals speaking a mixed lect of Hindi and English, popularly known as “Hinglish”. Based on predator-prey, sociolinguistic theories, salient local ecological factors and the rural-urban divide in India, we propose a new mathematical model of interacting monolingual Hindi speakers, Hindi/English bilinguals and Hinglish speakers. The model yields globally asymptotic stable states of coexistence, as well as bilingual extinction. To validate our model, sociolinguistic data from different Indian classes are contrasted with census reports: We see that purported urban Hindi/English bilinguals are unable to maintain fluent Hindi speech and instead produce Hinglish, whereas rural speakers evidence monolingual Hindi. Thus we present evidence for the first time where an unrecognized mixed lect involving English but not “English”, has possibly taken over a sizeable faction of a large global population. This is in direct contrast to perspectives of English encroachment, as well as what is traditionally believed in India about bilinguism: while the approach presented here, a three-species predator-prey model with rural settings as a refuge, is significantly distinct from earlier language competition modeling efforts.

Keywords: language dynamics — three species food chain — diachronic

1 Introduction

Language competition is the central driver of diachronic language change in populations: Language competition models have traditionally considered two competing languages, each spoken by a monolingual group in a population, with the possibility of a bilingual group. Applications of these models has shown that English has out-competed a number of languages such as Scottish Gaelic in Scotland, Welsh in Wales and Mandarin in Singapore. However these models fail to account for mixed code development outcomes. One of the most prominent regions within discussions of ongoing English language shift is India, given its massive population, British colonial history and contemporary role in outsourcing/IT. Mathematical modeling of population-based language shift (generally towards English) have been the subject of intense investigation [1, 34, 38, 32, 40], seeking to explain the process and products of language shift, and typically considering situations of two competing languages, each spoken by a

*Department of Mathematics, Clarkson University, Potsdam, New York, USA
†Department of Language and Linguistics, University of Essex, Colchester, UK.
‡Department of Psychology, Rutgers University, Piscataway, New Jersey, USA
§Department of Mathematics, Clarkson University, Potsdam, NY 13699, USA.
monolingual group, sometimes with an intermediate bilingual group. These models have been successful in replicating and predicting the dynamics of language competition and the resultant decline of one language in specific bilingual socio-historical contexts, such as Scottish Gaelic in Scotland, Welsh in Wales and Mandarin in Singapore, all of which were out-competed by English [1, 2]. Language competition models dealing with two languages frame language shift as motivated by the relative status (prestige) of each language [2]. Mathematically, this leads to dynamics of a stable fixed point \((K,0)\): one of the languages diachronically attains a steady level, while the other goes extinct. This outcome scenario results in wholesale diachronic language shift across a population. Meanwhile, other research has examined the role of bilingualism on language competition: mathematical analysis shows that bilingualism can prove stable coexistence of two languages, if the status of the less prestigious language is protected [36, 42, 35], a scenario reflective of empirical stable bilingual settings. The role of population size and geographical distance on language competition have also been explored [39, 43], demonstrating that sociolinguistic factors strengthen the predictive power of models: language shift is a multifaceted phenomenon. However, past model-based language competition research is strongly oriented towards modeling historically completed changes wherein “good” models accurately match historical data, rather than model building oriented towards contemporary settings where the final outcomes are unknown, and thus far models which explore mixed language outcomes are absent.

We investigate and model contemporary Hindi and English competition in India, in order to explore the range of future language competition outcome scenarios. The Indian situation is very different from contexts like Wales, Scotland and Singapore: the Hindi-speaking population consists of monolingual Hindi speakers and Hindi/English bilinguals, but, as we demonstrate with Indian census data for 40+ years, a diachronically almost non-existent monolingual English speaking population. Recently language competition models have explored competition between a bilingual group and a monolingual group [19], akin to the Indian situation. However, all current language competition models, when applied to our setting, would yield either extinction of monolingual Hindi speakers, or a coexistence state where the bilingual population greatly outnumbers the monolingual Hindi speaking population, because, according to current modeling hypotheses, more speakers will shift from Hindi to English given the higher status of English in India. However, both of these scenarios are extremely unlikely for India, given that 59% of Indians reside under the poverty line [13], with almost no access to English. This prevents complete English acquisition and thus, full language shift to English within lower class populations. Furthermore, one notices in contemporary urban India, that “bilinguals” often use a mixed form of Hindi and English, popularly known as “Hinglish”. Despite informal accounts that “Hinglish” is rapidly gaining tread, no definitive linguistic description or speaker profiles exist. Hinglish is a colloquial umbrella-term encompassing a variety of perspectives on its linguistic structure: it is contrastively framed as isolated English or Hindi borrowings which are grammatically incorporated into the host language, conscious/stylistic code switching involving sequential alternations between codes, and as a mixed language [3, 14]. It is more likely that the linguistic nature of Hinglish varies by sub-population, such that the Hinglish spoken by communities with limited English access will manifest more in terms of borrowings and loanword adoption, while urban communities with more extensive English access may demonstrate a deeper integration involving codeswitching at multiple levels (e.g. including morphemic integration). Across all of these practices, we refrain from locating Hinglish as a creole [41]: there are no linguistic criteria which separate creoles structurally from canonical monoparental languages, and ecologically this setting lacks the geographic isolation from native English, the population-based segregation and the time depth within which traditional creoles have developed. In lieu of the above, a new line of reasoning is called for, to model Hindi/English competition in India. We next examine local Indian ecological factors affecting language competition and selection, and processes and motivations for language innovation and deviation from established practices.

While traditional Hindi/English bilingual older speakers from elite settings are mixed in their evaluations of Hinglish, it is valued by urban, affluent Indian youth over monolingual Hindi or English speech [5]. It is socially salient as demonstrated in informal discourse, popular handbooks, fiction novels, TV shows and films [8, 9, 10, 12, 11], and holds considerable local and covert prestige [18]: while not prescriptively a “proper” code, it indexes a middle ground between upper and lower classes, values and broader dispositions, as a modern but locally grounded way of representing oneself. In addition, there is a reframing in contemporary language evolution research which argues that language hybridization and horizontal language contact (contra vertical monoparental language evolution frameworks) are both natural and common processes [41, 44]. Collectively, the value and social meaning of Hinglish motivate its diachronic uptake, and offer a viable alternative outcome for Indian language shift. However, given the lack of consensus (and remarkably limited discussion) regarding Hinglish linguistic structure or pop-
ulation demographics, and given our modeling agenda, our first goal is to formally define a ‘Hinglish’ speaker, and illustrate their practices with authentic data. We proceed by exploring conversational data from an Indian sub-population, in an attempt to observe interactional trends of the “Hinglish” speaker.

2 An investigation in “Hinglish”:

2.1 Bigg Boss & NDTV

Before we begin, we define the following:

Definition 2.1. (Monolingual Hindi class) Can produce monolingual Hindi, English restricted to limited inclusion of historical indigenizations and contemporary loanwords.

Definition 2.2. (Hindi/English bilingual class) Can produce monolingual Hindi and monolingual English.

Definition 2.3. (Hinglish class) An urban sub-population who cannot produce monolingual Hindi, and/or monolingual English, only Hinglish.

Through differences in monolingual versus codeswitching practices, we aim to isolate from the traditional, census defined, Hindi/English bilingual population a sub-population of Hinglish speakers, who do not speak in monolingual Hindi. We focus on the popular Indian reality TV show Big Boss [BB]. The show is based on the same lines as MTV Real World in the US, and Big Brother in the UK. Essentially participants are isolated from the outer world, with tasks and obstacles used to successively eliminate participants until a ‘winner’ is declared. Participants are from a range of backgrounds including film/TV, sports, beauty, crime and international fame. The context is ostensibly Hindi-only, as a stated rule, with repercussions for English transgressions. The episodes thus represent naturalistic language within a context encouraging and overtly valuing monolingual Hindi. The backgrounds of the participants reveal that there are only two monolingual Hindi speakers in the group, with the rest claiming to be Hindi/English bilinguals. However, given the rules of the game show, and the background of the participants, we expect that the group will be able to communicate amongst each other in monolingual Hindi. Examining two complete seasons, consisting of thirty two participants, we uncover that this is far from true. No participant is able to consistently produce monolingual Hindi, while Big Boss, who arguably has the only scripted role as taskmaster and is only present via his voice, produced no English in season 6.

This warrants the question: Is anyone in this group (except Big Boss) capable of producing monolingual Hindi? To answer this we further examine the “confession room” scenarios in the two seasons. These are essentially one-on-one discourses between a single participant and big boss. In these scenarios, big boss speaks to the participant in the room, in monolingual Hindi. However, only three contestants are able to consistently respond in monolingual Hindi, of which two of these are the monolingual Hindi speakers identified earlier. Thus there is only one purported Hindi/English bilingual, who successfully produces monolingual Hindi (and thus fits with our definition of Hindi/English bilingual). Every other participant exhibits the types of mixed practices mentioned earlier ( that is code-switching, single word insertions etc).

From this sample population of thirty three people (thirty two participants plus big boss), among which we have identified twenty nine Hinglish speakers, two bilingual and two monolinguals, specific interactional outcomes are observed.

1) The Hinglish speaker does not deviate from mixing, whether he/she speaks to a bilingual or a monolingual.

2) The bilinguals and monolinguals mix code when speaking with Hinglish speakers, and do not produce monolingual Hindi. That is the monolingual Hindi speakers, use many more single word English insertions when speaking with a Hinglish speaker, than when speaking with a bilingual producing monolingual Hindi (such as with bigg boss in the confession rooms) or amongst themselves. The
bilingual does the same.

3) The bilinguals and monolinguals are able to produce monolingual Hindi when speaking with other bilinguals and monolinguals.

Since the Hinglish speakers and bilinguals in BB are all from urban India, we decide to next identify a sub-population of Hindi speakers in rural India, and observe closely their language practices. To this end, we examine televised interviews with rural Hindi Belt speakers conducted by NDTV, a well-known news channel in India. The interviews consist of NDTV correspondents traveling to rural parts (some of which are extremely segregated) of India and speaking with rural folk about a host of socio-economic issues. We observe that by and large the rural production monolingual Hindi. English was almost entirely absent in these 189 rural speakers, who can be characterized as lower class and minimally educated villagers. When present, English was revealed as single word insertions, historical indigenizations and/or technology/government related words (e.g. police, telephone) which arguably are established loans into Hindi for all speakers. Our key observations are the following.

1) Rural Indians are able to produce monolingual Hindi when speaking with other bilinguals and monolinguals, thus fitting in with our definition of a monolingual Hindi speaker.
2) Rural Indians do not produce Hinglish when speaking with other bilinguals and monolinguals.

Given these observations, our next goal is to use predator-prey, sociolinguistic and evolutionary linguistic theories to hypothesize certain principles that could explain our observations.

2.2 Mutation, Linguistic Innovation, Fitness and refuge

In evolutionary biology a mutation is defined as any alteration in a gene from its natural state, wherein fitness is defined as the ability of a species to survive and reproduce [23]. Exemplar based theories of cognition and language change offer a means of understanding the emergence of mutations (innovations) within language shift: language shift takes place through the selective (both conscious and unconscious) replication of some linguistic structures and practices over time, wherein accumulated daily production and perception practices (continued selection) diachronically feed into cognitive storage and activation thresholds and permit mutations to both emerge (via contact between classes) and survive (via continued selection). Thus, depending on the selection process, one possible outcome is a mutation or innovation, here defined as the emergence of Hinglish as obligatory code switching. This has been explored recently [21, 20, 33].

The audience design model [31] interprets individual situations of language performance as primarily related to the addressee and the speaker’s socio-interactional goals. Style shifts towards the addressee’s language (convergence) are a resource for showing shared background, affiliation, and perspective (empathy), while divergent style shifts signal the opposite. Prestige will play into contexts of potential convergence, such that speakers will shift towards the code of the higher prestige group if the code is available to them. Moreover, in language contact, “borrowings are very often asymmetric and the socially non-dominant language is affected much more than the dominant one”. We thus hypothesize the following interactional and diachronic outcomes for cross-class interactions:

**Hypothesis 2.4. (Interlocutor, Prestige and Fitness Hypothesis)**

1) Monolingual Hindi speakers who wish to show affiliation with the identity, background and outlook associated with Hinglish (higher prestige, more modern) will synchronically shift towards Hinglish when interacting with Hinglish speakers. Regardless of whether monolingual English is seen equally or more prestigious than Hinglish, this is their only shift-based option, to include minimal English, as they do not command extensive enough English to converge to monolingual English. Diachronically, Hindi monolingual interactions with either of the other two classes will lead to Hinglish conversion, not Hindi/English based conversion, again because of the lack of English access.

2) Hinglish speakers are unable to produce monolingual Hindi synchronically with either Hindi monolinguals or Hindi/English bilinguals producing monolingual Hindi (no shift-based convergence), and because of prestige (fitness), they will not diachronically convert to monolingual Hindi status.

3) Hindi/English bilinguals, when interacting with Hindi monolinguals, may synchronically shift to monolingual Hindi but will not diachronically convert to the Hindi monolingual class because of the prestige their English competence commands.

The Indian urban/rural segregation is also a salient ecological factor mediating Hinglish adoption: the rural NDTV data demonstrates monolingual Hindi, suggesting that despite a lower fitness, monolingual
Figure 1: These figures demonstrate what is traditionally believed about language practice in India, as reported via census data. The figure on the left demonstrates the traditional scenario of movement from the Hindi class to the bilingual class, given the higher prestige of English in India. Here we are using the data from the 1961 census. The figure on the right demonstrates that according to the 2001 census data, the numbers as well as proportions of bilinguals is increasing in the population.
Figure 2: This figure demonstrates what we believe is actually taking place. The population is behaving similar to a three species food chain, such as a buzzard-snake-titmouse food chain, as opposed to a two species competitive system. Therein the buzzard predates on both the snake and the titmouse, the snake only on the titmouse, the titmouse is purely prey. The Hinglish speaker, due to his greater fitness (demonstrated here with thicker arrows) has climbed to the top of this chain. He can convert monolinguals to Hinglish, whereas the bilingual cannot easily convert them to the bilingual class, given they lack the resources for complete English uptake. However, rural Indians who remain in their villages are in a protected refuge, thus protected from the Hinglish speaker.
Hindi can be maintained. We explain this maintenance despite lower fitness by defining the rural setting as a protected refuge from interactions with other classes, which would otherwise provoke conversion given relative fitness. We first define the following,

**Definition 2.5. (prey refuge)** A prey refuge is defined as any strategy used by prey to decrease the risk of predation.

Two types of refuges have been considered [22]. One is where a constant number of prey are protected. This has no influence on the dynamics. The second is where a constant proportion of the prey are protected. This is known to stabilise the dynamics. Also, Indian Hindi-speaking rural migration to urban areas is estimated at about 20% of all Hindi speakers, often seeking work. We consider this migration, and adopt the prey refuge concept to our setting to hypothesise the following:

**Definition 2.6. (Rural Refuge Hypothesis)**
1) Hindi monolingual conversion to other classes is contingent upon migration. The Hinglish speaker, defined as an urban dweller, only has access to maximally influence the 20% of Hindi speakers who are urban immigrants.
2) The remaining 80% of Hindi monolinguals who do not leave their villages are “protected” (a rural refuge) from interactions and fitness-based conversion and thus remain in Hindi monolingual status.

### 2.3 Data collection and Analysis Methods

Three types of spoken language data were examined, Bigg Boss (BB) reality show data (2011-12), 25 rural NDTV interviews (2000-2006) and 30 additional sociolinguistic interviews conducted by the authors (2013).

All English insertions (excluding mixes e.g. no tokens like [relgari], literally railway car, with rel- as a historical borrowing from English rail that has undergone phonetic and morphological indigenization, and indigenized terms like timepass which are not found outside of India) were extracted and tallied for word count based on standard orthographic conventions for the BB and the sociolinguistic interviews, while we also extracted English mixes and indigenized terms to create the NDTV results.

**Bigg Boss:** We examined, from seasons 5 and 6, the 32 contestants who were on the show from the first episode of the season, the 2 seasonal celebrity hosts, and the mysterious Bigg Boss who appears by voice only. Cumulatively, 7,843 English words were used across the two seasons [Season 5: 4,221 English words, Season 6: 3,622 English words]. In Season 5 the 17 participants averaged 248 English words across the 14 episodes [SD = 194, Range = 26-748] while in season 6 the 15 participants averaged 226 English words across the 14 episodes [SD = 147, Range = 22-448]. Only two participants had an extended Hindi-only interaction together: Sidhu’s Confession Room discussion with Bigg Boss illustrated monolingual Hindi, demonstrating that both, as members of the Hindi/English bilingualism class, could produce monolingual Hindi when isolated from Hinglish participants.

**NDTV:** Interviews (conducted 2002-2006) with rural Hindi Belt speakers conducted by NDTV, a New Delhi based broadcast network, were examined following the same data extraction and methods used for the BB data. Examination of 26 episodes resulted in 14 30 minute episodes of viable data from the rural population. Of these, 68:04 minutes of speech by 189 rural Hindi speakers yielded only 80 English words. Moreover, most of these English words were historical loans (e.g. timepass), and/or are arguably stable borrowings, having replaced the Hindi terms, even in rural areas, for official, government and technical terms (e.g. report, police, telephone, photo).

### 2.4 The Mathematical Model

Based on the above hypothesis, we consider an ODE model to represent the dynamics of the Hindi speaking Indian population. We consider three distinct classes of individuals in the population, $M, B$ and $H$. The $H$ variable represents the number of monolingual Hindi speakers. The $B$ variable represents the number of English Hindi bilinguals and the $M$ variable represents the number of Hinglish speakers, or those mixing code.
The model is given by the following system of equations

\[
\begin{align*}
\frac{dM}{dt} &= (m_1M^2 + m_2MB + m_3MH)(1 - (M + B + H)/K) \\
&+ a_1MH + a_2MB - a_3M^2 - a_4M, \\
\frac{dB}{dt} &= (m_4B^2 + m_5BH)(1 - (M + B + H)/K) - a_2MB \\
&+ a_5BH - a_6B - a_7B^2, \\
\frac{dH}{dt} &= (m_6H^2)(1 - (M + B + H)/K) - a_1MH - a_5BH \\
&- a_8H.
\end{align*}
\]

(1)

To the above system we feed reasonable initial conditions.

We look at the first equation in (1), which is for the change in numbers of the Hinglish speakers. This change is due to sexual reproduction, which is then logistically controlled. Furthermore mating between a Hinglish speaker and a monolingual or bilingual only results in a Hinglish speaker, hence we have the term

\[(m_1M^2 + m_2MB + m_3MH)(1 - (M + B + H)/K)\]  (2)

where \(m_1, m_2\) and \(m_3\) are the rates at which a Hinglish speaker might mate with a Hinglish speaker, a bilingual or a monolingual. One can think of \(0 < a_1 < 1\), as precisely a refuge parameter. Thus we see that \((1 - a_1)\), are the constant proportion of the monolingual Hindi speakers that are protected, that is those who do not leave their refuges. While \(a_1\) is the proportion who are subject to interaction with Hinglish speakers (primarily due to their migration to urban areas). The \(a_1MH + a_2MB\) are gain terms due to interaction with monolinguals and bilinguals and the \(-a_3M^2 - a_4M\) are loss terms due to death and intraspecific competition. The equations for the change in bilinguals and monolinguals are derived on similar principles.

3 Results

3.1 Mathematical Results

In this section we describe certain mathematical results pertaining to our proposed model. The system (1) has eight possible steady states. \(E_0 = (0, 0, 0), E_1 = (0, B_1^0, 0), E_2 = (0, B_2^0, 0), E_3 = (0, B_3^0, H_3), E_4 = (M_3^0, 0, H_3), E_5 = (M_3^0, 0, 0), E_6 = (M_6^0, 0, 0), E_7 = (M_6^0, B_6^0, 0), E_8 = (M^*, B^*, H^*).\) The existence of individual states is obtained under certain conditions on the parameters. We omit the details here. We are primarily interested in three states. The bilingual extinction state \(E_4 = (M_3, 0, H_3)\), the coexistence state \(E_8(M^*, B^*, H^*)\), and the Hinglish extinction state \(E_3 = (0, B_2, H_2)\). The others are not realistic from our viewpoint. We show that under certain parameter restrictions \(E_4\) and \(E_8\) are globally asymptotically stable, and that \(E_3\) is unstable.

**Theorem 3.1.** Consider the three species population model described via (1). The bilingual extinction state given by \(E_4 = (M_3, 0, H_3)\) is globally asymptotically stable for all \((M, H)\) such that \(M \leq H\), as long as \(2m_1 + a_1 + 3m_6 < 2a_3\).

**Proof:** There is a fixed point \(E_4 = (M_3, 0, H_3)\) found trivially. In this case system (1) reduces to the following.

\[
\begin{align*}
\frac{dM}{dt} &= (m_1M^2 + m_3MH)(1 - (M + H)/K) + a_1MH \\
&- a_3M^2 - a_4M, \\
\frac{dH}{dt} &= (m_6H^2)(1 - (M + H)/K) - a_1MH - a_8H.
\end{align*}
\]

(3)
We now apply the Dulac criterion [28], with Dulac function $B = 1$ on the reaction terms above.

\[
\nabla \cdot (B f) = \frac{\partial}{\partial M} ((m_1 M^2 + m_3 MH)(1 - (M + H)/K) + a_1 MH \\
- a_3 M^2 - a_4 M) \\
+ \frac{\partial}{\partial H} (m_6 H^2)(1 - (M + H)/K) - a_4 MH - a_6 H) \\
= (2m_1 - 2a_3)M + (a_1 + 3m_6)H - \frac{3m_1}{K} M^2 - \frac{4m_6}{K} H^2 \\
- \frac{2m_1}{K} + \frac{4m_6}{K} M H - a_4 - a_6 \\
\leq (2m_1 - 2a_3 + a_1 + 3m_6)H - \frac{3m_1}{K} M^2 - \frac{4m_6}{K} H^2 \\
- \frac{2m_1}{K} + \frac{4m_6}{K} M H - a_4 - a_6 \\
\leq 0
\]

(4)

This follows via the condition in Theorem 3.1. Thus the Dulac criterion is satisfied, and this excludes the possibility of any closed orbits in $M \leq H \leq K$, as long as $2m_1 + a_1 + 3m_6 \leq 2a_3$. The result follows.

**Theorem 3.2.** Consider the three species population model described via (1). The coexistence steady state given by $E_8 = (M^*, B^*, H^*)$ is globally asymptotically stable, as long as there exist positive constants $0 < k < 1$, $\theta_1$ and $\theta_2$ such that $k(a_2 + a_3) > m_1 a_1, k(a_2 + a_7 \theta_1) > m_2 a_1, k(a_2 \theta_1) > m_2, k(a_2 \theta_2) > m_3, k(a_2 \theta_3) > m_1 a_1,$ and for all $(M, B, H)$ such that $k(a_2 + a_3) M \geq K a_1, k(a_2 + a_7 \theta_1) B \geq (K a_2 - a_6 \theta_1)$ and $k(a_2 \theta_1) H \geq (K a_2 - a_6 \theta_2)$ hold simultaneously.

**Proof:**

Note, $a_1 MH \leq K a_1 H - a_1 H^2$, and the same is true of the state variables $M, B$. Using this we obtain.

\[
\frac{dM}{dt} + \theta_1 \frac{dB}{dt} + \theta_2 \frac{dH}{dt} = (m_1 M^2 + m_2 MB + m_3 MH + \theta_1 m_4 B^2 + \theta_1 m_5 BH + \theta_2 m_6 H^2) \\
(1 - \frac{M + B + H}{K}) + a_1 MH + a_2 MB \\
- a_3 M^2 - a_4 M - a_2 \theta_1 MB \\
+ a_2 \theta_1 BH - a_6 \theta_1 B - a_7 \theta_1 B^2 - a_2 \theta_2 MH - a_5 \theta_2 BH - a_6 \theta_2 H \\
\leq (m_1 M^2 + m_2 MB + m_3 MH + \theta_1 m_4 B^2 + \theta_1 m_5 BH + \theta_2 m_6 H^2) \\
+ Ka_1 M - a_1 M^2 + Ka_2 B - a_2 B^2 - a_3 M^2 - a_4 M - a_2 \theta_1 MB \\
+ a_2 \theta_1 KH - a_2 \theta_1 H^2 - a_6 \theta_1 B - a_7 \theta_1 B^2 - a_1 \theta_2 MH - a_2 \theta_2 BH \\
- a_6 \theta_2 H \\
\leq (a_2 \theta_1 K - a_6 \theta_2) H + (Ka_1 - a_4) M + (Ka_2 - a_6 \theta_1) B \\
+ (m_1 M^2 + m_2 MB + m_3 MH + \theta_1 m_4 B^2 + \theta_1 m_5 BH + \theta_2 m_6 H^2) \\
- (a_2 \theta_1 H^2 + (a_1 + a_3) M^2 + (a_2 + a_7 \theta_1) B^2 + a_2 \theta_1 MB + a_1 \theta_2 MH) \\
- a_2 \theta_2 BH \\
\leq 0
\]

(5)

The last inequality follows via the constraints in Theorem 3.2.
Theorem 3.3. Consider the three species population model described via (1). The Hinglish extinction state given by \( E_3 = (0, B_2, H_2) \) is unstable, as long as \( a_4 < a_2B_2 + a_1H_2 \).

**Proof:** The local stability analysis of \( E_3 \) yields stability under various inequalities holding, one of which is

\[
(m_2B_2 + m_3H_2)(B_2 + H_2) + a_4K > (m_2 + a_2)B_2K + (m_3 + a_1)H_2K
\]  
(6)

Now due to the carrying capacity, we have that \( M_2 + B_2 + H_2 \leq K \), thus we have

\[
(m_2B_2 + m_3H_2)(K) + a_4K > (m_2 + a_2)B_2K + (m_3 + a_1)H_2K
\]  
(7)

dividing above by \( K \) and subtracting \( m_2B_2 + m_3H_2 \) from both sides yields, for (6) to hold we would require

\[ a_4 > a_2B_2 + a_1H_2 \]  
(8)

Violating the above yields instability.

### 3.2 Simulation Results

We next perform numerical simulations to see what the long time dynamics are of the various classes in system (1). In Fig 2 we select the following parameter range: \( a_1 = 0.01; a_3 = .2, a_5 = 0.004, a_2 = 0.06, a_4 = 0.1, a_6 = 0.01, a_7 = 0.01, K = 480000000, m_2 = 0.01 \times 0.25, m_3 = 0.001 \times 0.25, m_1 = 0.25, m_4 = 0.01 \times 0.25, m_5 = 0.001 \times 0.25, m_6 = 0.03 \times 0.25, \)

Also we choose our initial conditions as \( M_0 = 2, B_0 = 4000000, H_0 = 180000000 \), as per the census data for 1961. In Figure 3 all parameters, and initial conditions remain the same except \( a_2 = 2 \). Also in order to accurately capture the actual data, we multiply the LHS of the equations for the Hinglish speaker by \( 10^{-8} \times 28 \), the LHS of the equation for the bilinguals by \( 10^{-8} \times 12 \), and the LHS of the equation for the monolinguals by \( 10^{-8} \times 7 \). Thus in this simulation the uptake parameters for the Hinglish speakers are different from the loss parameters of the monolinguals and bilinguals.

### 3.3 Experimental results for Hinglish

Brief (5-15 minute) sociolinguistic interviews with 24 north Indians from a range of socioeconomic backgrounds were conducted with informants who all 1) acquired Hindi at home, and 2) currently claim fluency in Hindi and English (and occasionally a third language). Informants were explicitly instructed to respond only in Hindi, and Hindi-administered prompts explored informal topics (e.g. childhood, opinions on familial decisions and options in India (e.g. nuclear vs. joint families), pop media (e.g. cricket team preferences)). Importantly, these are not registers or topics which target informants would not acquire in a Hindi-saturated home environment: we were not setting informants up to ‘fail’ by inspecting registers and topics which are not common authentic Hindi socialization settings for this community. Further, speakers have metalinguistic awareness of words as belonging to discrete languages, while they may not recognize the source language for, e.g. a syntactic structure. Given our interest in speakers conscious ability to produce monolingual Hindi speech, we operationalized the research question of monolingual Hindi fluency by examining English insertions at the lexical level.

Ethical consent preceded data collection; interviews were digitally recorded and then orthographically transcribed, and indigenized forms, historical mixes and borrowings were excluded from the English tally. Based on total (Hindi and English) word count [\( \text{Mean} = 995.33, \text{Range} = 172-1963, \text{SD} = 456.55 \)], but importantly, all of these purportedly Hindi/English bilinguals used some English for a Hindi-only task: on average, 18.26 % of their responses were English words (\( \text{Mean} = 171.37, \text{Range} = 3-444, \text{SD} = 125 \)), and English output included individual words as well as larger phrases.

Beyond the 24 respondents from the Hindi Belt who had acquired Hindi at home, we had 6 respondents who self-identified as Hindi bi-/multilingual whose background and practices separated themwhile we report on them separately based on their language acquisition background, their linguistic behavior is also relevant, as they are likely to feed into census reports of India-wide Hindi/English population strength. Of these, three of the four did not acquire Hindi at home, yet in the Hindi-focused task they
Figure 3: We assume that 1961 is a reasonable starting point. We first present a coexistence state. We see that the model captures data well till about 1991, after this the Hinglish speaker rises exponentially and proportionally overtakes the bilingual population, who demonstrate a post-1991 decline in both proportion and raw population. Thus within our scenario incorporating Hinglish speakers, the proportion of bilinguals and monolinguals is likely declining, as people have transitioned from these classes into the Hinglish class.
Figure 4: This is a bilingual extinction state. It shows that bilingual extinction is possible for this choice of parameters by 2021. Here we do not include census data.
illustrated similar code mixing behavior (English rates within target Hindi monolingual production of 12.24-13.39%). The fourth outlier, claiming Hindi fluency, was incapable of producing monolingual Hindi (with output of 93.49% English for the Hindi monolingual task). This suggests that census data, based on self-reports, are even more divergent from actual competencies for speakers who do not report Hindi as a 'home' language, and further underscores our supposition that Indian census data is not representative of actual Hindi/English bi-/multilingual competencies in India. We do not argue against the existence of Hindi/English bilinguals, merely suggesting that their rate is exaggerated in census data, if a Hinglish class is incorporated.

Thus we provide experimental evidence confirming the model assumption of a Hinglish class comprised of a subset of the census population of Hindi/English bilinguals. We confirm that Hinglish practices are uncovered in informants who consider themselves for census purposes Hindi/English bilinguals. Broadly, this task, explicitly seeking and instructing towards monolingual Hindi production, did not uncover a single speaker who could, for a relatively brief conversation, produce monolingual Hindi. Informants were unequal in their English insertions output (see Methods). This confirms our model findings of diachronically changing Hindi competence at the population level, and not to changes in the social contexts within which English insertions are sanctioned, socially contrastive or simply overtly valued, while it confirms, in private discourse, that monolingual Hindi is problematic for the Hinglish class, while they are treated, within Indian censuses, as Hindi/English bilinguals.

4 Discussion and Conclusion

While earlier research has focused on deriving models that match existing census data and only explore canonical ‘languages’, we investigate the emergence of a hybrid Hinglish obligatory codeswitching population based on the Indian context and model future diachronic change scenarios based on this hypothesis. Local ecological conditions related to the urban/rural segregation, and the non-uniform access to English are factors which encourage the emergence and growth of this Hinglish class, while census reports mask this emerging population and practice by not exploring hybridized mixed practices, and by prioritizing self-reports over more direct measures of multilingual repertoires. More broadly, unbalanced competency across multilingual repertoires and limited access to English can lead to obligatory codeswitching in other settings: the ecological factors and outcomes addressed here are not unique to the Indian context, and instead reflect a pattern which better explains both creole development and the emergence and staying power of other hybrid mixed practices, e.g. Singlish and Spanglish.

Specifically, we propose a language competition model which divides the population into three distinct classes: monolingual Hindi speakers, bilingual Hindi/English speakers and Hinglish speakers (who we view as a mutants/innovators). We hypothesize that the three separate classes are interacting similar to a three species food chain, rather than species in direct competition. Within this vein of reasoning, we hypothesize that Hinglish speakers are fitter than monolingual Hindi speakers and bilingual Hindi/English speakers: Hinglish speakers influence both other classes towards Hinglish, and Hinglish speakers, through interactions with other classes, can provoke the spread of Hinglish over either Hindi monolingualism or Hindi/English bilingualism. Fitness here is also predictive, and relates to unequal ability to convert others into one’s class. When Hinglish speakers interact with monolingual Hindi speakers, the Hinglish speaker will out-compete the bilingual, only needing to introduce a few English words to the Hindi speaker in order to provoke the latter’s conversion to Hinglish status. In contrast, bilingual influence on the monolingual towards bilingual status would require the monolingual speaker acquiring full English competency (production of monolingual English), a less likely scenario in the model outcomes. We also introduce the prey refuge framework to protect up to 80% of rural Hindi monolinguals from interactions with the Hinglish speakers. Under these assumptions, the model examines possible outcome scenarios, and broadly predicts that even a very small number of initial Hinglish speakers can quickly take over a large fraction of the population (a defining trait of mutants with higher fitness than the average population). Our hypothesis that the Hinglish speaker has actually climbed to the top of this chain, due to his fitness, is reflected in the model outcomes. The model proposed here supports states of coexistence as well as bilingual extinction. Significantly, the latter is completely contradictory to current census data.

Modeling Hindi/English diachronic language shift in India based on census data fails to capture the reality of the historic, contemporary, and hypothesized future possible scenarios. Contemporary Hindi-only reality show data from urban speakers uncovers almost uniformly mixed Hinglish practices. This finding is replicated in our independent sociolinguistic interviews: urban informants who claim Hindi/English
bilingualism are unable, in a focused Hindi-only conversational task, to produce monolingual Hindi speech, and instead demonstrate Hinglish code switching. These data directly challenge census reports by reframing these speakers who claim bilingual proficiency as unable to produce monolingual Hindi. Further it directly supports our hypothesis for the presence of a Hinglish population (monolingual Hindi is unavailable to these speakers), who our model predicts are increasing in population strength diachronically. Our hypothesis for a rural stronghold for Hindi monolingualism is confirmed with rural NDTV interview data, which does not demonstrate Hinglish, instead reflecting fluent monolingual Hindi practices.

We have presented models here which are distinct from most language competition modeling goals in that they do not reflect nor attempt to match census data, and are not focused on refining parameters per historical records. Instead, we hypothesize, supported by three types of data, the emergence of a Hinglish-speaking population, and illustrate the range of possible outcomes given this class of speakers. In such, our research better reflects the reality of the Indian context and sociolinguistic findings from other (post)colonial and language contact scenarios, and makes predictions for future Indian fluencies in these populations. We do not focus on parameter refinements, and instead investigate how this distinct three species food chain has formed in the Hindi-speaking Indian population, and where it may lead. However, it is worth analysing the mathematical results, and their connections to realistic diachronic scenarios. Theorem 3.2 predicts a coexistence state is possible, within these three groups. This is probably the most realistic outcome. Although a bilingual extinction state is predicted by Theorem 3.1, this is probably unrealistic as there will always be bilinguals who essentially will be able to produce monolingual Hindi, because of say professions that require it. Theorem 3.3 is extremely interesting. It essentially says that for $a_1, a_2$ large enough, an introduction of even a very small number of Hinglish speakers can take over a large faction of the population. Since $a_1, a_2$ are the uptake parameters of monolinguals and bilinguals into the Hinglish class, we are essentially claiming that Hinglish speakers whose fitness is above a certain threshold will always establish themselves in the population.

Our work also differs from earlier language death/extinction research: we do not claim a future extinction of Hindi as a likely outcome, but, based on the models, suggest that Hindi will survive precisely because the rural context functions as a refuge. Meanwhile, as future work we aim to investigate spatially explicit models to predict densities and sites of contact and conversion to Hinglish in urban areas, and Hindi maintenance in protected rural villages. We also aim to explore modeling refuges from a couple of different ways. We might also try to model various other socio-economic factors, that might affect language uptake such as education level, modeled perhaps via linear migration terms between classes. The range of future outcomes based on our models explains the recent Hinglish explosion in urban areas, the Hindi monolingual retention in rural areas, as well as Hindi/English bilingual attrition, all of which are reflected in the linguistic data, while the model builds on these definitions to predict possible future language shift-based outcomes in India.

References

[1] Abrams DM, Strogatz SH (2003) Linguistics: Modelling the dynamics of language death. Nature 424(6951):900.
[2] Zhang M, Gong T (2013) Principles of parametric estimation in modeling language competition. Proc Natl Acad Sci USA 110:9698-9703.
[3] Gardner-Chloros P (2009) Code-Switching (Cambridge University Press, Cambridge).
[4] Sarhimaa A (1999) Syntactic transfer, contact-induced change, and the evolution of bilingual mixed codes: Focus on Karelian-Russian language alternation (Finnish Literature Society, Helsinki).
[5] Chand V (2011) Elite positionings toward Hindi: Language policies, political stances and language competence in India. Journal of Sociolinguistics 15(1):1-30.
[6] Relph MK (2011) Make room for Hinglish. in New York Times (Hearst, New York City).
[7] Agha A (2003) The social life of cultural value. Language & Communication 23:231-273.
[8] John BK (2007) Entry from backside only: Hazaar fundas of Indian-English (Penguin Books, New Delhi).
[9] Kothari R & Snell R eds (2011) Chutnefying English: The phenomenon of Hinglish (Penguin Books India, New Delhi).
[10] Chauhan A (2013) Those pricey Thakur girls (HarperCollins Publishers India, Noida).
[11] Si A (2010) A diachronic investigation of Hindi-English code-switching, using Bollywood film scripts. International Journal of Bilingualism 15(4):388-407.
[12] Shankar S (2004) Reel to real: Desi teens’ linguistic engagement with Bollywood. Pragmatics 14(2/3):317-335.
[13] Bhandari (2010) Asian Development Bank. Special chapter: The rise of Asia’s middle class. In Key Indicators for Asia and the Pacific 2010.
[14] Auer P ed (1998) Code-switching in conversation: Language, interaction and identity (Routledge, London).
[15] Auer P (1999) From code-switching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. International Journal of Bilingualism 3(4):309-332.
[16] Muysken P (2013) Language contact outcomes as the result of bilingual optimization strategies. Bilingualism: Language and cognition 16(4):709-730.
[17] Myers-Scotton C (1993) Common and uncommon ground: Social and structural factors in codeswitching. Language in Society 22(4):475-503.
[18] Trudgill P (1972) Sex, covert prestige and linguistic change in the urban British English of Norwich. Language in Society 1(2):179-195.
[19] Isern F and Fort J (2014) Language and cognition 16(4):740-742. Language extinction and linguistic fronts. J.R.Soc. Interface 11: 20140028.
[20] Loreto V, Mukherjee A and Tria F (2012) On the origin of the hierarchy of color names. Proc Natl Acad Sci USA. 109 (18):6819-6824.
[21] Pucci L, Gravino P, Servedio V (2014) Modeling the Emergence of a New Language: Naming Game with Hybridization. Self organising systems, Lecture Notes in Computer Science. 8221: 78-89.
[22] Kar TK (2005) Stability analysis of a predator-prey model incorporating a prey refuge. Communications in Nonlinear Science & Numerical Simulation 10:681-691.
[23] Burger R (2000) The mathematical theory of selection, recombination and mutation (Chichester, Wiley).
[24] Paradis M (2004) A neurolinguistic theory of bilingualism (John Benjamins, Amsterdam).
[25] Schmid MS (2007) The role of L1 use for L1 attrition. Language attrition: Theoretical perspectives, eds Kpke B, Schmid MS, Keijzer M, & Dostert S (John Benjamins, Amsterdam), pp 135-153.
[26] Maschler Y (2000) Toward fused lects: Discourse markers in Hebrew English bilingual conversation twelve years later. International Journal of Bilingualism 4(4):529-561.
[27] Matras Y (2000) Fusion and the cognitive basis for bilingual discourse markers. International Journal of Bilingualism 4(4):505-528.
[28] Strogatz S, (1994) Nonlinear dynamics and Chaos: With applications to physics, biology and chemistry, Cambridge, Mass.
[29] Potowski K (2013) Language maintenance and shift. The Oxford handbook of sociolinguistics, eds Bayley R, Cameron R, & Lucas C (Oxford University Press, Oxford), pp 321-339.
[30] Bell A (2001) Back in style: reworking audience design. Style and sociolinguistic variation, eds Eckert P & Rickford JR (Cambridge University Press, Cambridge), pp 139-169.
[31] Bell A (1984) Language style as audience design. Language in Society 13:145-204.
[32] Nowak MA,Komarova NL,Niyogi P (2002) Computational and evolutionary aspects of language. Nature 417(6889):611617
[33] Lieberman E, Michel JB, Jackson J, Tang T, Nowak M. (2007) Quantifying the evolutionary dynamics of language. Nature 449(7163): 713-716
[34] Kandler A (2009) Demography and language competition. Hum Biol 81(2-3):181210.
[35] Mira J, Seoane LF, Nieto JJ (2011) The importance of interlinguistic similarity and stable bilingualism when two languages compete. New J Phys 13:033007.
[36] Baggs I, Freedman HI (1990) A mathematical model for the dynamics of interactions between a unilingual and a bilingual population: Persistence versus extinction. J Math Sociol 16(1):5175.
[37] Mira J, Paredes A (2005) Interlinguistic similarity and language death dynamics. Europhys Lett 69(6):10311034.

[38] Minett JM, Wang WS-Y (2008) Modeling endangered languages: The effects of bilingualism and social structure. Lingua 118(1):1945.

[39] Patriarca M, Heinsalu E (2009) Influence of geography on language competition. Physica A 388(2-3):174186.

[40] Mufwene SS (2001) The Ecology of Language Evolution.

[41] Mufwene SS (2008) Language Evolution: Contact, Competition and Change, Continuum International Publishing Group.

[42] Wang WS-Y, Minett JW (2005) The invasion of language: Emergence, change and death. Trends Ecol Evol 20(5):263269.

[43] Schulze C, Stauffer D (2007) Competition of languages in the presence of a barrier. Physica A 379(2):661664.

[44] Schreier D and Hundt M (2013) English as a contact language, Cambridge, UK, Cambridge university press.