Accepted Manuscript

What is India speaking? Exploring the “Hinglish” invasion

Rana D. Parshad, Suman Bhowmick, Vineeta Chand, Nitu Kumari, Neha Sinha

PII: S0378-4371(16)00023-6
DOI: http://dx.doi.org/10.1016/j.physa.2016.01.015
Reference: PHYSA 16776

To appear in: *Physica A*

Received date: 26 August 2015
Revised date: 21 December 2015

Please cite this article as: R.D. Parshad, S. Bhowmick, V. Chand, N. Kumari, N. Sinha, What is India speaking? Exploring the “Hinglish” invasion, *Physica A* (2016), http://dx.doi.org/10.1016/j.physa.2016.01.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

*Highlights (for review)

Highlights:

1) Field evidence is collected to show a class of ``Hinglish" speakers in India.
2) A three species language competition model is derived including Hinglish speakers.
3) The model shows a few Hinglish speakers can invade the population.
4) The model shows Monolinguals, Bilinguals and Hinglish speakers can also coexist.
5) Turing instability can occur in the spatially explicit three species model.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT
*Manuscript
Click here to view linked References

WHAT IS INDIA SPEAKING? EXPLORING THE "HINGLISH" INVASION

Rana D. Parshad and Suman Bhowmick
Department of Mathematics,
Clarkson University,
Potsdam, New York 13699, USA.
Vineeta Chand
Centre for Language Development Across the Lifespan,
Department of Language and Linguistics,
University of Essex,
Colchester CO4 3SQ, UK.
Nitu Kumari
School of Basic Sciences,
Indian Institute of Technology Mandi,
Mandi, Himachal Pradesh 175 001, India.
Neha Sinha
Center for Molecular and Behavioral Neuroscience,
Rutgers University,
197 University Avenue, Newark, New Jersey 07102, USA.

Abstract. Language competition models help understand language shift dynamics, and have effectively captured how English has outcompeted various local languages, such as Scottish Gaelic in Scotland, and Mandarin in Singapore. India, with a 125 million English speakers boasts the second largest number of English speakers in the world, after the United States. The 1961-2001 Indian censuses report a sharp increase in Hindi/English Bilinguals, suggesting that English is on the rise in India. To the contrary, we claim supported by field evidence, that these statistics are inaccurate, ignoring an emerging class who do not have full bilingual competence and switch between Hindi and English, communicating via a code popularly known as "Hinglish". Since current language competition models occlude hybrid practices and detailed local ecological factors, they are inappropriate to capture the current language dynamics in India. Expanding predator-prey and sociolinguistic theories, we draw on local Indian ecological factors to develop a novel three-species model of interaction between Monolingual Hindi speakers, Hindi/English Bilinguals and Hinglish speakers, and explore the long time dynamics it predicts. The model also exhibits Turing instability, which is the first pattern formation result in language dynamics. These results challenge traditional assumptions of English encroachment in India. More broadly, the three-species model introduced here is a first step towards modeling the dynamics of hybrid language scenarios in other settings across the world.

1. Introduction

1.1. **Motivation.** Language competition is a central driver of diachronic language change and shift [1]. Methods and ideas from statistical physics, evolutionary biology and game theory have been extremely effective in analysing the evolution of

Key words and phrases. Language competition, Mixed languages, Three-species food chain, stability, Turing instability.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

languages, see [2, 3, 4, 5, 6] and the references within. Modeling of population-based language shift (generally towards English) has also been the subject of intense investigation [6, 7, 8, 9, 10, 11, 12]. The statistical physics approach has recently been in the forefront due to large amounts of digital information or "big data", that is now readily available to be analysed by such methods [13]. e.g., the google books corpus, Twitter and Facebook. The availability of such data allows one to analyze word frequencies across millions of words in various languages, and time spans, in an attempt to fit various statistical laws to them [14]. This enables us to rigorously infer how and in what capacity a language is changing. A language could be changing due to competition at the lexical level, similar to intraspecies competition [15], where individuals of a certain species, or genes in a gene pool, compete amongst each other. For example, it has been suggested that at the lexical level, competition between words proceeds similarly to gene selection [15, 16].

Languages also compete with *other* languages [6, 17, 18, 19, 9, 20]. Thus another approach has been to consider language competition, similar to interspecies competition, such as two predators from different species, competing for a common prey [21]. In this framework, many models have traditionally considered two competing languages, each spoken by a Monolingual group, with the possibility of a Bilingual group [6, 26, 22, 23, 24, 25, 29]. These models have been successful in replicating and predicting the dynamics of language competition and the resultant decline of one language in specific contexts: English has out-competed Scottish Gaelic in Scotland, Welsh in Wales and Mandarin in Singapore [8, 27]. In order to preserve endangered languages, there has been very interesting recent work, using a control theoretic approach [6, 26]. Incorporating the role of population size and geographic distance on language competition [17, 28, 29] demonstrates that sociolinguistic factors strengthen the predictive power of models: language shift is a multifaceted phenomenon.

However, language competition modeling is strongly oriented towards historically completed changes wherein "good" models accurately match historical data, rather than models oriented towards contemporary settings where the final outcomes are unknown or undocumented. Furthermore, although there are some current models that do consider local ecological factors [17, 18, 19], there are *almost none* which address the types of challenges Indian language competition presents. For language dynamic models to stay relevant, they must be able to handle real-world changing scenarios, and to adapt, when current modeling frameworks mis-predict. The primary objectives of this manuscript are to:

- Provide a background for the modern Indian context, emphasizing *why* current language dynamics models are innapropriate herein,
- Introduce new evidence/data which motivate the need for a mixed class in modeling
- Introduce a new model that includes this class of mixed language speakers,
- Focus on three species type predator-prey dynamics,
- Model the urban/rural divide, operationalized via a refuge setting,
- Model intense competition for resources such as jobs among Bilinguals.
- Explore various diachronic scenarios based on the above.

1.2. **Background.** Modern day India is extremely heterogeneous, with a population of 1.13 billion encompassing six language families, language policies which dictate schooling in up to three languages, which are themselves embedded within

ACCEPTED MANUSCRIPT

a traditional caste system and emerging class system linking to language competencies and access. This paper thus focuses on a narrower setting within India, the Hindi Belt, for several reasons. Hindi, an Indo-European language, arguably holds more social power than any other indigenous Indian language as the planned replacement for English in early post-colonial governance. While this was not realized, Hindi continues to hold considerable sway as an imagined lingua franca [30], and it continues to be both numerically and proportionally the largest indigenous language community: 422 million (41% of the population [1]) [31] . The Hindi Belt is a swath across north/central India encompassing the capital, New Delhi, and including a majority of the Indian population, in which Hindi has a stronghold [32, 33]. More broadly, a large proportion of L1 Hindi speakers all over India, are from these states.

1.3. **Current language models and local ecological factors.** The Hindi-speaking population consists of Monolingual Hindi speakers and Hindi/English Bilinguals, but, as Indian census data for 40+ years argues [31], a diachronically almost non-existent Monolingual English speaking population. Thus a current language competition model [22, 8, 9] applied to the Hindi Belt would be

$$\begin{aligned} \frac{dB}{dt} &= \gamma B^{\alpha} M^{\beta} \\ \frac{dM}{dt} &= -\gamma B^{\alpha} M^{\beta} \end{aligned} \tag{1}$$

where B, M stand for the numbers of Hindi-English Bilinguals and Hindi Monolinguals at a given time t, and where $\alpha \geq 1$ is a parameter that measures the difficulty of English to attract speakers, and $\beta \geq 1$ is a parameter that measures the resistance of Hindi to lose speakers [22]. $\alpha = \beta = 1$ is a simple choice [9]. We would provide suitable initial conditions drawn from census data, $B(0) = B_0, M(0) = M_0$. Note, this model has only one stable fixed point $(B^*, 0)$ $(B^* = B_0 + M_0)$, and would thus predict *the extinction of Monolingual Hindi speakers.* This is because, according to current modeling hypotheses, speakers will shift from using Monolingual Hindi to learn English and thus become Bilingual, given the higher status of English in India, see Fig. 2 A. However, this scenario is *extremely unlikely*, given that nationally, 26% of rural Indians and 13% of urban Indians reside under the poverty line [34] with almost no access to English, while the Hindi Belt arguably has experienced even less social progress, and hence would be driving these numbers, with comparatively *higher* poverty rates (and less English access) [35], see Table 1. Collectively, these will prevent complete English acquisition within lower class Hindi Belt communities, as language acquisition requires rich and sustained access to linguistic material which is unavailable to these (largely) lower class populations. Thus this socioeconomic factor —accounting for socioeconomic disparity —is not reflected in (1) —making (1) an inaccurate model to predict language shift in the Hindi Belt.

[1]The Indian Census is admittedly problematic: at times, numerous other codes have been treated as dialects of Hindi, rather than as distinct languages, a political move arguably accomplished to bolster the purported strength of Hindi. Meanwhile, 2011 Census findings related to speaker strength at both the national and state level remain unavailable at the time of submission (Summer 2015).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

	% of rural population BPL	% of urban population BPL	
India as a whole	25.7	13.7	
Bihar	34.06	31.23	
Jharkhand	40.84	24.83	
Uttar Pradesh	30.4	26.06	
Madhya Pradesh	35.74	21	

TABLE 1. For India as a whole and a sample of four Hindi Belt states, the poverty rates (% below the poverty line: BPL) are higher in rural settings. Note, these Hindi Belt states have higher poverty rates than the national average.

Poverty and English access also link to urbanization: unlike other developed nations, India's very low urban ratio (projected to remain under 45% in 2030 [36]) provides a rural setting which blocks, or at least hinders English access, acquisition and ultimate growth [37]. This urban/rural divide of the Indian population, in conjunction with limited rural migration into urban centres [36], is an important local ecological factor which we will draw on in modeling, as a refuge setting that restricts English development. While this is not the first work to introduce a refuge setting [17, 38], the Indian setting requires further considerations, discussed next. However, before addressing these note that this factor —accounting for a strong urban/rural division —is not reflected in (1) —making (1) an inaccurate model to predict language shift in the Hindi Belt.

Language competition models do not typically examine factors which would provoke the demise of English in the local setting. However, making sense of Indian census data requires considering this possibility, and ecological factors which would support limited English development —or actual loss. We draw on national age demographics: India is a very young nation [36], wherein job competition —already fierce—is likely to have a larger effect in the near future, as youth reach productive working age. We hypothesize that jobs, as a limited resource, may link to language competencies in the following ways: speakers who lack English competency will be at a disadvantage in job competition, which would perpetuate any poverty, rural living, and limited education; speakers who acquire English competency for the job market, but fail to secure a position which uses this language skill, will see a diachronic loss in their English competency, and will be less likely to pass English on to their children. Collectively, these encapsulate our interpretation of how there may be a loss term associated with the Bilingual Hindi/English class, and they are supported by development projections, which argue that increased economic growth puts less skilled workers out of work, thereby increasing the poverty [36]. Since there is clearly no loss term for the Bilingual in (1), it again is an inaccurate model to predict language shift in the Hindi Belt.

1.4. **Mixed language practice.** Another practical and theoretically under explored issue which motivates the current modeling effort involves the *type* of language practices found in the Hindi Belt. While traditionally dismissed as rare and unnatural, contemporary language evolution research argues that language hybridization is a common process [1, 39]. Supporting this, "profound and widespread"

ACCEPTED MANUSCRIPT

horizontal language contact is historically attested in India [40]. Contemporary language hybridization is also visible in unmarked code-switching (alternating between two or more languages in a single conversation) between English and regional Indian vernaculars. In the Hindi Belt, Hinglish is the most prominent form of hybrid communication. "Hinglish" is a colloquial umbrella-term [41, 42, 43] spanning isolated borrowings (1) indigenized Indian English forms (2) within otherwise Monolingual Hindi or English, to rich code-switching practices unintelligible to Monolingual Hindi or English speakers (3).

1. *Jahan unka* transfer *hua tha* (Where he was transferred to.)
2. *Rahul*, tension *mein hain* (Rahul is tense.)
3. Don't think *wo* capable *hai* based on again *jo tum padh sakte ho, wo kaisa hai* real life *mein kabhi nahi pata chalega* (I don't think he was capable, based on what you can read about him/that, you would never know how he is in real life.)

We refrain from locating Hinglish as a creole or as an established fused lect [1, 44, 45, 46]. It thus contrasts to Nagamese, an Assamese lexified pidgin-turned-creole found in Nagaland and Indo-Portuguese, a creole found in Southwestern coastal India and Sri Lanka [47]. Hinglish commands national and international recognition, is used across a range of genres and mediums, e.g. informal discourse, popular handbooks, fiction novels, advertising, TV shows and films [41, 42, 43, 48, 49], and demonstrates nascent codification efforts [50]. Ideologically, it indexes a middle ground between upper and lower classes, values and broader dispositions, as a modern but locally grounded way of presenting oneself, while this positive value is challenged by purists [41, 30]. Some urban elite youth, canonical Bilinguals per census reports, *even consider Hinglish their mother tongue* and claim only limited Monolingual Hindi competence [30].

Hinglish, as the most socially prominent of emergent hybrid codes in the Hindi Belt setting, is potentially affecting the most widely spoken indigenous language, Hindi, yet it is not accounted for within census reports. Census reports may instead be providing a false perspective of contemporary language competition and linguistic vitality in the Hindi Belt. Thus a realistic language competition model for the Hindi Belt *must* take this class of Hinglish speakers into account. However, this is not reflected in (1), exposing another way (1) will mis-predict Hindi Belt language dynamics.

As a first step in building an accurate language competition model for the Hindi Belt, we need to consider three interacting classes of speakers, isolating from the traditional, census defined, Hindi/English Bilingual population a sub-population who do not speak Monolingual Hindi and/or English, and who should be reclassified as Hinglish speakers. Thus we define

- Monolingual Hindi class: Can produce Monolingual Hindi, English restricted to limited inclusion of established indigenizations and loanwords.
- Hindi/English Bilingual class: Can produce Monolingual Hindi and Monolingual English.
- Hinglish class: An urban sub-population who cannot produce Monolingual Hindi, and/or Monolingual English, only Hinglish.

We beleive these classes are behaving similar to a three species food chain [51], see Fig. 2.

ACCEPTED MANUSCRIPT

2. A two species ODE model

As pointed out earlier the problems with applying (1) to the Indian scenario are that

- The only stable fixed point is $(B^*, 0)$, or extinction of Monolingual Hindi speakers, which is unrealistic.
- The lack of considering local ecological factors, such as Bilingual competition and the existence of a Hinglish class of speakers.

For the time being we will not consider Hinglish speakers. Assume that there are only Hindi-English Bilinguals and Hindi Monolinguals in the Hindi belt [52]. What might a reasonable model for the language dynamics be, given that (1) is not apt? The critical problem in (1) is that there is no "loss" term in the equation for the Bilingual. They only gain speakers as the Monolinguals loose speakers, resulting in Monolingual extinction. However, competition in India is intense, essentially due to a huge population, and only limited resources. Jobs in almost all prestigious sectors require English proficiency, hence Bilingual status and competition solely within this class. This needs to be reflected in a reasonable model.

2.1. Derivation, equilibrium and stability.

We propose the following model to describe the language dynamics between Hindi-English Bilinguals and Hindi Monolinguals in the Hindi Belt region of India,

$$\begin{aligned} \frac{dB}{dt} &= B\left[\frac{\varepsilon M}{B+M} - \frac{d_2 B}{B+M}\right] \\ \frac{dM}{dt} &= M\left[\frac{-\varepsilon B}{B+M} + 1 - \frac{M}{K}\right] \end{aligned} \tag{2}$$

We assume that the gain or loss of speakers to a certain class occurs by direct interaction between the speakers of that class with the proportion of speakers of the other classes, in the whole population. We are modeling change within speakers, though it's only allowed for the Bilingual and Monolingual classes. The Bilingual class gains speakers from the Monolingual class according to the direct product $B \times \left(\frac{M}{B+M}\right)$, the rate of gain is measured by the parameter ε. Here we model this horizontal transmission term $B \times \left(\frac{M}{B+M}\right)$, according to the forms derived in [53]. Furthermore, we model the extreme competition among Bilinguals via a loss term given by $B \times \left(\frac{B}{B+M}\right)$ with the rate of loss measured by the parameter d_2. Note this competition term depends on the number of Bilinguals in the population, thus if there were very few Bilinguals, $\left(\frac{B}{B+M}\right)$ would be small, hence there would be very little competition, however this is not so due to the large number of Bilinguals suggested by census reports. The Monolingual class only looses speakers to the Bilingual class according to direct product $-B \times \left(\frac{M}{B+M}\right)$, but the Monolinguals are also logistically controlled.

The model (2) has steady states $(0, 0)$, $(0, K)$ and (B^*, M^*). $(0, K)$ is unstable. The non trivial equilibrium point is given by $(B^*, M^*) = (\alpha M^*, M^*)$, where $\alpha = \frac{\varepsilon}{d_2}$ and $M^* = K[1 - \frac{\varepsilon^2}{d_2+\varepsilon}]$. For positivity/feasibility of the equilibrium point we require $1 - \frac{\varepsilon^2}{d_2+\varepsilon} > 0$.

ACCEPTED MANUSCRIPT

Further we find the equilirium point (B^*, M^*) is locally asymptotically stable if $d_2 < \varepsilon^2$. This is seen by applying the Routh Hurwitz criterion for stability [54].

The $Trace(J^*) = -\left[\frac{d_2 B^* M^*}{(B^*+M^*)^2} + \frac{M^*}{K}\right] < 0$. Thus for stability we require the determinant to be positive $Det(J^*) = \frac{(d_2+\varepsilon)\alpha}{\alpha+1}\left[\frac{M}{K} + (\frac{\alpha\varepsilon}{\alpha+1})(\frac{1}{(\alpha+1)^2} - 1)\right]$, and this will be positive if $\frac{M^*}{K} + \frac{\varepsilon\alpha}{\alpha+1}(\frac{1}{(1+\alpha)^2} - 1) > 0$, or equivalently if $1 - \frac{\varepsilon^2 d_2^2}{(\varepsilon+d_2)^3} > 0$. Using this in conjunction with the feasibility conditions, gives the requisite condition.

2.2. **Model fitting to census data.** To validate the robustness of our ODE model, we try and fit it to the actual Indian census data. To this end we use a nonlinear optimization routine. We find that parameters $d_2 = 1.81, \varepsilon = 1.092$ best fit the Indian census data given in table 2. The total error here is 0.3784. The best fit is shown in Figure 1.

Note, we do not agree that the language dynamics in India are evolving according to (2), instead rather like (3) – discussed next – due to the presence of Hinglish speakers. Fitting (2) to real census data is important for model building, as well as to have certain base parameters around which we can vary parameters and run simulations to explore various diachronic scenarios. Moreover, before introducing a Hinglish class, it is necessary to prescribe the correct dynamics between Monolingual Hindi speakers and Bilinguals, taking local ecological factors in India into account.

Year	Bilingual Population	Monolingual Population
1961	3800000	187000000
1971	8500000	237000000
1981	17000000	300000000
1991	29000000	381000000
2001	35000000	473000000

TABLE 2. Bilingual and Monolingual population numbers from the Indian census data 1961-2001. The 2011 census data is still not available publicly.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

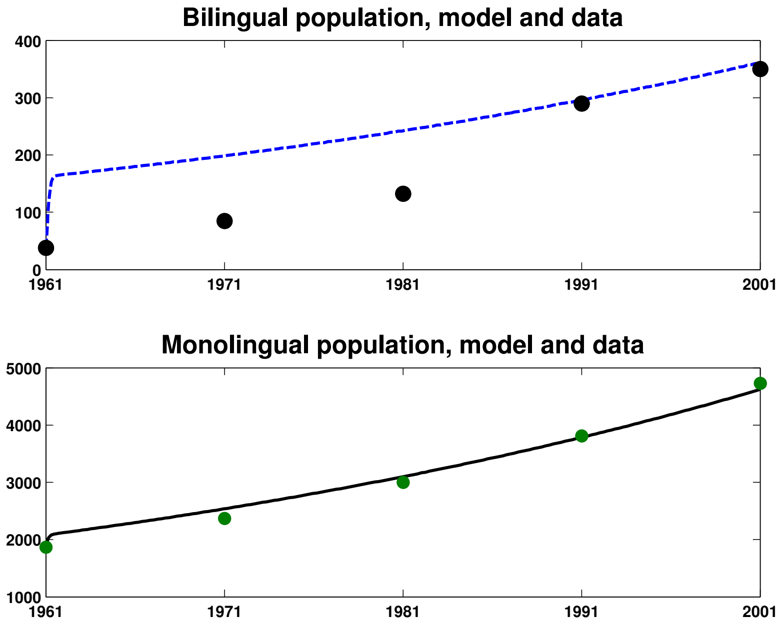

FIGURE 1. Simulations from our model (2) that best fit the Indian census data. The dots are the actual data from the census, while the solid lines are the output from our model (2). The scale in the top panel y-axis is 1 unit = 1 million. While in the bottom panel y-axis it is 1 unit = 100,00. Note again that the key novelty in (2) is that competition among bilinguals is introduced. While the model fits the monolingual class census data extremely well, it does not fit the bilingual class census data as well.

3. Sociolinguistic evidence for "Hinglish"

We next present 3 types of field data that concretely point towards the existence of a Hinglish class, and demonstrate a rural refuge setting. We use this data to hypothesise certain interactional trends, which we then use to build a mathematical model that incorporates both this Hinglish class and a rural refuge setting, and explore its various diachronic outcome scenarios. We will demonstrate the presence of a Hinglish class and investigate their interactional trends with authentic data from the Indian media as well as independent sociolinguistic interviews that we conducted.

3.1. **Field Evidence collection methods.** Three types of spoken language data were examined, 2 complete seasons of Bigg Boss (BB) reality show data (2011-12), 26 rural NDTV interviews (2000-2006), and 24 sociolinguistic interviews conducted by the authors (2013). Given that lexical access is first and most severely affected in language attrition [55], we focused on the lexical level as a benchmark for separating hybrid code-mixing from Monolingual practices (i.e. un-established English

ACCEPTED MANUSCRIPT

lexical insertions). For BB and sociolinguistic interview data Hindi and English words were tallied following standard orthographic conventions, excluding established forms, including borrowings, e.g. *report, police, telephone, photo*; attested mixes e.g. [relgari], *railway car*, with *rel-* a historical borrowing from English rail that has undergone phonetic and morphological indigenization to be combined with the Hindi term gari or car; and indigenized Indian English terms, e.g. *timepass*, not attested outside of India. The NDTV results included established borrowings, mixes, and indigenizations in order to explore the degree of English encroachment into purportedly Monolingual Hindi settings.

Bigg Boss: We examined, from seasons 5 and 6, the 32 contestants who were on the show from the first episode of the season, 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 contestants averaged 248 English words each across the 14 episodes [SD = 194, Range = 26-748] while in season 6 the 15 contestants averaged 226 English words across the 14 episodes [SD = 147, Range = 22-448].

Sociolinguistic Interviews: We conducted brief (5-15 minute) sociolinguistic interviews with 24 north Indians from a range of socioeconomic backgrounds who all 1) acquired Hindi at home, and 2) report Hindi/English Bilingualism per census criteria. Informants were explicitly instructed to respond only in Hindi, via informal prompts (e.g. childhood, opinions on familial decisions and options in India, sports and pop media) targeting participants who self-assess as Hindi-English Bilinguals per census criteria. Ethical consent preceded data collection; interviews were digitally recorded and then orthographically transcribed. Based on total (Hindi and English) word count [Mean = 995.33, Range = 172-1963, SD = 456.55], all of these purportedly Hindi/English Bilinguals used some English for this focused Hindi-only task: on average, English comprised 18.26 % of each response (Mean = 171.37, Range = 3-444, SD = 125).

NDTV: 26 NDTV-conducted interviews (2002-2006) were examined, resulting in 14 episodes of viable data from the rural Hindi Belt, within which, the 68:04 minutes of speech by 189 rural Hindi villagers yielded only 80 English words. Moreover, all of these were historical loans and/or established borrowings.

3.2. **Interpretation of Field Evidence.** The popular Indian reality TV show Bigg Boss [BB] is based along the same lines as MTV Real World in the US, and Big Brother in the UK. BB is ostensibly Hindi-only, with repercussions for English transgressions, and thus represents authentic language/speech within a context encouraging and overtly valuing Monolingual Hindi. Given the rules of the game show and background demographics, all contestants should have been able to speak in Monolingual Hindi. However, a tally of English lexicon *revealed zero contestants who were able to consistently produce Monolingual Hindi*, while Bigg Boss, who arguably has the only scripted role as taskmaster, produced Monolingual Hindi. The following interactional patterns were uncovered, which are useful for modeling.

1. Hinglish speakers, who are defined as such based on their language, did not deviate from mixing, whether speaking to a Bilingual (regardless of whether their interlocutor is code-switching or not) or a Monolingual.

2. Likewise, Bilinguals and Monolinguals code-switched when speaking with Hinglish speakers, and did not produce Monolingual Hindi. The Monolingual Hindi speakers demonstrated increased single word English insertions when speaking with Hinglish

ACCEPTED MANUSCRIPT

speakers, compared to when they were speaking with a Bilingual producing Monolingual Hindi (such as with Bigg Boss) or amongst themselves. The Bilingual did the same, moving toward more mixed practices when interacting with Hinglish speakers.
3. Bilinguals and Monolinguals were able to produce Monolingual Hindi when speaking with other Bilinguals and Monolinguals: their code-switching was conditioned.

To confirm the above practices we conducted brief sociolinguistic interviews with 24 census reporting Hindi/English Bilinguals from the Hindi Belt, from a range of socioeconomic backgrounds, who were *explicitly instructed to respond only in Hindi*. We did not uncover a single speaker who produced Monolingual Hindi. Instead all our subjects exhibited Hinglish: on average, 18.26% of their speech was in English. This data confirms that a subset of self-reporting Hindi/English Bilinguals *are actually Hinglish speakers*.

Since these BB participants and subjects for the sociolinguistic interviews were all from urban settings, and access to English is unequal across rural and urban India, we next explored contemporary televised rural data from NDTV, a well known news channel in India. In these, Bilingual NDTV correspondents travel to rural Hindi Belt villages (some extremely segregated) to speak with lower class and minimally educated villagers. The 189 villagers interviewed produced Monolingual Hindi. English was only uncovered in single word historical indigenizations and/or technology/government terms (e.g. police, telephone) which are established loans into Hindi for all speakers. This field data demonstrates that rural Hindi Belt villagers meet our Hindi Monolingual definition. The field evidence for three classes and specific interactional patterns, in conjunction with predator-prey, sociolinguistic and evolutionary linguistic theories, were used to develop principles for modeling Hindi Belt language dynamics, discussed next.

3.3. **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 [56]. Exemplar-based theories of cognition and language change [1, 57, 58] offer a means of understanding the emergence of mutations (innovations) within language shift and have been incorporated within recent language dynamics modeling of two-class competition [59, 60, 61]. Hybrid languages, while never incorporated into population-based shift models, are suggested to have fitness distinct from traditional languages [20]. It has also been suggested that at the lexical level, competition between words proceeds similarly to gene selection [15, 16]. Interactionally, the audience design model of speech accommodation [62] 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 also plays into contexts of potential convergence, such that speakers shift towards the code of the higher prestige group if that code is available to them. We draw on these to hypothesize the following cross-class interactional and diachronic outcomes.

I) **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

ACCEPTED MANUSCRIPT

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, limited English insertions, given limited English access. Diachronically, Hindi Monolinguals' interactions with either of the other two classes will lead to Hinglish conversion, over Bilingual conversion, again because of limited English access.
2) Hinglish speakers, because they lack full proficiency in Monolingual Hindi and English, will synchronically demonstrate no shift-based convergence to Monolingual practices with interlocutors using Monolingual Hindi or English, and because of relative 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 given the prestige and tangible job advantages their English competence offers.

The Indian urban/rural segregation is also a salient ecological factor mediating Hinglish adoption [63]: the rural NDTV data demonstrates Monolingual Hindi, suggesting that despite a lower fitness, Monolingual Hindi can be maintained. We explain this maintenance despite lower fitness by defining the rural setting as a refuge, akin to prey refuges [64] from predator-prey literature. Given that there is a fair amount of Hindi-speaking rural migration to urban areas, in search of work, we adopt the the refuge concept and hypothesize:

II) **Rural Refuge Hypothesis**
1) Hindi Monolingual conversion to the Hinglish class is contingent upon migration. The Hinglish speaker, as an urban dweller, only has access to maximally influence the percentage of Hindi speakers who are urban immigrants, who migrate from their villages mostly in search of work.
2) The remaining Hindi Monolinguals who do not leave their villages are "protected" (a rural refuge) from interactions and fitness-based conversion and thus maintain Hindi Monolingual status.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

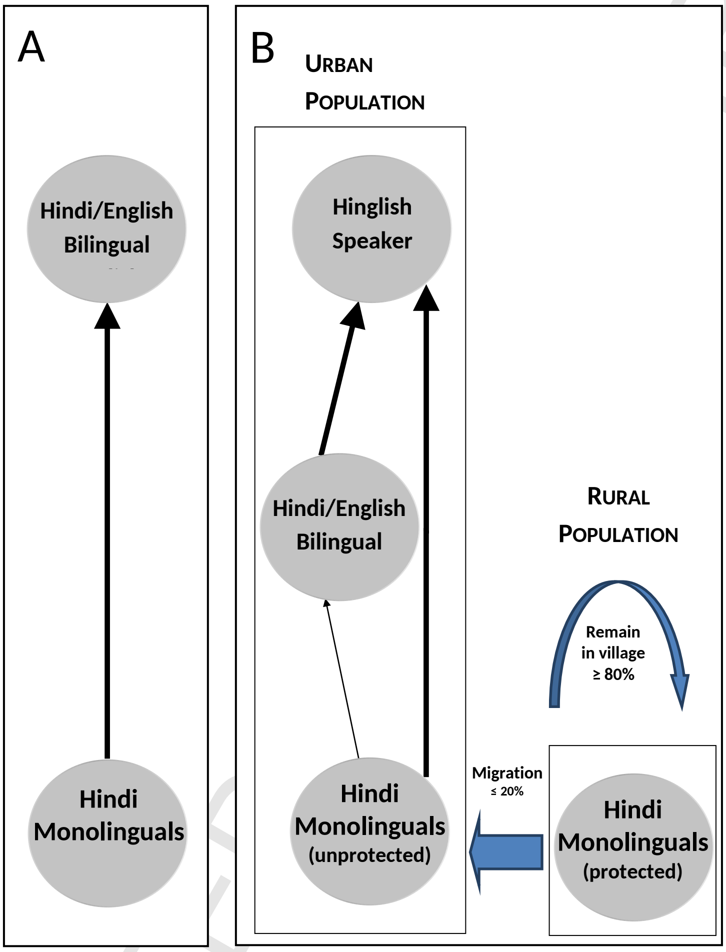

FIGURE 2. Fig. 2 A demonstrates what is traditionally believed about language practices in India, as reported via census data, with movement from the Hindi Monolingual class to the Bilingual class, given the higher prestige of English in India. Fig. 2 B demonstrates what we believe is actually taking place. The population is behaving similar to a three-species generalist top predator-specialist middle predator-prey food chain [51]. The Hinglish speaker (the top predator), due to his greater fitness (demonstrated here with thicker arrows) has climbed to the top of this chain. He can convert Monolinguals (the prey) to acquire Hinglish in a much shorter time than it takes Bilinguals (the middle predator) to convert Monolinguals to acquire English, and thus become Bilingual. Given that many Monolinguals lack the resources for complete English uptake, they will never become Bilingual but are able to learn a number of English words, hence become Hinglish speakers. Note that rural Hindi Monolingual Indians who remain in their villages are in a refuge, thus protected from Hinglish predation.

ACCEPTED MANUSCRIPT

4. A three species ODE model incorporating Hinglish speakers

4.1. **Formulation.** Based on the above hypotheses, and support from past modeling efforts [29], we develop 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, respectively the number of Hinglish, Hindi/English Bilingual and Monolingual Hindi speakers.

$$
\begin{aligned}
(3)\quad \frac{dH}{dt} &= H\left[\frac{\varepsilon_1 a M}{H+B+M} + \frac{a_1 B}{H+B+M} - d_1\right] = Hg_1(H,B,M) \\
\frac{dB}{dt} &= B\left[\frac{-a_1 H}{H+B+M} + \frac{\varepsilon M}{H+B+M} - \frac{d_2 B}{H+B+M}\right] = Bg_2(H,B,M) \\
\frac{dM}{dt} &= M\left[\frac{-\varepsilon_1 a H}{H+B+M} + \frac{-\varepsilon B}{H+B+M} + 1 - \frac{M}{K}\right] = Mg_3(H,B,M)
\end{aligned}
$$

We assume that the gain or loss of speakers to a certain class occurs by direct interaction between the speakers of that class with the proportion of speakers of the other classes, in the whole population. Thus the Hinglish class gains speakers from the Monolingual class according to the direct product $H \times \left(\frac{M}{H+B+M}\right)$, the rate of gain is measured by the parameter ε_1. Here we model this horizontal transmission term $H \times \left(\frac{M}{H+B+M}\right)$, and others, according to the forms derived in [53]. The parameter $0 \le a \le 1$, models refuge for the Monolinguals. That is a measures what proportion of the Monolinguals will be susceptible to interaction with the Hinglish speaker. Likewise the Hinglish class also gains speakers from the Bilingual class according to the direct product $H \times \left(\frac{B}{H+B+M}\right)$, the rate of gain is measured by the parameter a_1. Note, we have framed the Hinglish class as being the fittest of all classes. Thus we assume someone can leave the Hinglish class only due to physical death, modeled via the $-d_1 H$ term. The Bilingual class gains speakers from the Monolingual class according to the direct product $B \times \left(\frac{M}{H+B+M}\right)$. The rate of gain is measured by the parameter ε, furthermore they loose speakers to the Hinglish class according to the direct product $H \times \left(\frac{B}{H+B+M}\right)$, where the rate of loss is measured by the parameter a_1. Furthermore we assume that there is extreme competition for jobs involving English (call centers, etc.) and thus there is a loss term in the Bilingual class that measures this, given by $B \times \left(\frac{B}{H+B+M}\right)$ with the rate of loss measured by the parameter d_2. Lastly the Monolingual class only looses speakers to both the Bilingual class and the Hinglish class, but are also logistically controlled.

It is important to note here that although we assume competition in the Billingual class, no competition is assumed in the Hinglish class. The reason for this is the earlier interlocuter hypothesis that we make. Note we are modeling change within speakers, though it's only allowed for the Bilingual and Monolingual classes. That is, once a person becomes a Hinglish speaker he does not leave that class except due to natural death. Thus a Hinglish speaker is unable to gain a job requiring fluent English, and will also not revert to Monolingual Hindi status. This also fits with what they say for a generic model in [29].

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

4.2. **Equilibrium points and stability.** The nontrivial critical point (H^*, B^*, M^*) is given by

$$M^* = K\left[\frac{\alpha_1+\alpha_2+1-(\varepsilon_1 a\alpha_2+\alpha_1\varepsilon)}{1+\alpha_1+\alpha_2}\right],\ H^*=\alpha_2 M^*,\ B^*=\alpha_1 M^* \tag{4}$$

where α_1 and α_2 are defined as;

$$\alpha_1=\frac{d_1(\varepsilon+a_1)-aa_1\varepsilon}{d_1d_2+aa_1-d_1a_1},\ \alpha_2=\frac{\varepsilon_1 a+\alpha_1 a-d_1(1+\alpha_1)}{d_1}$$

For the positivity of the equilibrium point (H^*, B^*, M^*) we require that the following parametric restrictions hold;

$$\begin{aligned} d_1(\varepsilon+a_1) &> aa_1\varepsilon\\ d_1d_2+a_a1 &> d_1a_1\\ \varepsilon_1 a+\alpha_1 a &> d_1(1+\alpha_1)\\ \alpha_1+\alpha_2+1 &> \varepsilon_1 a\alpha_2+\varepsilon\alpha_1 \end{aligned}$$

The evaluation of the Jacobian at equilibrium is standard. If we let $b_{i,j}$ denote the respective entries of the Jacobian Matrix, with $i,j=1,2,3$, we see that if the following parametric restrictions hold,

$$\begin{aligned} H^*(a_1-d_2)-M^*(\varepsilon+d_2) &> 0\\ \varepsilon_1 aH^*+\varepsilon B^* &< (H^*+B^*+M^*)^2\\ 0<H^*(\varepsilon_1 a-\varepsilon) &< \varepsilon M^*\\ a_1H^*+M^*(a_1-\varepsilon_1 a) &> 0\\ b_{13}b_{32}-b_{33}b_{12} &< 0\\ b_{31}(b_{11}+b_{33})+b_{12}b_{23} &> 0 \end{aligned}$$

then the nontrivial equilibrium point (H^*, B^*, M^*) is locally asymptotically stable. This is easily checked via application of the Routh Hurwitz's condition [54]. Note that one distinctly sees the effect of the refuge parameter a, in these conditions. We next simulate the system (3) for various parameters, to explore possible diachronic outcome scenarios.

4.3. **Conditions for a "Hinglish" invasion.** An interesting result is that if one consider the three species language model (3), and the following parametric restrictions hold

$$(\varepsilon_1 a-d_1)+(a_1-d_1)\frac{\varepsilon}{d_2}>0,\ d_2>\varepsilon^2 \tag{5}$$

then the equilibrium point $(0,\bar{B},\bar{M})$ is locally asymptotically stable in $\mathbb{R}^2$, but is a saddle in $\mathbb{R}^3$, with unstable manifold in the H direction.

This is because, if we consider (H,B,M), in $\mathbb{R}^3$, the Jacobian (J_H^*) of the Hinglish extinction state $(0,\bar{B},\bar{M})$ is given by

$$J_H^*=\begin{pmatrix} g_1(0,\bar{B},\bar{M}) & 0 & 0\\ \bar{B}\frac{\partial g_2}{\partial H} & \bar{B}\frac{\partial g_2}{\partial B}+g_2(0,\bar{B},\bar{M}) & \bar{B}\frac{\partial g_2}{\partial M}\\ \bar{M}\frac{\partial g_3}{\partial H} & \bar{M}\frac{\partial g_3}{\partial B} & \bar{M}\frac{\partial g_3}{\partial M}+g_3(0,\bar{B},\bar{M}) \end{pmatrix}$$

Solving for the eigenvalues we obtain from the roots of the characteristic equation that the eigenvalue corresponding to the H direction $\lambda_1=g_1(0,\bar{B},\bar{M})$. The other two λ_2,λ_3 are found to be negative. Now if $\lambda_1=g_1(0,\bar{B},\bar{M})>0$, then we have

ACCEPTED MANUSCRIPT

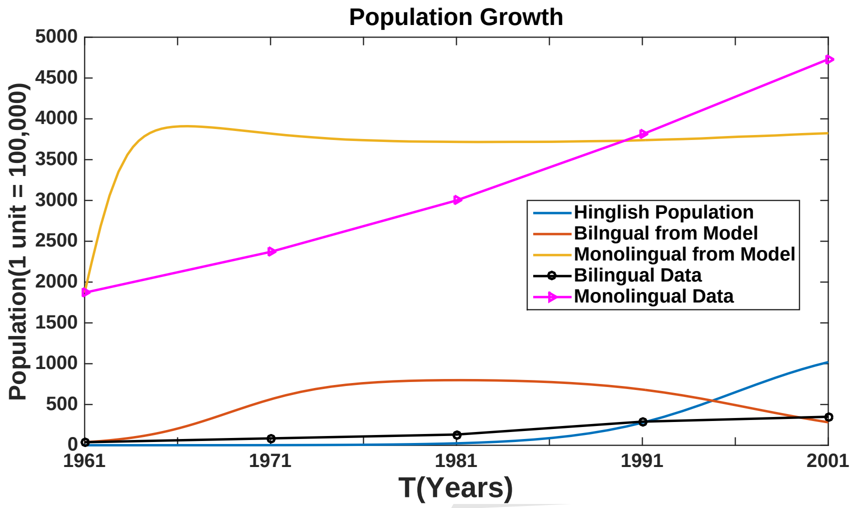

FIGURE 3. In this simulation we use the best fit parameters from our optimisation routine, $\varepsilon = 1.092$, $d_2 = 1.81$, $\varepsilon_1 = 1.2$, $d_1 = .0561$, $a = .05$, $a_1 = 1.3900$, K=4000. The initial conditions used were $H_0 = 1, M_0 = 1870, B_0 = 47$. Thus we assume there are about a 100,000 Hinglish speakers in 1961, where the number of Monolinguals is 187 million, and Bilinguals is 4.7 million. The initial number of Monolinguals and Bilinguals are drawn from the 1961 Indian census data. Our goal here is to simulate the effect of a few Hinglish speakers on the population (while keeping the best fit parameters from our optimisation routine). What we observe in the simulation is that due to the presence of the Hinglish speakers, the Monolinguals and Bilingual numbers actually decrease from about 1965 onwards. The Monolinguals rise again from 1971, but the Bilingual numbers continue to drop. Interestingly, by 2001 the Monolingual numbers are similar to actual census reports of about 400 million. However, Bilinguals are down to a few hundred thousand, with Higlish speakers being in the hundreds of millions.

instability in the Hinglish direction. The sufficient condition for the positivity of this eigenvalue is derived to be $(\varepsilon_1 a - d_1) + (a_1 - d_1)\alpha = (\varepsilon_1 a - d_1) + (a_1 - d_1)\frac{\varepsilon}{d_2} > 0$.

The above condition tells us that if the rate of gain of speakers for the Hinglish speakers $\varepsilon_1 a + \alpha a_1$ is more than their death rate d_1, or equivalently if their fitness is high enough, then even a few Hinglish speakers will be able to establish themselves sizeably in the population. That is a "Hinglish" invasion will take place.

ACCEPTED MANUSCRIPT

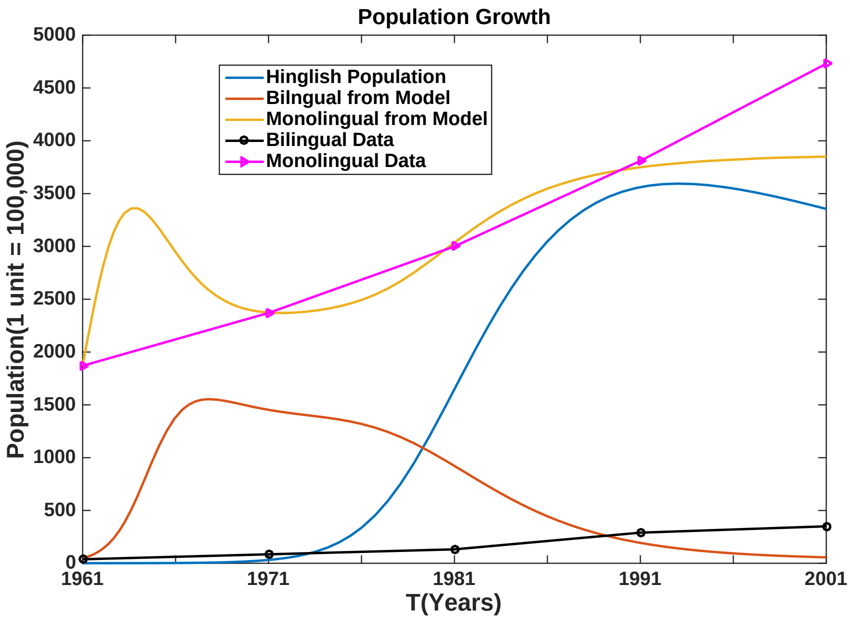

FIGURE 4. In this simulation we adjust parameters and choose $\varepsilon = .4$, $d_2 = 1.81$, $\varepsilon_1 = 1.1$, $d_1 = .11$, $a = .09$, $a_1 = 1.63$, K=4000. The initial conditions used were $H_0 = 1, M_0 = 1870, B_0 = 38$. Thus we assume there are about a 100,000 Hinglish speakers in 1961, where the number of Monolinguals is 187 million, and Bilinguals is 3.8 million. The initial number of Monolinguals and Bilinguals are drawn from the 1961 Indian census data. Our goal here is to simulate the effect of a few Hinglish speakers on the population (while deviating from the best fit parameters our optimisation routine). What we observe here is that the Hinglish speakers take much longer to start to increase in numbers. It is not till about 1990, that there numbers increase distinctly. By 2001 we see that the Bilingual numbers are close to census reports of about 40 million, and Monolingual numbers are also at about 370 million. However, there are about a 100 million Hinglish speakers in the population, more than twice the number of Bilinguals. This is very interesting as 1990 corresponds with the time that mass scale economic liberalisation started in India, leading to an English outbreak.

ACCEPTED MANUSCRIPT

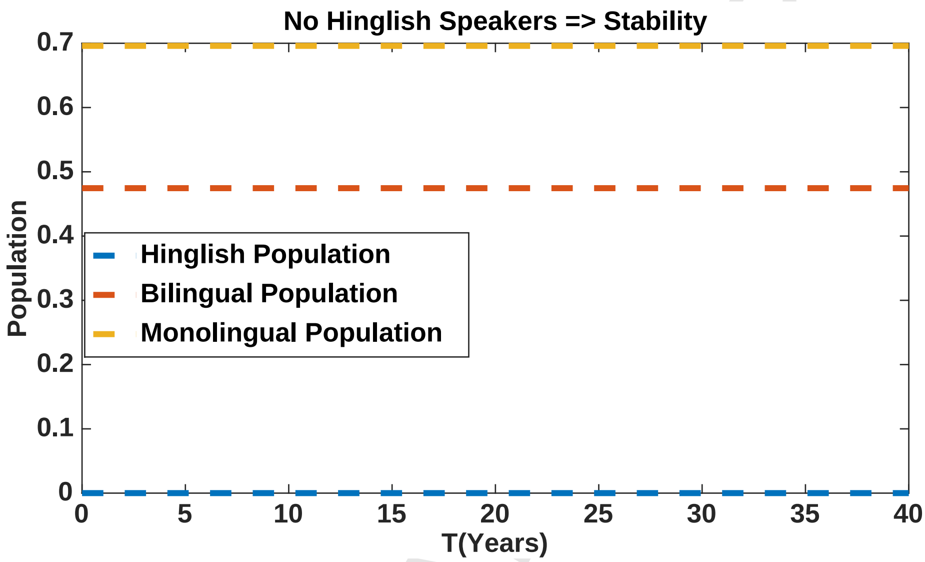

FIGURE 5. Here we numerically illustrate (B, M) are stable in the absence of H, that is in $\mathbb{R}^2$. If there are no Hinglish speakers $(H = 0)$ B, M do not deviate from equilibrium, under a small perturbation.

4.4. **Bilingual Extinction State.** In the absence of Bilingual speakers, the model (3) reduces to,

$$\frac{dH}{dt} = H\left[\frac{\varepsilon_1 a M}{H+M} - d_1\right]$$
$$\frac{dM}{dt} = M\left[\frac{-\varepsilon_1 a H}{H+M} + 1 - \frac{M}{K}\right]$$

The equilibrium point for the above is given by $(H^*, M^*) = (\beta M^*, M^*)$, where $\beta = \frac{\varepsilon_1 a - d_1}{d_1}$ and $M^* = K[1 - \frac{\varepsilon_1 a\beta}{\beta+1}]$ Thus for positivity of the Bilingual extinction state we require; $\beta = \frac{\varepsilon_1 a - d_1}{d_1} > 0$ and $[1 - \frac{\varepsilon_1 a\beta}{\beta+1}] > 0$; This yields $\varepsilon_1 a - d_1 > 0$ and $1 + \frac{1}{\beta} > a\varepsilon_1$

The Jacobian is given by;

$$J^*_B = \begin{pmatrix} -\frac{H^*M^*a\varepsilon_1}{(H^*+M^*)^2} & \frac{H^{*2}a\varepsilon_1}{(H^*+M^*)^2} \\ -\frac{M^{*2}a\varepsilon_1}{(H^*+M^*)^2} & M^*[\frac{a\varepsilon_1 H^*}{(H^*+M^*)^2} - \frac{1}{K}] \end{pmatrix}$$

Note, $Trace(J^*_B) = -\frac{M^*}{K} < 0$ and $Det(J^*_B) = \frac{a\varepsilon_1 H^* M^{*2}}{K(H^*+M^*)^2} > 0$. Therefore via standard stability theory the Bilingual extinction state is locally asymptotically stable in the $H - M$ plane. We now ask the following question: what is the effect of introducing a few Bilingual speakers into the population? That is if we consider $\mathbb{R}^3$. In this case we obtain that the Jacobian (J^*_B) of the Bilingual extinction state $(\tilde{H}, 0, \tilde{M})$ is given by;

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

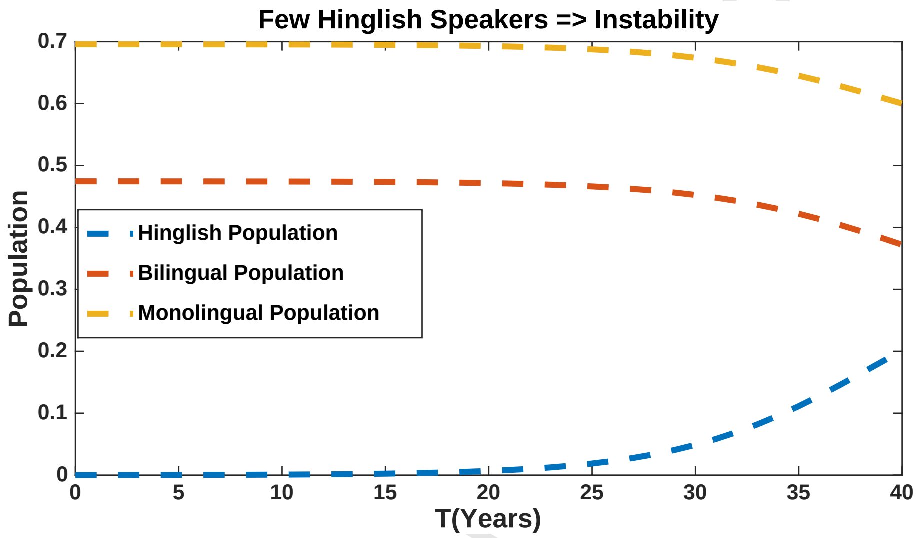

FIGURE 6. Here we numerically illustrate that although (B, M) are stable in the absence of H, that is in $\mathbb{R}^2$, under certain parameter restrictions, there is actually a saddle node dynamics in $\mathbb{R}^3$. Essentially this demonstrates that if the fitness of the Hinglish speakers is high enough, then even a few Hinglish speakers will be able to establish themselves sizeably in the population, and a "Hinglish" invasion will take place.

$$J_B^* = \begin{pmatrix} \tilde{H}\frac{\partial g_1}{\partial H} + \tilde{g}_1 & \tilde{H}\frac{\partial g_1}{\partial B} & \tilde{H}\frac{\partial g_1}{\partial M} \\ 0 & \tilde{g}_2 & 0 \\ \tilde{M}\frac{\partial g_3}{\partial H} & \tilde{M}\frac{\partial g_3}{\partial B} & \tilde{M}\frac{\partial g_3}{\partial M} + \tilde{g}_3 \end{pmatrix}$$

Solving for the eigenvalues we obtain from the roots of the characteristic equation that the eigenvalue corresponding to the B direction $\lambda_2 = g_2(\bar{H}, 0, \bar{M})$. The other two λ_1, λ_3 are found to be negative. If $a_{22} = g_2(\bar{H}, 0, \bar{M}) < 0$, this will result in stability in the Bilingual class over time. $a_{22} < 0$ is equivalent to $(\beta+1)d_2 > \beta a_1 - \varepsilon$. This condition tells us that if the competition coefficient among the Bilinguals d_2 is more than the net gain of speakers to the Bilingual class, then a few Bilinguals will not be able to establish themselves in the population.

5. THE PDE MODEL

5.1. **Formulation.** Populations disperse in space, while space can also itself influence their interactional dynamics and language outcomes. Thus it is important to consider the effect of geographical factors on the language dynamics of the three classes. Similar to the approach taken in [22, 9, 21, 17, 19].

We consider a reaction diffusion version of (3), where the local random movement of the Hinglish speakers, Bilinguals and Monolinguals is modeled via diffusion. The interaction terms between speakers in the three classes are kept the same as in (3), and so we obtain the following system of partial differential equations

ACCEPTED MANUSCRIPT

$$
\begin{aligned}
(6) \qquad \frac{\partial H}{\partial t} &= D1\Delta H + H\left[\frac{\varepsilon_1 a M}{H+B+M} + \frac{a_1 B}{H+B+M} - d_1\right] \\
\frac{\partial B}{\partial t} &= D2\Delta B + B\left[\frac{-a_1 H}{H+B+M} + \frac{\varepsilon M}{H+B+M} - \frac{d_2 B}{H+B+M}\right] \\
\frac{\partial M}{\partial t} &= D3\Delta M + M\left[\frac{-\varepsilon_1 a H}{H+B+M} + \frac{-\varepsilon B}{H+B+M} + 1 - \frac{M}{K}\right]
\end{aligned}
$$

Here $D1, D2, D3$ are the diffusion coefficients, that is the speeds with which the populations disperse. The spatial domain is $\Omega \subset \mathbb{R}^N$, $N = 1, 2$. Neumann boundary conditions $\nabla H \cdot \mathbf{n} = \nabla B \cdot \mathbf{n} = \nabla M \cdot \mathbf{n} = 0$ on $\partial\Omega$, are prescribed, as are suitable positive initial conditions $H_0(x), B_0(x), M_0(x)$.

The reaction terms here provide no obstacle to global existence of solutions, and apriori estimates on the solutions is established via standard methods [65]. This says there is a global attractor $\mathcal{A}$ for the system, which is the repository of *all* the long time dynamics. It attracts all bounded subsets of $L^2(\Omega)$ in the $L^2(\Omega)$ metric. The attractor in certain parameter regime, is a simple one point attractor, that is the spatially homogenous steady state (H^*, B^*, M^*).

5.2. **Turing Instability and Turing patterns.** Turing instability occurs when the positive interior equilibrium point for an ODE system is stable in the absence of diffusion, but becomes unstable due to the addition of diffusion. This phenomena is also referred to as diffusion driven instability, [66]. To see if Turing instability exists in (6), we first linearize the system (6) about the homogenous steady state, by introducing both space and time-dependent fluctuations around the feasible steady state solution (H^*, B^*, M^*), and apply the well known procedure [67] to derive the instability conditions in three species models. The result is that the diffusive language model (6), with equilibrium point (H^*, B^*, M^*), *does indeed posess Turing instability*. Thus we can get Turing patterns by choosing an appropriate range of parameters. This results demonstrate that different diffusion rates of the Bilinguals, Monolinguals and Hinglish populations can cause geographic inhomogeniety akin to [17]. This is a basis for the formation of rural and urban centers. Some Turing patterns are shown in the following simulations,

ACCEPTED MANUSCRIPT

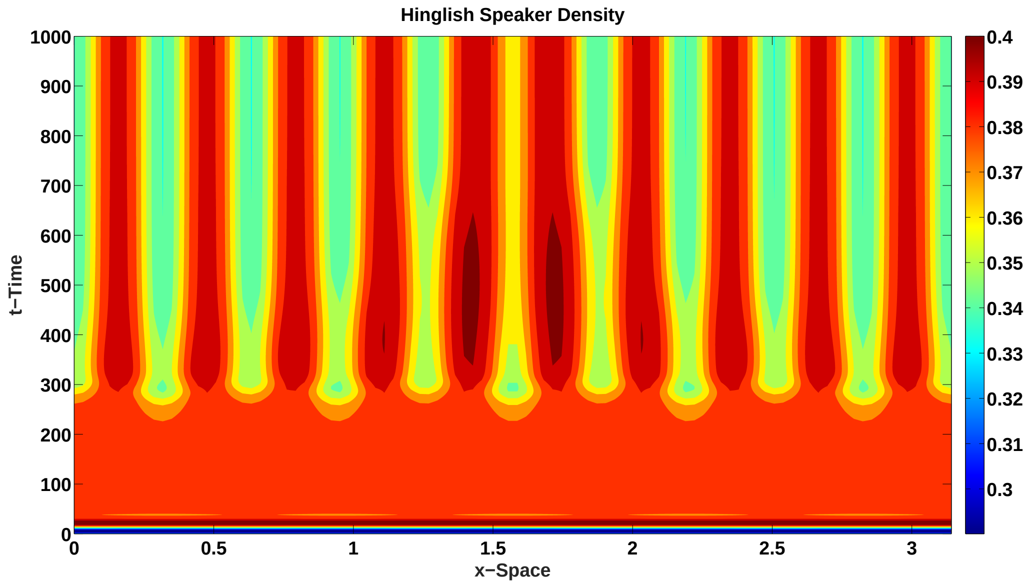

FIGURE 7. In this experiment we consider an initial condition where the equilibrium is perturbed by $0.005cos^2(5x)$. Here we perform a 1-d simulation. The parameters used are $a = 1.000; a_1 = 0.3800; \varepsilon = 0.910; \varepsilon_1 = 1.190; d_1 = 0.381; d_2 = 0.31; D1 = 0.01; D2 = 0.01; D3 = 0.00001, K = 1;\ H^* = 0.336; B^* = 0.050; M^* = 0.120$. What we observe is that a stripe like Turing pattern is formed. The simulation starts of with the spatially homogenous steady state $H^* = 0.336; B^* = 0.050; M^* = 0.120$, but soon the Turing instability picks up, and leads to a spatially inhomogenous state. A contour plot in the $x - t$ plane is shown for the density of the Hinglish speakers. The simulation is run till $t = 1000$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

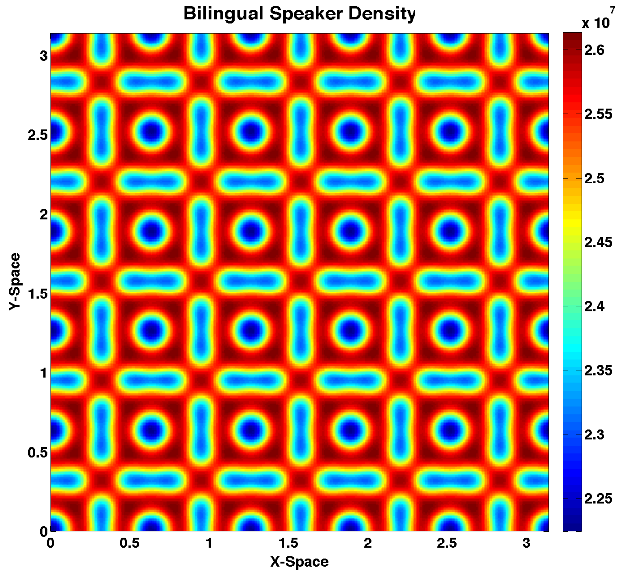

FIGURE 8. In this experiment we consider an initial condition where the equilibrium is perturbed by $0.005cos^2(5x)cos^2(5y)$. Here we consider a 2-d simulation. The parameters used are $a = 1, a_1 = 0.38, \varepsilon = 0.91, \varepsilon_1 = 1.19, d_1 = 0.381, d_2 = 0.31, D1 = 0.01, D2 = 0.01, D3 = 0.00001, K = 4 \times 10^8$. What we observe is that a maze like Turing pattern is formed. The simulation starts of with the spatially homogenous steady state but soon the Turing instability picks up, and leads to a spatially inhomogenous state. Here a contour plot of the density in the $x - y$ plane is shown for the density of the Bilingual speakers. The simulation is run till $t = 3000$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

6. Discussion and Conclusion

While earlier language competition research has focused on deriving models that match existing census data and only explores canonical "languages", we investigate the effect of a hybrid Hinglish obligatory code-switching population based on three types of empirical field evidence. Contemporary Hindi-only reality show data and independent sociolinguistic interviews demonstrate Hinglish code-switching, supporting our hypothesized Hinglish population. Our hypothesis for a rural stronghold for Hindi Monolingualism is confirmed with rural NDTV interview data, which does not demonstrate Hinglish. Based on this field evidence we model Indian language shift via a three-species predator-prey framework with a prey refuge, which better reflects the reality of the Indian context and sociolinguistic findings from other (post)colonial and language contact scenarios. Furthermore we model excessive competition for resources such as jobs, within the Bilingual class through a loss parameter. The inclusion of a rural refuge with limited urban movement provides a way to operationalize the pervasive strength of the Hindi population and the links between interactional patterns, fitness and conversion within the Hindi Belt. Our application of a three-species predator-prey model, field evidence and subsequent introduction of localized ecological factors conditioning Indian language dynamics are intended as a first step toward the quantitative modeling of hybrid language competition and Indian language dynamics.

The modeling results show via (4) that a coexistence state is possible within these three classes, and is probably the most realistic outcome. Although a Bilingual extinction state is possible, this is less statistically likely. There will always be true Bilinguals, given religious, cultural and business domains that require Monolingual Hindi or English. This is also in line with the coexistence results of [29, 7, 26]. The result via the parameteric restriction (5) is perhaps the most interesting. This says that for ε_1 (the rate of gain of the Monolinguals into the Hinglish class), significantly larger than d_1 (the death rate of Hinglish speakers), there is a saddle node dynamics, with instability in the Hinglish direction. This means that an introduction of even a very small number of Hinglish speakers can lead to them establishing a large population, if they can convert Monolinguals to Hinglish speakers "efficiently" enough. This implies that Hinglish speakers whose fitness is above a certain threshold will *always* establish themselves in the population. This then is requirement, for a "Hinglish invasion". The Turing instability result seen via Figs. 7 - 8 demonstrates that different spreading speeds of the Bilinguals, Monolinguals and Hinglish populations can lead to geographic inhomogeneity. This is a basis for the formation of rural and urban centers.

There are many possibilities for future reseacrh. Note in reality, there is also "backward" migration. That is, a small percentage of people may move *back* to villages from urban areas, and there are also rural-urban commuters. Some people will often commute 2-3 hours a day (by trains etc.) to go into cities for work, and commute back to their villages at night. It is conceivable that these people could bring Hinglish to rural areas. This is not considered in the current model. It would also be interesting to consider the effect of spatial refuges, and spatially dependent diffusion coefficients for the populations. While we focus on the Indian setting, unbalanced multilingual competency and limited access to the dominant or newly developed hybrid practices can lead to obligatory code-switching in other settings: the ecological factors and outcomes addressed here are not unique to the

ACCEPTED MANUSCRIPT

Indian context. Instead, they reflect a dynamic process involving various types of interactions and classes which are conditioned by their local ecology, Collectively, these better explain both creole development and the emergence and staying power of other hybrid mixed practices, e.g. Singlish in Singapore and Taglish in the Philippines [1]. In addition, localized ecological factors need to be considered in future language dynamics modeling efforts more broadly.' The other approach one might take is to try and model possible controls, that might suppress Hinglish and reinforce true Bilingualism, similar to the recent approaches in [6, 26].

Another future direction of interest to researchers applying statistical physics methods to language dynamics, would be the analysis of corpora that contains Hinglish, in order to test if various statistical laws such as Zipf's law hold for emerging varieties of English like that in India, and/or for hybrid varieties like Hinglish. Lastly, while our modeling approach builds on various definitions to predict possible future language dynamics in the Indian Hindi belt, it also provides a means to explore hybrid language competition dynamics in *other settings* across the world.

7. Acknowledgments

We would like to acknowledge that we use national Indian census data on Hindi/English Bilingual population numbers provided to us by David Clingingsmith for the years 1961 and 1991, and by William Cotter for the years 1971 and 1981. We would also like to acknowledge valuable comments from the two anonymous referees that helped us greatly improve the quality of the manuscript.

References

[1] S.S. Mufwene, Language Evolution: Contact, Competition and Change, Continuum International Publishing Group, London, 2008.

[2] V. Loreto, L. Steels, Social dynamics: emergence of language, Nat. Phys. 3 (2007) 758-760.

[3] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Rev. Math. Phys. 81 (2009) 591-646.

[4] V. Loreto, A. Baronchelli, A. Mukherjee, A. Puglisi, F. Tria, Statistical physics of language dynamics. J. Stat. Mech. Theor. Exp. 2011 (2011) P04006.

[5] M. Nowak, D. Krakauer, The evolution of language, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 8028-8033.

[6] L.F. Nie, Z.D. Teng, J.J. Nieto, I.H. Jung, Dynamic Analysis of a Two-Language Competitive Model with Control Strategies, Mathematical Problems in Engineering Volume 2013, Article ID 654619 (2013) 1-13.

[7] C. Walters, A reaction-diffusion model for competing languages, Meccanica 49 (2014) 2189-2206.

[8] D.M. Abrams, S.H. Strogatz, Modelling the dynamics of language death, Nature 424 (2003) 900.

[9] A. Kandler, Demography and language competition, Hum. Biol. 81 (2009) 181-210.

[10] J.M. Minett, W.S. Wang, Modeling endangered languages: The effects of bilingualism and social structure, Lingua 118 (2008) 19-45.

[11] M.A. Nowak, N.L. Komarova, P. Niyogi, Computational and evolutionary aspects of language, Nature 417 (2002) 611-617.

[12] W.S. Wang, J.W. Minett, The invasion of language: Emergence, change and death, Trends Ecol. Evol. 20 (2005) 263-269.

[13] M. Perc, Evolution of the most common English words and phrases over the centuries, J. R. Soc. Interface 9 (2012) rsif20120491.

[14] A. Petersen, J. Tenenbaum, S. Havlin, H. Stanley, M. Perc, Languages cool as they expand: Allometric scaling and the decreasing need for new words, Sci. Rep. 2 (2012).

[15] M. Perc, Self-organization of progress across the century of physics, Sci. Rep. 3 (2013).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[16] T. Kuhn, M. Perc, D. Helbing, Inheritance patterns in citation networks reveal scientific memes, Phys. Rev. X 4 (2014) 041036.
[17] M. Patriarca, E. Heinsalu, Influence of geography on language competition, Physica A 388 (2009) 174-186.
[18] E. Bakalis, A. Galani, Modeling language evolution: Aromanian, an endangered language in Greece, Physica A 391 (2012) 4963-4969.
[19] J. Pinasco, L. Romanelli, Coexistence of languages is possible, Physica A 361 (2012) 355-360.
[20] Z. An, Q. Pan, G. Yu, Z. Wang, The spatial distribution of clusters and the formation of mixed languages in Bilingual competition, Physica A 391 (2012) 4943-4952.
[21] J.D. Murray, Mathematical Biology, Springer-Verlag, Berlin, 1993.
[22] F. Isern, J. Fort, Language extinction and linguistic fronts, J. R. Soc. Interface 11 (2014) 20140028.
[23] J. Mira, A. Paredes, Interlinguistic similarity and language death dynamics, Europhys. Lett. 69 (2005) 1031-1034.
[24] J. Mira, L.F. Seoane, J.J. Nieto, The importance of interlinguistic similarity and stable Bilingualism when two languages compete, New J. Phys. 13 (2011) 033007.
[25] I. Baggs, H.I. Freedman, A mathematical model for the dynamics of interactions between a unilingual and a Bilingual population: Persistence versus extinction, J. Math. Sociol. 16 (1990) 51-75.
[26] L.F. Nie, Z.D. Teng, J.J. Nieto, I.H. Jung, State impulsive control strategies for a two-languages competitive model with bilingualism and interlinguistic similarity, Physica A 430 (2015) 136-147.
[27] M. Zhang, T. Gong, Principles of parametric estimation in modeling language competition, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 9698-9703.
[28] C. Schulze, D. Stauer, Competition of languages in the presence of a barrier, Physica A 379 (2007) 661-664.
[29] E. Heinsalu, M. Patriarca, J.L. Leonard, The role of Bilinguals in language competition, Advances in Complex Systems 17 (2014) 1-16.
[30] V. Chand, Elite positionings toward Hindi: Language policies, political stances and language competence in India, J. Socioling. 15 (2011) 1-30.
[31] Census of India, Abstract of Speakers Strength of Languages and Mother Tongues 1961-2001, Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, Government of India, Delhi, 2001.
[32] C. Jarelot, The rise of the other backward classes in the Hindi belt, J. Asian Stud. 59 (2000) 86-108.
[33] C. LaDousa, In the mouth but not on the map: visions of language and their enactment in the Hindi belt, J. Pragmat. 36 (2004) 633-661.
[34] Asian Development Bank, Special chapter: The rise of Asia's middle class, in: Key Indicators for Asia and the Pacific, 2010, pp. 3-48.
[35] R.P. Gosal, A century of population change in India: A spatial perspective, in: S. Nangia, N.C. Jana, R.B. Bhagat (Eds.), State of Natural and Human Resources in India, Volume 2, Concept Publishing Company, New Delhi, 2010, pp. 361-379.
[36] D. Graddol, English Next India, British Council, London, 2010.
[37] M. Azam, A. Chin, N. Prakash, The returns to English language skills in India, Econ. Devel. Cult. Change 61 (2013) 335-367.
[38] R.D. Parshad, V. Chand, N. Sinha, N. Kumari, What is India speaking: The "Hinglish" invasion, arXiv preprint arXiv:1406.4824 (2014).
[39] D. Schreier, M. Hundt, English as a contact language, Princeton University Press, Princeton, New Jersey, 2013.
[40] S.N. Sridhar, Language contact and convergence in South Asia, in: B.J. Kachru, Y. Kachru, S.N. Sridhar (Eds.), Language in South Asia, Cambridge University Press, Cambridge, 2008, pp. 235-252.
[41] R. Kothari, R. Snell, Chutnefying English: The phenomenon of Hinglish, Penguin Books India, New Delhi, 2011.
[42] A. Chauhan, Those pricey Thakur girls, Harper Collins Publishers India, Noida, 2013.
[43] A. Si, A diachronic investigation of Hindi-English code-switching, using Bollywood film scripts, Int. J. of Biling. 15 (2010) 388-407.
[44] P. Gardner-Chloros, Code-Switching, Cambridge University Press, Cambridge, 2009.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[45] P. Auer, Code-switching in conversation: Language, interaction and identity, Routledge, London, 1998.

[46] A. Sarhimaa, Syntactic transfer, contact-induced change, and the evolution of Bilingual mixed codes: Focus on Karelian-Russian language alternation, Finnish Literature Society, Helsinki, 1999.

[47] I. Smith, Pidgins, Creoles, and Bazaar Hindi, in: B.J. Kachru, Y. Kachru, S.N. Sridhar (Eds.), Language in South Asia, Cambridge University Press, Cambridge, 2008, pp. 253-268.

[48] B.K. John, Entry from backside only: Hazaar fundas of Indian-English, Penguin Books, New Delhi, 2007.

[49] S. Shankar, Reel to real: Desi teens' linguistic engagement with Bollywood, Pragmatics 14 (2004) 317-335.

[50] B. Mahal, The Queen's Hinglish: How to speak pukka, Harper Collins, Glasgow, 2006.

[51] A. Hastings, T. Powell, Chaos in a three-species food chain, Ecology 72 (1991) 896-903.

[52] D. Clingingsmith, Industrialization and Bilingualism in India, J. Hum. Resour. 49 (2014) 73-109.

[53] L. Cavalli-Sforza, M. Feldman, Cultural Transmission and Evolution: A Quantitative Approach, Cambridge University Press, Cambridge, UK, 1981.

[54] S. Strogatz, Nonlinear dynamics and Chaos: With applications to physics, biology and chemistry, Perseus, Cambridge, Massachusetts, 1994.

[55] M.S. Schmid, Language attrition: Theoretical perspectives, in: B. Koopke, M.S. Schmid, M. Keijzer, S. Dostert (Eds), John Benjamins, Amsterdam, 2007, pp. 135-153.

[56] R. Burger, The mathematical theory of selection, recombination and mutation, Wiley, Chichester, 2000.

[57] J. Pierrehumbert, Exemplar dynamics: Word frequency, lenition, and contrast, in: J. Bybee, P. Hopper (Eds.), Frequency effects and the emergence of linguistic structure, John Benjamins, Amsterdam, 2000, pp. 137-157.

[58] Y. Matras, Fusion and the cognitive basis for Bilingual discourse markers, Int. J. of Biling. 4 (2000) 505-528.

[59] L. Pucci , P. Gravino, V. Servedio, Modeling the emergence of a new language: Naming game with hybridization, in: Self-Organising Systems, Springer Berlin Heidelberg, 2014, pp. 78-89.

[60] V. Loreto, A. Mukherjee, F. Tria, On the origin of the hierarchy of color names, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 6819-6824.

[61] E. Lieberman, J.B. Michel, J. Jackson, T. Tang, M. Nowak, Quantifying the evolutionary dynamics of language, Nature 449 (2007) 713-716.

[62] A. Bell, Language style as audience design, Language in Society 13 (1984) 145-204.

[63] D. Britain, Space and Spatial Diffusion, in: J. K. Chambers, P. Trudgill, N. Schilling-Estes (Eds.), The Handbook of Language Variation and Change, Oxford, Blackwell, 2002, pp. 603-637.

[64] T.K. Kar, Stability analysis of a predator-prey model incorporating a prey refuge, Commun. Nonlinear. Sci. Numer. Simul. 10 (2005) 681-691.

[65] R. Temam, Infinite-Dimensional Dynamical Systems in Mechanics and Physics, 2^{nd} ed, Applied Mathematical Sciences 68, Springer, New York, 1997.

[66] A. Turing, The chemical basis of morphogenesis, Phil. Trans. R. Soc. B 237 (1952) 37-72.

[67] K.J. White, C.A. Gilligan, Spatial heterogeneity in three species, plant-parasite-hyperparasite systems, Phil. Trans. R. Soc. B 353 (1998) 543-557.

ACCEPTED MANUSCRIPT