JASSS

Monojit Choudhury, Anupam Basu and Sudeshna Sarkar (2006)

Multi-Agent Simulation of Emergence of Schwa Deletion Pattern in Hindi

For information about citing this article, click here

Abstract

Recently, there has been a revival of interest in multi-agent simulation techniques for exploring the nature of language change. However, a lack of appropriate validation of simulation experiments against real language data often calls into question the general applicability of these methods in modeling realistic language change. We try to address this issue here by making an attempt to model the phenomenon of schwa deletion in Hindi through a multi-agent simulation framework. The pattern of Hindi schwa deletion and its diachronic nature are well studied, not only out of general linguistic inquiry, but also to facilitate Hindi grapheme-to-phoneme conversion, which is a preprocessing step to text-to-speech synthesis. We show that under certain conditions, the schwa deletion pattern observed in modern Hindi emerges in the system from an initial state of no deletion. The simulation framework described in this work can be extended to model other phonological changes as well.

Keywords:

Introduction

1.1

Recently, there has been a revival of interest in computational methods for investigating language change. This is partly because of the facts that languages do not fossilize and it is nearly impossible to verify any hypothesis regarding language change through laboratory experiments. Therefore, the only possibilities left with the researchers in diachronic linguistics are to collect as much historical data as possible and try to establish their theories based on this data as well as indirect evidence gathered from disciplines like cognitive sciences, biological sciences and other social sciences. Computational techniques provide a way to circumvent some of these problems, at least partially, by facilitating precise modeling of the problem and virtual experimentation.

1.2

There are two complementary views to a problem in diachronic linguistics for which computational models can be designed — *functional* and *emergent*. To explain why a particular linguistic change has taken place, in the functional model, it suffices to show that under a given set of causal forces (identified independent of the model), the linguistic structure that is

functionally optimal (identified computationally from the model) coincides with what is observed in reality. This has been traditionally modeled as a constrained optimization problem, which in turn can be solved using standard computational methods like genetic algorithms, simulated annealing, numerical methods, linear or non-linear programming etc. As an example, let us consider the problem of explaining the universal principles observed in the vowel inventories of the languages all over the globe[1]. Liljencrants and Lindblom (1972) put forward a constrained optimization model, where the vowels are represented as points within a bounded two-dimensional acoustic plane and the optimization function is to minimize the mean of the inverse-square Euclidean-distances between the vowels. This explanation was motivated by the acoustic distinctiveness principle (Schwartz et al 1997). They used numerical simulations for solving the optimization problem and arrived at vowel systems closely resembling the naturally occurring ones. Ke, Ogura and Wang (2003) used a genetic algorithm as an optimization tool to model the vowel as well as tone systems, and arrived at similar results, but from more general principles.

1.3

A functional model is a necessary, but not a sufficient explanation, because it does not tell us how the optimization might have taken place in reality (Oudeyer 2005). There are at least three reasons for which this question deserves a non-trivial explanation: first, a naïve Darwinian search with random mutations might not be sufficient to explain the emergence of a complex pattern; second, the speakers are generally oblivious to the fact that the language they speak is undergoing some structural change and they participate in the process being quite unaware of it; and third, language change takes place in a distributed environment without any central control. Therefore, one must be able to provide a self-organizing model of language change based on realistic assumptions about the language users and their interactions. Multi-agent simulation (MAS) and dynamical system models are two popular techniques for providing such explanations. A MAS model for the emergence of the vowel inventories has been described by de Boer (2001), which shows how starting from a completely random distribution of the vowels, a group of speakers can arrive at a shared vowel system that has properties similar to vowel inventories of real languages. However, unlike Liljencrants and Lindblom (1972) and Ke et al (2003), de Boer's work does not assume any global functional optimization principle; the only assumption it makes is about an inherent drive for successful communication among the speakers. Recent works (Smith 2005; Oudeyer 2005) have shown that even this assumption may not be necessary for explaining the emergence of structural patterns.

1.4

In spite of their wide popularity, most of the computational models developed for explaining language change and evolution suffer from a serious drawback: they are too simplistic to model the reality (Hauser et al 2002). Most of the models have been developed for explaining certain general linguistic phenomena illustrated on toy languages rather than realistic language change that are validated against real data (of course there are exceptions to this, some of which will be discussed below). This is because more often than not, reality is too complex to yield a tractable computational model. On the other hand, models that are validated against real language change data can provide further insight into the phenomena investigated, giving more reasons to believe the plausibility of the model. In this work, we make an attempt to address this issue by modeling the emergence of a phonological phenomenon called schwa deletion for Hindi and compare the results with real language data. Despite the fact that we make several simplifying assumptions to retain the computational tractability and understandability of the models developed, the emergent pattern closely resembles the observed schwa deletion pattern in modern Hindi.

1.5

Schwa deletion refers to the context dependent deletion of the mid-central vowel schwa (Ohala 1983) and is an important issue in Hindi phonology that must be tackled suitably in order to develop a grapheme-to-phoneme converter for Hindi text-to-speech synthesis (Choudhury and Basu 2002; Kishore and Black 2003; Narasimhan et al 2004). The diachronic nature of this phenomenon (Misra 1967; Choudhury et al 2004) and its non-triviality (Kishore and Black 2003; Narasimhan et al 2004) have been pointed out by several researchers working on linguistic and computational aspects of this problem. Choudhury et al (2004) provided a functional explanation for the schwa deletion pattern in Hindi and proposed a constrained

optimization model for the same based on syllable minimization. However, like other optimization models (discussed above), their model too is silent about how such a complex optimization might have taken place without any conscious effort by the speakers and central regulation. We try to answer these questions through MAS.

1.6

The MAS model described here assumes that the speakers have an inherent tendency to reduce the duration of the schwas for faster communication (see [Lupyan and McClelland 2003](#) for example), thereby providing a bias towards schwa deletion in the system. However, as we shall see, there is no assumption or bias embedded in the system towards the specific schwa deletion pattern of Hindi. We show that under certain parameter settings, the schwa deletion pattern that emerges in the MAS model is similar to the pattern observed in reality. In fact, several interesting phenomena, like dialectal variation and the S–shaped dynamics that are observed during real language change, automatically emerge in the system. The sensitivity of the emergent pattern to the parameter settings as well as the vocabulary, and its stability over different runs of the simulation for the same parameter settings imply that the emergence is not an outcome of some random factors in the model.

1.7

Thus, to summarize, the objectives of the current work are to describe a general MAS framework for simulating language change, and to show that the emergence of the schwa deletion pattern of modern Hindi can be captured by the model. It should be noted that the current work does not make any claim about the emergence of the schwa deletion phenomenon; rather it tries to give a diachronic explanation for the schwa deletion pattern that is observed. In other words, the work does not explain why there is schwa deletion. Assuming that there is a tendency for schwa deletion, it shows why in certain contexts the schwas are deleted and why in certain other contexts it is not.

1.8

The paper is organized as follows: Section [2](#) tries to place the current work in the context of the existing research findings by discussing the basic issues in language change, previous attempts of modeling language change through MAS, schwa deletion in Hindi and an optimization model of the problem ([Choudhury et al 2004](#)). The MAS framework is described in the next section. Section [4](#) presents the experimental results and their analyses. The concluding section summarizes the contributions of this work and discusses possible enhancements of the current model. In this paper, Hindi scripts are written using Roman characters following the ITRANS convention ([Chopde 2001](#)). To avoid IPA symbols, we will use ITRANS to represent the pronunciations, according to which "*a*", pronounced as "e" in "the", represents the schwa (/ax/ in ARPAbet) and "*A*" represents the sound of "a" in "after' (/ae/ in ARPAbet).

🌐 Background

Language Change

2.1

The phenomenon of language change is formally studied under *diachronic linguistics* (also known as *historical linguistics*). Apart from the questions of the causes, effects and the course of language change, diachronic linguistics also studies the languages of the past. *Synchronic linguistics* on the other hand studies languages as they are/were at a particular point of time. The generative tradition[2] views the grammar of a language from the synchronic perspective as "an instantiation of the initial state of the cognitive system of the language faculty with options specified" ([Chomsky 1995](#)). In other words, the grammar is defined as a fixed set of *principles*, the so-called *Universal Grammar* (UG), and each language is just a particular instantiation of the principles to certain parameters. In this principles–and–parameters framework, language acquisition is the process of setting the right parameters after observing a set of triggers, i.e. linguistic inputs; and language change is the process of variation in the parameter values over time ([Roberts 2001](#)).

Forces in Language Change

2.2

Languages are stable over a large period of time. Children acquire their parents' (target) grammars without any error. The transmission of a particular language from one generation to another thus takes place with hardly any imperfection. Given these facts, how can one explain the observation that languages change spontaneously, often without any external influences? Apparently, we are faced here with a logical paradox of language change — the biggest mystery of diachronic linguistics (Lightfoot 1991, 1999; Clark and Roberts 1993; Niyogi and Berwick 1998). To explain this paradox, we have to look deeper into the nature of language transmission and other forces governing language change. Figure 1 schematically represents how a language is transmitted from one generation of speakers to another. The bubbles denote the I-language (the grammar or internal model of the language) of the speakers, using which they generate linguistic expressions or the E-language. Language is said to change, when the grammar G_{n+1} acquired by the learners is different from the grammar G_n of the teachers. However, we do not have a direct access to the internal grammars i.e. I-language of the speakers and we can hypothesize an event of language change by analyzing the E-language over a period of time. This leaves sufficient room for several contradictory explanations to co-exist that may explain the same historical data, but are hard to validate in general. We explain the different possibilities through Figure 1.

2.3

The E-language is directly accessible to all the users of a language, whereas the I-language of an individual is accessible only to that particular individual. Therefore, all communications take place through the E-language. In Figure 1, the vertical arrows represent learning and the T-shaped arrows represent communication between speakers through the E-language. The thick black arrows represent language acquisition by children. It is possible that this process is imperfect, leading to a different G_{n+1} (Andersen 1973). However, language change can be explained even if the process of language acquisition is assumed to be robust and perfect. The E-language (i.e. the trigger) from which the $n+1$ generation learns the language can be different from the E-language from which the nth generation learnt the language. The E-language can change due to three reasons:

1. The speakers of the nth generation produce an E-language that is slightly different from their I-language as far as the statistical distribution is concerned (Lightfoot 1991); this means the thick white arrows initiate the language change.
2. There is a contact with speakers of some other language (the dashed arrow) and the E-language observed by the children is a mixture of the outputs from both the grammars G_n and G'_n (Kroch and Taylor 1997).
3. Adults themselves learn from the new E-language, which can be an outcome of language contact or some other socio-cultural event. This is represented in the diagram by gray arrows.

2.4

Thus, corresponding to the four different types of arrows in the diagram, we get four basic possible causes of language change. In reality, the situation is often much more complicated, where several different causes interact at different levels of linguistic structures. Facts like children learn from other children and languages are associated with caste, pride and social hierarchies add further complexities (see Labov 1972 for an account of social theories of language change). It should be noted that although in Figure 1 the individuals of the same generation are shown to have the same I-language, synchronic variation is usually observed in all linguistic systems (Ohala 1989).
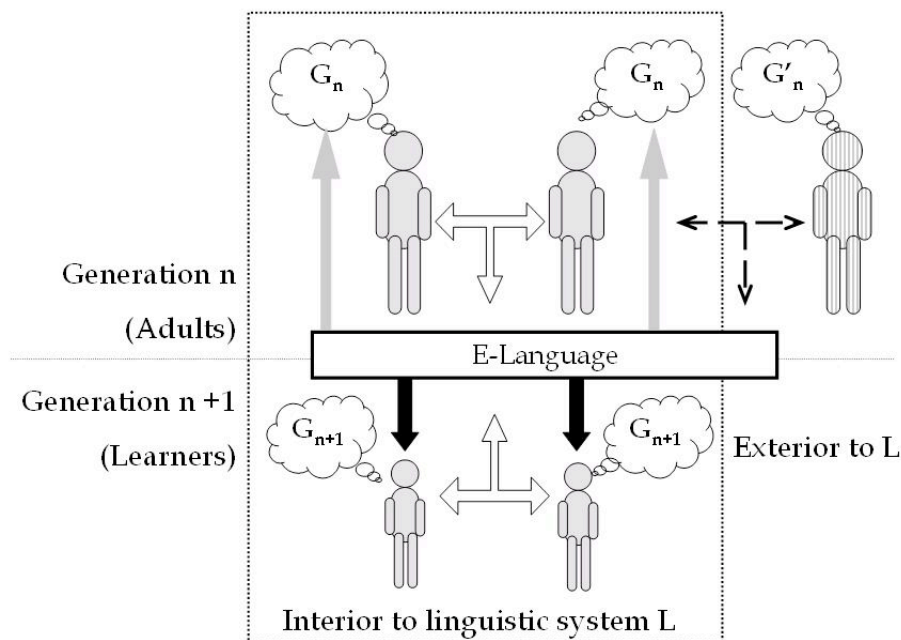
Figure 1. Language transmission in an open linguistic system. One or combination of several factors (shown by arrows in the diagram) can be responsible for a language change – the case where the grammar G_{n+1} is different from G_n.

The Course of Language Change

2.5

Several independent lines of research suggest that language change often proceeds along an S-shaped trajectory, also known as the logistic curve (Bailey 1973; Weinreich et al 1968). Initially, one of the forms is stable and a competing form occurs rarely in the language. The frequency of occurrence of the competing form increases slowly in the beginning of the process. Then an exponential growth is observed over a period of few generations, at the end of which the older form is completely driven out by the new variant. Figure 2 shows the rise in the use of the auxiliary do in English over a period of three centuries, which exhibits the S-shape pattern. The S-shaped trajectory for language change has been independently confirmed for several cases such as the shift of the words from one tone class to another in the Chaozhou dialect of Chinese (Chen and Wang 1975), the loss of verb-second syntax in English (Kroch 1989), French (Fontaine 1985) and Spanish (Fontana 1993), etc. See Briscoe (2000) and Kroch (2001) for further discussions on this.
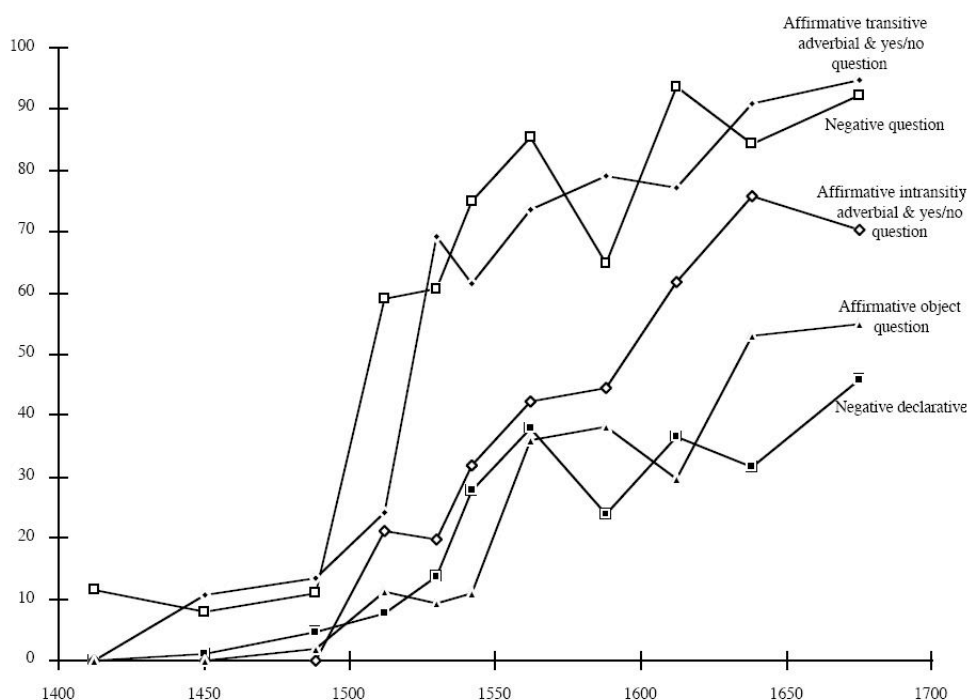
Some other observations on the course of language change include the *gradualness* both across the population and the lexicon (in case of sound changes), directionality etc (see [Bhat 2001](#) for an overview).

MAS Models of Language Change: A Review

2.6

The earliest examples of MAS modeling for language change goes back to 1960s, when Klein and his colleagues developed a general framework for Monte Carlo simulation of language change ([Klein 1966](#), [1974](#)) and demonstrated it on Tikopia and Maori languages ([Klein et al 1969](#)). These extremely detailed simulations tried to model every aspect of the concerned population, including the demographic distributions, social structures and interaction patterns. However, it was not until recently that the MAS models got into the mainstream research in diachronic linguistics, presumably due to the successful application of these models in a closely related domain of language evolution. Although the questions about language change are quite different from the questions about the origin and evolution of language, it is a well accepted fact that the two domains cannot be studied in complete isolation ([Parisi and Cangelosi 2002](#)). Moreover, the computational techniques and the models applied in the latter case can be adapted directly for the former with very little or no change at all.

2.7

There are several scholarly surveys on computational modeling of language evolution and origin ([Perfors 2002](#); [Steels 1997](#); [Wagner et al. 2003](#)). These surveys primarily focus on the simulation based models of language evolution, although some also discuss language change as a possible area where simulation models can be successfully applied (for example see [Christiansen and Dale 2003](#); [Wagner et al. 2003](#)). In a recent survey, Wang and Minett ([2005](#)) discuss both synthetic and analytical models of language evolution and change. They describe in depth a few works on dynamical system models of language change. Niyogi ([2006](#); [2002](#)) introduced a unified dynamical system formalism for language change and has analyzed different works with respect to his model.

Basic Concepts in MAS

2.8

Typically, a MAS experiment is designed as a population of interacting agents. The structure of the agents depends on the objective of the experiment and is normally kept simple for computational tractability, as well as clarity. Depending on the hypothesis being tested, the agents might interact genetically (through mutation and reproduction), linguistically (through language games as in Steels [2001](#)) or both. MAS based models must respect the following criteria to yield realistic models ([Steels 1997](#))

- *The distributed system constraint*: MAS should be conceived as a distributed system, where no agent has a complete view of the behavior of all the other agents and neither any agent has the power to control the behaviors of all others.
- *The limited rationality constraint*: Agents should not have direct access to each other's internal states. They can exchange information through interactions or inspection and speculation.
- *The open system constraint*: There should be an influx and outflux of agents in the population and an influx and outflux of meaning. Not only does this model the realism with respect to human populations and cultures, but this is also necessary to explain partly what kind of structures will emerge and why there is an evolution in the structure.

2.9

Wagner et al ([2003](#)) suggested a classification of MAS models based on two independent features — *situated-ness* and *structured-ness*. In a *situated* simulation the agents have a causal connection with their environment or the 'artificial world'; whereas in a *nonsituated* simulation,

the agents do not have direct or indirect interaction with the environment or other agents except for linguistic communication. In a *structured* communication, the agents send/receive signals which are composed of sequentially structured smaller units such as syllables, words etc. On the other hand in a *non-structured* communication, the signal can be considered as a single unit with no structural information. Thus, according to this classification, any MAS can be placed into one of the following categories:

- nonsituated & unstructured, as in the models of de Boer (2001); Livingstone & Fyfe (1999)
- nonsituated & structured, as in the models of Hare and Elman (1995); Oudeyer (2005)
- situated & unstructured, as in Oudeyer (1999) and Smith (2005)
- situated & structured, as in Cangelosi and Parisi (2001)

As we shall see shortly, MAS for language change are usually nonsituated, but can be both structured and unstructured.

2.10

Turner (2002) presents a different classification of the MAS systems on the basis of the structural details of the experiments across three dimensions: (1) agent representation, (2) agent interaction and knowledge acquisition, and (3) evolution. Recall that for explaining the causal forces behind language change one must explain the interaction patterns of the speakers and the language acquisition process (the arrows in Figure 1). Going by this, we have three different MAS models

- *The iterated learning model or the vertical model*: Normally there is one teacher and one learner. The information flows vertically as the learner acquires the teacher's language. Thus, only the thick black arrows (Figure 1) are modeled. See Kirby (2001); Hare and Elman (1995).
- *The horizontal model*: A population of agents tries to communicate with each other and learn from the success and failures. In some models the agents might age and die (removed from the system). Thus, the white thick arrows and the gray arrows are modeled (Figure 1). See de Boer (2001); Livingstone & Fyfe (1999).
- *The hybrid model*: Children learn from their teachers, and fellow learners. Adults learn from other adults, but at a much slower rate. Thus, it models the white, black and gray arrows (Bodik 2003; Smith and Hurford 2003).

Examples

2.11

We describe here three representative examples of MAS based models of language change, two of which have been developed to explain cases of real language change, and the third one tries to explain certain general issues. The examples have been chosen arbitrarily and not on the basis of their significance or popularity.

2.12

Dras, Harrison and Kapicioglu (Dras et al 2002; Harrison et al 2002) developed a MAS model to capture Turkic vowel harmony that is observed along the tongue height and tongue backness dimensions. The agents were capable of learning new words, mispronouncing or mishearing a word, making errors in favor of harmony etc. They reported an S-shaped pattern of language change with respect to harmony patterns in the simulation environment, under certain parameter settings.

2.13

Hare and Elman (1995) put forward a connectionist model of language change, where an artificial neural network was trained with a language data and the output of this network was used to teach a second network and the process continued. The duo tried to model the change in English verb morphology and came up with results consistent with the observed historical development. Although this is not a MAS model, the flow of information from the teacher to the learner is comparable to the vertical model or the iterated learning model (Kirby 2001) used in MAS.

2.14

Livingstone and Fyfe (1999) studied the nature of synchronic variation and dialect diversity using the MAS model. Unlike the previous two works, which attempted to model realistic language change, the aim of Livingstone and Fyfe was to identify the factors leading to dialect diversity. They enhanced de Boer's horizontal MAS model (where the agents play imitation games) with the concept of spatial locality and showed that distinct dialects emerge in different localities. Their work is significant because it obviated the need for sociological theories of dialect diversity.

Schwa Deletion in Hindi

2.15

Hindi is a language of the Aryan or Indo-Iranian branch of the Indo-European family of languages spoken in north and central India. It has its root in Prakrit, which evolved from old Sanskrit and thus, a large part of Hindi vocabulary has been borrowed from Sanskrit. Hindi is written from left to right using the *Devanagari* script, where all the vowels are marked overtly around the consonant using diacritical marks except for the inherent vowel *a*. However, this inherent vowel *a* (also known as the *schwa*) is often deleted during pronunciation — a phenomenon referred to as *schwa deletion* in linguistics. For example, the word *saradAra* (chief or head) is pronounced as /sar-dAr/ deleting the second and final *a*. Schwa deletion is an important issue in Hindi phonology that must be addressed while designing grapheme-to-phoneme converter for Hindi speech synthesizers (Kishore and Black 2003; Choudhury and Basu 2002; Choudhury *et al* 2004).

2.16

The context for schwa deletion in Hindi has been summarized by Ohala (1983) based on empirical observations as follows.

$$\alpha \to \Phi \ / \ VCC? \ \_\_ \ CV$$

Condition: Deletion should not violate the phonotactic constraints.
Convention: The rule applies from left to right.

2.17

The above rule says[3] that a schwa is deleted (maps to null) if it is preceded by VC or VCC and followed by CV, where V and C stand for vowels and consonants respectively. Table 1 shows some of the Hindi words and their possible schwa deletion patterns based on Ohala's rule. The second column shows the possibilities when just the rewrite rule is followed. However, out of these possibilities, some are discarded based on the convention, for which Ohala did not give any linguistic explanation. The remaining possibilities are shown in the third column. Interestingly, this rule does not capture word-final schwa deletion in Hindi, which must be considered before application of Ohala's rule (Narasimhan et al 2004). Therefore, we also allow deletion word finally, i.e. in a context like VCC? __ $, where $ stands for the end of a word. The schwa deletion pattern also interacts with the morphology of the word and the rule stated above is for mono-morphemic words only. The treatment of morphology is out of the scope of the current work.

Table 1: Schwa deletion patterns in Hindi and the application of Ohala's rules. The ITRANS encoding has been used. '-' represents syllable breaks

Written form	Possibilities after Ohala's rule	Possibilities after application of the convention
bachapana	*bach-pan, ba-chap-na*	*bach-pan*
bachapanA	*ba-chap-nA, bach-pa-nA*	*bach-pa-nA*
sundarI	*sun-drI*	*sun-drI*

A Constrained Optimization Model

2.18

That schwa deletion is a diachronic phenomenon is well attested in the literature ([Misra 1967](#)). The deletion of a vowel leads to the reduction in the number of syllables in a word (e.g. /*sa-ra-dA-ra*/ has 4, whereas /*sar-dAr*/ has only two syllables), and thus reduces the duration and thereby the effort of articulation. Therefore, the schwas that were originally present in Sanskrit words were deleted during casual and fast speech ([Ohala 1983](#)). The deletions that neither affect the perception, nor the proper identification of a word were acceptable and they replaced the older forms of the words where all the schwas were present. This is a functionalist explanation of the phenomenon of schwa deletion and thus can be modeled as an optimization problem, where the speaker wants to minimize the effort by deleting as many schwas as possible, but the listener accepts only those forms which are not confused with other words.

2.19

Choudhury, Basu and Sarkar ([2004](#)) have shown that the schwa deletion pattern in Hindi can be modeled as a multi-objective constrained optimization problem. To formulate a single objective function for optimization, they modeled other optimization criteria as constraints. Thus, ease of articulation led to the constraint that the phonotactics of the language should not be violated (which is also the precondition in Ohala's rule), and ease of perception was modeled as a constraint, which disallowed formation of any syllable in the schwa-deleted word that has an onset distinct from the syllable onsets in the undeleted form. The optimization criterion was syllable minimization, which implies deletion of as many schwas as possible. Choudhury et al showed that this optimization model entails Ohala's rule for schwa deletion in Hindi, except for the left to right convention that remained unexplained.

The Multi-agent Framework

3.1

In this section, we describe the MAS framework developed for modeling language change. The MAS setup consists of a population of *linguistic agents*, which interact with each other through *language games*. The framework, although generic, has been designed keeping in mind the experiments to be conducted and therefore certain features of the agent model are kept at the bare minimum to model only relevant phonological processes. We would like to emphasize however, that this is not a handicap and the framework can be extended to capture other phonological as well as syntactic phenomena.
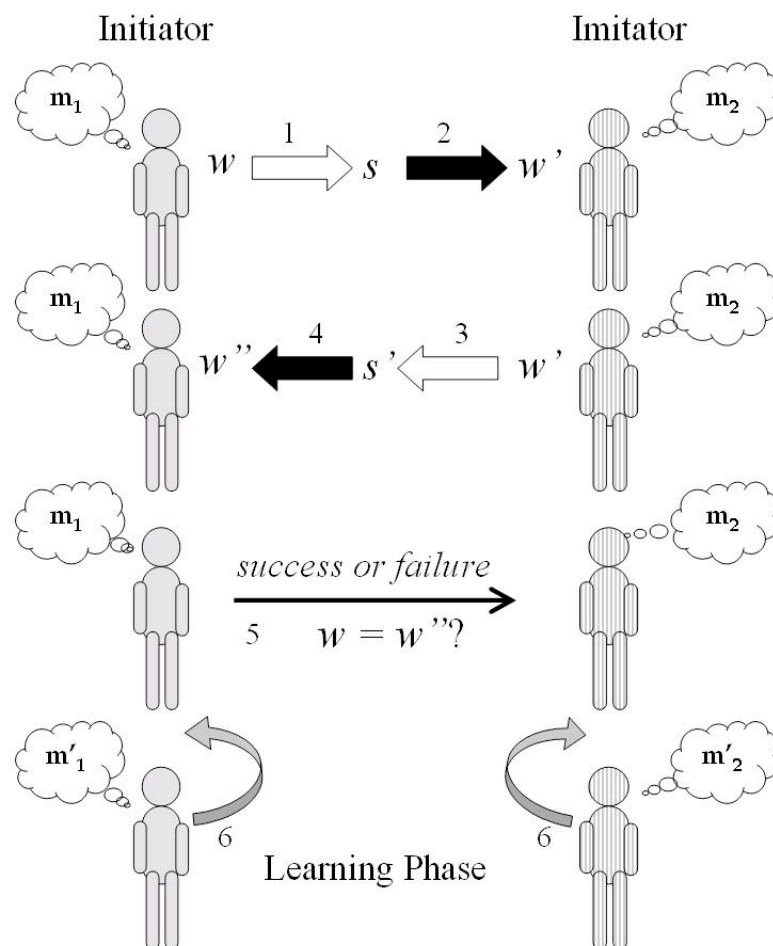
Figure 3. Schematic representation of an *imitation game*. The arrows represent events, which are numbered according to their occurrence and oriented according to the direction of information flow. The thick white and black arrows represent the process of articulation and perception respectively. The thin black arrow represents extra-linguistic communication and the gray arrows represent learning.

Imitation Games

3.2

Imitation games ([de Boer 2001](#)), a special type of language game, are played by two linguistic agents. The agents are identical in every respect except for, possibly, their language models. The basic framework of the current work is similar to the imitation game model, which is schematically represented in Figure 3. Nevertheless the current work differs significantly in the details of the linguistic agent described subsequently. The two agents playing the imitation game are known as the *initiator* and the *imitator*. The *initiator* chooses a particular linguistic unit w (a phoneme, word, or a sentence depending on the objective of the simulation) and generates a signal s corresponding to w using its *articulator model* **A** (described [below](#)). The imitator perceives the signal and tries to map it to some valid linguistic unit w' in its language model using the perceptor model **P** (described [below](#)). It then produces a signal s' corresponding to w', which then reaches back to the *initiator*. The *initiator* now tries to map s' to some linguistic unit w'' in its language model using **P**. If w (the original message) is same as w'' (the perceived message after the imitation game), then the game is considered to be *successful*; otherwise it is a *failure*. The initiator conveys this information to the imitator extra-linguistically. The agents then update their language models based on the result of the language game and its past history.

3.3

At this point, it might be useful to correlate this model with the existing hypotheses regarding language change and scrutinize some of the assumptions made here, which we summarize below.

- Recall that this is a horizontal model of language change ([2.8](#)), where we do not make any distinction between the adult and the learning phases.

- We do not model the noise in the communication channel. In other words, the signal *s* produced by the initiator reaches the *imitator* without any distortion. This is hardly the case in reality, where noisy communication channel introduces further complications into the system and demands robust communication pattern. However, this calls for a realistic noise model to be incorporated, which is beyond the scope of the current work.
- The extra-linguistic communication of the success of a game (step 5 in Figure 3) is a simplification in the sense that in reality when an uttered word is unclear to the listener, often he/she prompts the speaker to repeat, or wrong interpretations might go unnoticed as well. The detailed modeling of interaction patterns between individuals is too complicated to be considered in this work, but we do mention here that recent investigations (Smith 2005) have shown that successful communication can emerge in a system even without this assumption.
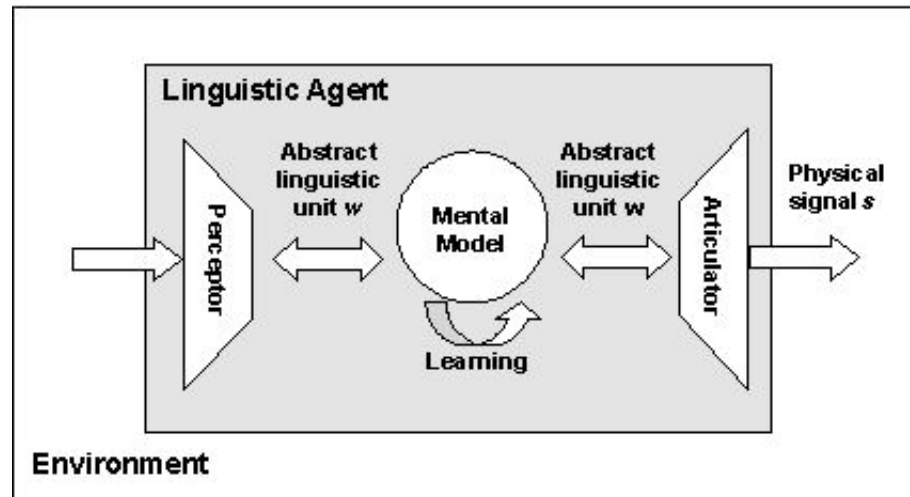

Figure 4. The architecture of a linguistic agent

Linguistic Agent Model

3.4

An agent (Russell and Norvig 2003) is composed of a *sensor-actuator* system, where the sensor helps the agent to get inputs from the environment and the actuator helps it to change the environment by some action. The agent also has a central control system that *decides* its actions based on the inputs and helps it adapt to its environment through appropriate learning. A *linguistic agent* is a special type of agent that acts in a linguistic environment. Figure 4 shows the block diagram of a linguistic agent. Formally, a linguistic agent **LA** is defined as a 4-tuple

$$\text{LA}: <\text{M, A, P, L}>$$

where

  M: Mental model of the agent (a set of mental states)
  A: The articulator model (or actuator)
  P: The perceptor model (or sensor)
  L: $\text{M} \rightarrow \text{M}$ is the learning algorithm that maps a mental state to another mental state

3.5

Ideally, in a MAS based model of language change, one would like to model **A**, **P**, **M** and **L** as close to human articulatory, perceptual, language representation and language acquisition mechanisms as possible. However, this is impossible partly because of the complexities of these phenomena and partly because of our incomplete knowledge of these faculties. Therefore, several simplified assumptions are made about each of the components of **LA** so that the MAS becomes practically realizable and at the same time the model remains realistic and powerful enough to facilitate the emergence of the desired linguistic features. Moreover, simplification is often desirable, as it makes the model transparent facilitating the study of the cause and effect relationships between the observed and the modeled.

Mental Model

3.6

Mental model **M** is defined as the set of all possible mental states. At any particular instant, an agent adheres to one and only one of these mental states. Let m_i and m_j be two distinct mental states belonging to **M**. An agent can change its mental state from m_i to m_j through *learning*. Therefore, a learning mechanism **L** can be conceived as a function from the set **M** to itself that maps a mental state m_i to another mental state m_j. **L** also depends on other parameters like the past history of an agent (in terms of success in communication) and the outcome of the most recent language game. **L** can be considered as a part of the mental model as well, but for the simplicity of presentation, we define **L** separately. Here, we describe **M** for linguistic agents that share a common vocabulary, but possibly different surface forms (pronunciations) of the words.

3.7

Let **Z** be a finite set of alphabet corresponding to the phonemes (sound units) in a language. A word w is a finite string of phonemes. Let **W** be the set of words in a language. Realization of a word w, represented by $r(w)$, is a string of 2–tuples $<p_i, t_i>$, for $1 \leq i \leq n$, such that $w = p_1 p_2 \ldots p_n$, (i.e. the string of phonemes) and $t_i \in [0, 2]$ represents the duration of the phoneme p_i in its realization in some abstract unit. A realization of **W**, represented as $r(\mathbf{W})$, is obtained by replacing each element of **W** (say w), by a realization $r(w)$. A mental state m_i consists of a realization of **W**, which we shall represent as $r_i(\mathbf{W})$ or simply **R**$_i$. Below, we illustrate this using a concrete example from Hindi.

3.8

Example 1: **Z** = {a, A, …, k, m, r, h, …} is the set of all Hindi phonemes. **W** is the set of all Hindi words, but for the purpose of illustration let us define **W** as {*hara*, *mara*, *amara*, *krama*}, a set of four words. A specific realization $r(hara)$ of the word *hara* looks like $<h,1>$ $<a,2>$ $<r,1.5>$ $<a,0>$, which means the word *hara* will be pronounced as *har*, where the first *a* will be long (duration 2), and the last *a* will be deleted (duration 0). The durations of the consonants *h* and *r* can be interpreted likewise. A typical mental state m_i will be comprised of one realization for each of the words in **W**, as given below (note that this realization is not representative of standard Hindi pronunciations).

$r_i(hara)$ = $<h,1>$ $<a,2>$ $<r,1.5>$ $<a,0>$,
$r_i(mara)$ = $<m,1>$ $<a,2>$ $<r,1.5>$ $<a,2>$,
$r_i(amara)$ = $<a,2>$ $<m,1>$ $<a,1.3>$ $<r,1.5>$ $<a,0.5>$,
$r_i(krama)$ = $<k,0.5>$ $<r,1.5>$ $<a,1.6>$ $<m,0.5>$ $<a,0>$

This particular state can also be represented as $r_i(\mathbf{W})$ or simply **R**$_i$. We can have another mental state m_j (also $r_j(\mathbf{W})$ or **R**$_j$) that looks like

$r_j(hara)$ = $<h,1>$ $<a,2>$ $<r,1.5>$ $<a,0>$,
$r_j(mara)$ = $<m,1>$ $<a,2>$ $<r,1.5>$ $<a,0>$,
$r_j(amara)$ = $<a,2>$ $<m,1>$ $<a,1.3>$ $<r,1.5>$ $<a,0.5>$,
$r_j(krama)$ = $<k,0.5>$ $<r,1.5>$ $<a,1.6>$ $<m,0.5>$ $<a,0>$

An agent can reach mental state m_i from a mental state m_j by learning to delete the schwa at the end of the word *mara*, thus reducing the duration of the word final *a* from 2 (long) to 0 (complete deletion). Note that as we allow the durations to assume any arbitrary value between 0 and 2, it leads to the possibility of an infinite number of mental states. However, by restricting the durations to a finite set of values (say 0 for deleted, 1 for short and 2 for long), we can restrict the number of mental models to a finite value (for this example, with 3 possible values for duration we have 3^{19} possible mental states that comprise **M**).

3.9

According to this definition of mental model, the agents share a common language represented by the universal set of words **W**. However, there exists variation in pronunciation among the agents, which is represented through the mental state the agent is in. This type of mental model also implies that the agents remember the pronunciation of each of the words by listing them separately, rather then learning a set of general phonological rules. Although this sounds

counterintuitive, there are at least two reasons for which this choice makes our model more general. First, by defining the type of phonological rules that an agent may learn we provide a bound on the possible types of variation in the pronunciations within an agent and between the agents. Second, in the absence of a complete knowledge about the phonological representation in human brain, it is better to avoid any inherent bias in the system towards any particular kind of rules. However, a possible disadvantage of non–generalization is that the agents have to learn the pronunciation of each of the words individually, which makes the convergence slow.

3.10

Apart from the realizations of the words, a mental state also consists of some memory where an agent can store its past experiences. In our case, the agents store the number of games played previously and the number of games which were successful. In other words, the agents remember how many times in the past they have been successful in communication, but they do not remember each of the games individually.

Articulator Model

3.11

The articulator model $\mathbf{A}$ is a procedure that maps a linguistic unit $r_i(w)$ from the mental state $\mathbf{R}_i$ of an agent to a physical signal s. We define $\mathbf{A}$ as a procedure and not a function, because there is an element of randomness in the mapping, such that the same unit $r_i(w)$ can be mapped to different physical signals at different time. The randomness models the imperfection of the human articulator that results in errors during speech production. For representational convenience, we will denote a signal s generated for a unit $r_i(w)$ as $\mathbf{A}(r_i(w))$ or simply $\mathbf{A_i}(w)$. Note that although $\mathbf{A}$ is identical for all agents, the articulatory behavior of an agent also depends on its current mental state and therefore, can be different for different agents.

3.12

The signal s is represented as a string of phonemes and phoneme–to–phoneme transitions, tagged with the corresponding durations in some abstract unit. So for a word $w = p_1p_2...p_n$, $\mathbf{A_i}(w)$ is a string of tuples $<q_j, u_j>$, where j varies from 1 to $2n - 1$. Here, q_j represents the phoneme $p_{\lceil j/2 \rceil}$ when j is odd and the transition from $p_{\lceil j/2 \rceil}$ to $p_{\lceil (j+1)/2 \rceil}$ when j is even. u_j represents the duration of the corresponding phoneme or transition. We also make the assumption that the duration of the consonants is 0. Only vowels and phoneme–to–phoneme transitions have non–zero durations. The following example illustrates the representation of a signal.

3.13

Example 2: Let w be *amara* (refer to Example 1 above). If the agent is currently in the mental state m_i (i.e. $\mathbf{R}_i$), then the corresponding realization of w (according to Example 1) is
$$r_i(amara) = <a,2> <m,1> <a,1.3> <r,1.5> <a,0.5>$$
The signal s corresponding to *amara* has 9 units (q_is) represented as follows:
$$a, a\text{-}m, m, m\text{-}a, a, a\text{-}r, r, r\text{-}a, a$$
Here, *x-y* represents the transition from phoneme *x* to *y*. The complete representation of the signal also includes the corresponding durations of the individual units, the duration of the consonants being 0. Therefore, a possible signal s generated for $r_i(w)$ has the following nature.
$$<a,1.8> < a\text{-}m,0.9> <m,0> <m\text{-}a,0.65> <a,1.3> <a\text{-}r,0.65> <r,0> <r\text{-}a,0.15> <a,0.3>$$
The mapping $\mathbf{A}$ takes place as follows. Suppose the word to be articulated is $r_i(w) = <p_1,t_1><p_2,t_2>...<p_n, t_n>$. Initially, the string $s = <q_1,u_1><q_2,u_2>...<q_{2n-1},u_{2n-1}>$ is generated, where each q_j is defined as above and the u_j are all 0. With respect to example 2, this corresponds to the string
$$<a,0> < a\text{-}m,0> <m,0> <m\text{-}a,0> <a,0> <a\text{-}r,0> <r,0> <r\text{-}a,0> <a,0>$$
Next, for all odd js that represent phonemes, u_j is assigned a value $t_{\lceil j/2 \rceil} + \mathbf{A_\epsilon}(t_{\lceil j/2 \rceil})$ if u_j is a vowel. The term $\mathbf{A_\epsilon}(t_{\lceil j/2 \rceil})$ is a small random perturbation that represents the articulatory error model. In the context of schwa deletion, the duration of each of the schwas in the realization is reduced with a probability p_r by a fixed amount d from the duration stored in the current mental state of the *speaker*. The durations of other vowels are kept unchanged. Suppose d is chosen to

be 0.2 and p_r is 0.4, then with probability 0.4 we reduce the duration of the schwas by 0.2. In the context of example 2, the duration of the first schwa is 2, which is reduced by 0.2, the duration of second schwa is 1.3, which is not reduced and the duration of the last schwa is again reduced by 0.2, leading to the following possibility.

<a,1.8> < a-m,0> <m,0> <m-a,0> <a,1.3> <a-r,0> <r,0> <r-a,0 > <a,0.3>

The duration of the transitions (i.e. u_j, when j is even) is initialized based on the durations of the neighboring vowels. Each transition is assigned half the duration of the adjacent vowel, when the vowel follows immediately, else it is assigned a further scaled down value. Thus, in our example, we get the following duration pattern:

<a,1.8> < a-m,0.9> <m,0> <m-a,0.65> <a,1.3> <a-r,0.65> <r,0> <r-a,0.15> <a,0.3>

3.14

Several simplifying assumptions have been made while designing the articulator model. We discuss each of them and provide motivation and justification for making such assumptions.

Assumption 1: The signal is represented as a sequence of phonemes and phoneme–to–phoneme transitions.
Justification: Analysis of speech signals show that there are steady states corresponding to the phonemes (especially the vowels and other sonorants), and between two phonemes there is a significant portion of the signal that represents phoneme–to–phoneme transition. This is a result of co–articulation and provides important cues for perception. Several concatenative speech synthesis systems utilize this fact and use diphones as the basic unit of synthesis (see Dutoit 1997 for an overview of such systems). This has been chosen as the representation scheme, because considering our objective, which is phone deletion, lower–level representations (for example formant–based as used in de Boer 2001) make the system computationally intensive without providing us any extra representational power. On the other hand, a syllable–level representation, which can provide a useful abstraction, calls for a definition of syllabification — a quite controversial concept (see Jusczyk and Luce 2002 for a general overview on issues related to human perception).

Assumption 2: Articulatory model has an inherent bias for schwa deletion, but not deletion of other vowels or consonants.
Justification: During casual speech, several articulatory phenomena are observed including vowel and consonant deletion, epenthesis, metathesis, assimilation and dissimilation. We refer to these as *articulatory errors* because such effects are unintentional, involuntary, and above all lead to the deviation from the correct pronunciation[4]. Nevertheless, the objective of the current work is to investigate the schwa deletion pattern and not general vowel deletion, or other types of sound changes. Incorporating these extra factors can further complicate the model leading to masking and interference of different factors. Moreover, we do not make any claims here regarding the emergence of the schwa deletion; we only claim that the model explains the specific pattern observed in Hindi schwa deletion. The investigation of the emergence of schwa deletion and other phonological phenomena is beyond the scope of the current work.

Assumption 3: Duration of consonants is 0.
Justification: Vowels and consonants can be placed on the *sonority hierarchy* (Clements 1990), which reflects their ability to form syllables and the possibility to be pronounced for a prolonged period. Thus, vowels placed at the top of the sonority scale almost always form the syllable nucleus and have a longer duration, whereas stops placed at the lower end of the scale can never form syllable nucleus and can never be lengthened. Nonetheless, several consonants, especially the sonorants can be lengthened as well as used as the syllable nucleus. Thus, the assumption that all the consonants have 0 duration is clearly incorrect. However, let us try to understand the objective of the current work. We want to model the deletion of the schwas. When a schwa gets deleted, the consonants which were part of the syllable have now to be placed within other syllables. For example, if both the schwas of the word *ha-ra* (- indicates syllable boundary) are deleted the resulting word *hr* is not pronounceable, whereas deletion of only one of the schwas gives rise to the patterns *hra* or *har* both of which are well formed. If we assume that the consonants have durations of their own, and can be perceived even without the transitions, in our model we have no way to claim that *hr* is unpronounceable. To the contrary, the assumption that consonants have no duration allows us to model the syllables around the vowels, even though there is no explicit reference to the syllables. In fact by assigning the

duration of the transitions on the basis of the neighboring vowel duration, we capture the fact that a syllable, including its onset and rime, is perceptible only if it is of sufficiently large duration (see below for details on the perceptor model).

Assumption 4: The duration of the schwas are reduced randomly, without considering the context.
Justification: The inherent bias towards fast speech is modeled through the tendency to reduce the duration of the schwas, but the articulator model does not accomplish this randomly. The duration is reduced by a fixed and predetermined amount d (which is an input to the simulation experiments) and a probability p_r (also an input). The randomness is with respect to the context in which the schwa is deleted. Stating it in another way, all the schwas in a word are equally likely to be deleted. This is a desired feature because we want to examine the emergence of the schwa deletion context and therefore, should refrain from providing any initial bias in the system towards deletion in certain contexts and not in others. We shall see that even without any initial context specific bias, the context for schwa deletion clearly emerges in the simulation experiments.

Perceptor Model

3.15

The perceptor model **P** maps a signal s to a word w in **W**. The perceptual mechanism can be divided into two distinct parts — *perception* and *cognition*. The former refers to the identification of the individual phonemes in the input signal s and the latter refers to the mapping of the string of phonemes so identified to a word w in the mental lexicon $\mathbf{R}_i$ of an agent. Although these two actions might proceed hand in hand in human beings, separating them out simplifies **P**. These two modules can be identified with the acoustic and the pronunciation models of automatic speech recognition systems (Jurafsky and Martin 2000, chapter 7), on which we do not elaborate here further. Given a signal $s = <q_1, u_1><q_2, u_2>…<q_n, u_n>$, the procedure **P** tries to perceive a phoneme p_i either from its realization $<q_{2i-1}, u_{2i-1}>$ or from the transitions $<q_{2i-2}, u_{2i-2}>$ or $<q_{2i}, u_{2i}>$. The probability of perception of a phoneme from a unit q_j depends on its duration u_j and also the neighboring phoneme for transitions. As u_j increases from 0 to 2, the probability also increases linearly from 0 to 1 according to the following equation.

$$\text{Prob}(p_j \text{ is correctly perceived from realization } q_j) = u_j / 2$$

3.16

Since the consonants are assigned a duration of 0, therefore, they can be perceived only from the transitions. In case of transitions, the probability of perception also depends on the relative *sonority* of the two phonemes, which we do not model here. However, since the transitions are assigned half the duration of the neighboring phonemes, we relax the probability for transition perception as follows.

$$\text{Prob}(p_j \text{ is correctly perceived from transition } q_j) = u_j$$

If a phoneme is correctly perceived from any of the three units (two neighboring transitions and the phoneme itself) than it is considered to be perceived, otherwise it is assumed that the listener has not heard the phoneme. In the current model, a phoneme that is not perceived correctly is assumed to be deleted (i.e. replaced by *nothing*). Once all the units in s have been analyzed, the complete string v of perceived phonemes is obtained. Below, we illustrate this process by an example.

3.17

Example 3: Let the signal s corresponding to the word $w = amara$ be (taken from Example 2)
   <a,1.8> < a−m,0.9> <m,0> <m−a,0.65> <a,1.3> <a−r,0.65> <r,0> <r−a,0.15> <a,0.3>
Let us estimate the probabilities of perception of the phonemes based on their realizations. For this, we first calculate the probabilities that the phoneme is not perceived from its realization and any of the neighboring transitions. We multiply these probabilities to identify the probability that the phoneme is not perceived from any of them. We subtract this quantity from 1 to get the actual perception probability of the phoneme. Table 2 illustrates the computations.

Table 2: Computation of perception probabilities. *perc* denote the

probability of perceiving a phoneme from a given unit

Phoneme p	Left transition		Realization		Right transition		Prob (p is perceived)
	perc	~perc	perc	~perc	perc	~perc	
a	0	1	0.9	0.1	0.9	0.1	$1 - (1 \times 0.1 \times 0.1) = 0.99$
m	0.9	0.1	0	1	0.65	0.35	$1 - (0.1 \times 1 \times 0.35) = 0.96$
a	0.65	0.35	0.65	0.35	0.65	0.35	$1 - (0.35)^3$ $= 0.96$
r	0.65	0.35	0	1	0.15	0.85	$1 - (0.35 \times 1 \times 0.85) = 0.7$
a	0.15	0.85	0.15	0.85	0	1	$1 - (0.85 \times 0.85 \times 1) = 0.23$

3.18

Thus, we see that for the given s, the first three phonemes are almost always perceived whereas the last is hardly ever perceived by any agent. The phoneme r is usually perceived. Therefore, the probability that v, the complete string of perceived phonemes, is *amar* can be computed by multiplying the probabilities of perceiving the first four phonemes (i.e. 0.99, 0.96, 0.96 and 0.7 respectively) and not perceiving the last a (i.e. $1 - 0.23 = 0.67$). This amounts to 0.49, which is the highest for all possible strings that could be perceived from the given s. Likewise, v is *mar* with probability 0.005, and *ra* with probability 3×10^{-7} (the least probable case). The perceived string of phonemes v might not be a valid word in **W**. The next task, therefore, is to map v to the *nearest* word w in **W** (the cognition step). This is accomplished by comparing v with realization of every word $r_i(w)$ in $\mathbf{R}_i$. This model of cognition has been adapted from (Boer 2002). A score is calculated based on the *minimum edit distance* (Jurafsky and Martin 2000, p 156) of w and v, keeping in mind the duration of the vowels in $r_i(w)$ as well. If a vowel has a short duration in $r_i(w)$ and it is deleted in v, the cost of the alignment is lower than the case when the vowel has a longer duration. The cost matrix used for calculating the minimum edit distance is given in Table 3. The word $w^* \in \mathbf{W}$ that has the lowest score (i.e. the word that is nearest to v) is then chosen as the output of the procedure **P**. If there are multiple words with the same minimum score, one of them is chosen at random. If the minimum score so obtained is larger than a threshold, then v is much different from any of the words in the agent's vocabulary. In such a case, the perception fails and no word is perceived corresponding to s. We illustrate the cognition process in Example 4.

Table 3: The cost matrix for calculating minimum edit distance between the perceived string of phoneme v and a word w in the mental lexicon. Φ stands for null or no phoneme. The case where schwa in w is aligned with nothing in v (last row, 2nd column) corresponds to a case of schwa deletion, which is penalized by t — the duration of the schwa according to the current mental state of the listener

	w: Phoneme p (other than schwa)	Schwa (Duration = t)	Φ
v			
Phoneme p (other then schwa)	0 (if match) 2 (otherwise)	2	2
Schwa	2	0	2
Φ	2	t	NA

3.19

Example 4: Let us consider an agent, whose current mental state is $\mathbf{R}_i$ given in Example 1. Let the perceived string v be *amar* (refer to Example 4). The minimum edit distance of v from *amara*

is calculated by finding out the best alignment, which in this case is

Table 4: Minimum distance alignment

w	a	m	a	r	a (0.5)
v	a	m	a	r	Φ
cost	0	0	0	0	0.5

Thus, the total cost of alignment between *amar* and *amara* is 0.5. The costs of alignment of *v* with the other words in R_i can be computed similarly. The results are displayed in Table 5. The scores are also displayed when calculated according to the mental state R_j (Example 1).

Table 5: Cost of alignment of *amar* with the different words in mental states R_i and R_j

Words	Cost of alignment	
	R_i	R_j
hara	4	4
mara	4	2
amara	0.5	0.5
krama	6	6

We observe that for both the mental states, the perceived string *amar* is mapped to the word *amara*, because this has the minimum cost of alignment. Stated differently, *amara* is the closest word to the string *amar*. However, if *v* was *mar* instead of *amar* (as figured out in example 3, this has a probability of 0.005), the perceived word would have been *mara*.

Assumption 5: A phoneme that has not been correctly perceived is assumed to be deleted. *Justification*: A phoneme that is not perceived correctly can be substituted for a similar sounding phoneme. For example, "par" can be heard as "bar", because /p/ and /b/ are similar sounding in the sense that both of them are labial stops. To incorporate this feature in our model, we need to define realistic phoneme–phoneme substitution probabilities, which are indeed considered while designing speech recognition systems. Firstly, this makes the perception model quite computationally intensive, increasing the simulation time significantly. Secondly, this reduces the chances of successful communication. Note that in reality the context (surrounding words) provides extensive clues for recognizing a word, which is completely absent in our model due to its limited scope. Thirdly, the only parameter considered here is phoneme and signal duration, which has a direct implication on deletion. The idea is not to deny the effect of a whole lot of other parameters on general human perception, but to focus specifically on the durational effects — which is arguably the most crucial factor in schwa deletion (Ohala 1983, Choudhury et al 2004).

Learning

3.20

Learning or *language acquisition* is the most crucial issue in MAS models and it is also the least understood one. There are several paradigms of learning e.g. rule generalization, neural networks, evolutionary algorithms etc. (see Cangelosi and Parisi 2002 and Niyogi 2006) for overviews of the different learning models), that can be modeled and compared. However, in our framework, we choose a very simple learning algorithm — learning from examples. The basic idea is as follows. An agent articulates different signals corresponding to a word in different language games. If a particular language game is successful, there is enough reason for the agent to believe that the articulated signal is well understood by the other agents and thus, the signal articulated is considered to be a successful example, which the agent remembers for future use. An agent is allowed to learn from the successful language games

only if it has been *quite successful* in its recent past, because a high failure rate indicates that the agent's model differs from most of the other agents in the population, implying that the apparent success of the recent language game might have been a result of random chance. The agents can similarly learn from their failures.

3.21

The steps and parameters involved in learning are summarized below.

1. Suppose the initiator articulates a signal s corresponding to a word w, which is perceived by the imitator as w' (which might be same as w). The imitator then articulates a signal s' corresponding to w', which is perceived by the initiator as w''. The game is successful if w and w'' are the same words.
2. If a game is successful, the initiator (imitator) might learn from the previous interaction by setting the duration of the phonemes of the word w (w') according to the durations in s (s').
3.
    - If the initiator's success in previous communications is greater than k_{sn}, then it sets the duration of the schwas corresponding to w in the current mental state to the duration of the realized schwas in s. But this is done only with a probability p_{sn}.
    - Similarly, if the imitator's success in previous communications is greater than k_{sm}, then it sets the duration of the schwas corresponding to w' in the current mental state to the duration of the realized schwas in s'. But this is done only with a probability p_{sm}.
4. If a game is a failure, the initiator (imitator) learns by increasing the duration with a probability p_{fn} (p_{fm}) only if its success rate is less than a threshold (or its failure rate is greater than a threshold, say k_{fm}).

The learning parameters are set to some predefined values at the beginning of a simulation experiment and are kept constant over a particular run of the simulation. Section [4](#) discusses the effect of one of these parameters on the emergent pattern.

The Simulation Set up

3.22

A population of N agents is initialized with identical mental states m_0, a realization $r(\mathbf{W})$, such that all the vowels have a duration of 2 units (the largest possible duration in the model). The initial state thus corresponds to the Sanskrit pronunciations, where all the schwas are pronounced. The simulation is continued for several rounds with R language games per round. At the end of each round the results of the simulation are manually checked to determine the convergence. There is also a provision to run a preset number of rounds without requiring manual intervention at the end of each round. The result file generated at the end of the simulation however records the mental states of the agents only at the end of each round. The simulation parameters, which include the initial lexicon, the deletion parameters d and p_r, and the learning parameters k_{sn}, k_{sm}, k_{fn}, p_{fm}, p_{fn}, p_{sm} and p_{sn} are specified in an input file. Also a *seed* is specified for the random number generator.

3.23

The steps in a language game are:

1. Two agents are selected from the population of N agents. One is given the status of *initiator* and the other the *imitator*.
2. The initiator chooses a word w at random from $\mathbf{W}$ and generates a signal s corresponding to w using the articulator model $\mathbf{A}$,
3. In the current model, the environment is assumed to be noise free and therefore, the imitator receives the same signal s.
4. Imitator uses the perceptor model $\mathbf{P}$ to map the signal s to a valid word in $\mathbf{W}$. Let the word perceived be w'.
5. Imitator generates a signal s' corresponding to w' using $\mathbf{A}$.
6. Initiator tries to perceive s' using $\mathbf{P}$. Let the perceived word be w''. If $w=w''$ then the game is successful, this message is conveyed to the imitator extra-linguistically.

7. Depending on the outcome of the game, both the initiator and the imitator may decide to learn (i.e. change there current mental states)
8. Finally, the agents update their mental states by registering the results of the last interaction as well as the learnt durations.

3.24

Before we present the simulation results, some of the assumptions and restrictions imposed on the model must be clearly articulated to avoid any confusion or wrong interpretation of the data. These are:

- *Closed system assumption*: In this model the agents do not leave or enter the system, neither do they age and die. Nevertheless, an open system with small emigration and immigration rates can be expected to show results similar to a closed system.
- *Simplistic agent model*: Articulatory, perceptory and learning mechanisms are too simple to model the reality. Agents can only reduce the duration of schwas and cannot make any other errors. Furthermore, it is assumed that the mental model consists of a set of words with possible realizations, rather than a general realization algorithm, which seems more plausible.
- *Simplistic interactions*: Words are uttered in isolation, which is seldom the case in reality. In fact much of the information for proper perception of a word comes from its context. All words are assumed to be equally probable, which is not true for real languages either.

3.25

Nonetheless, our claims here are: 1) the framework is extendible to overcome all of the above mentioned limitations, 2) The assumptions made are only to make the model simpler so that we can clearly study the effects of various parameter settings, and 3) most of assumptions have been made keeping in mind the area of study — schwa deletion in Hindi, rather than language change in general.

Experiments and Observations

4.1

Let us first enumerate the parameters that might affect the emergent schwa deletion pattern: 1) agent model, i.e. the learning, articulatory, perceptual mechanisms and the mental model, 2) vocabulary, 3) population size N, 4) the learning parameters like thresholds k_{sn} and k_{sm}, the learning probabilities p_{sn} and p_{sm} and 5) the deletion parameters p_r and d. The study of the effects of different agent models on the emergent pattern is out of the scope of this work. Also it has been found that the size of the population does not have any significant effect on the emerging pattern, it only determines the rate of convergence with larger time required for convergence for larger population. Below, we shall describe some of the significant observations of different experimental setups.

Vocabulary

4.2

The vocabulary or the lexicon **W** has an important impact on the emerging pattern. This is due to the fact that perception is based on the closest word in **W** corresponding to the given string of phonemes. Table 6a and 6b show two runs of the experiments under the same settings, except for the vocabulary. The apparent discrepancies in the emergent pronunciations in the case of 6a can be explained as follows. Since there is no word other than *amara* having the consonant *m*, identification of *m* itself allows identification of the word *amara*; similarly, *hra*, which can be often confused with *ra* is still perceived as *hara*, because the edit distance of *ra* to *hara* is less than that to *amara*. On the other hand, if **W** contains both *mara* and *amara*, deletion of the word initial *a* in *amara* is not preferred as it removes the distinction between the two words, resulting in a sharp decline in communication success. Therefore, presence of both the words helps in the emergence of the correct pattern.

Table 6: Two runs of the experiment with different lexica

Words	Emergent pronunciation	Correct pronunciation	Words	Emergent pronunciation	Correct pronunciation
hara	*hra*	*har*	*mara*	*mar*	*mar*
amara	*mar*	*amar*	*amara*	*amar*	*amar*
	Case 6a			**Case 6b**	

4.3

The strong effect of the **W** in the emerging pattern therefore, calls for the use of the complete Hindi lexicon during MAS experiments. We attempted such an experiment, but even after restricting the lexicon to the most frequent 8000 Hindi words extracted from a corpus, convergence demanded an astronomical simulation time. Typically, it has been observed that the number of games required for convergence in presence of |**W**| words in linear in |**W**|. The time required for one game increases linearly with |**W**| as during the perception step, the perceived string of phonemes v is compared with each of the words in **W** in terms of edit distance score. Therefore, the time to convergence can be shown to be approximately proportionate to the square of the size of **W**. Table 7 shows the time required for a few different parameter settings. Assuming that convergence requires 10M games per word (see below), the estimated time required for convergence for a 8000 word lexicon is 4×10^{15} seconds or 12×10^{7} years approximately (simulation is run on a Pentium 4 1.6GHz machine)!

4.4

Therefore, to nullify the effect of the vocabulary the experiments were conducted for a normalized lexicon, where only two consonants and two vowels were used in many possible combinations to generate the words, which resemble the structure of the real lexicon quite closely. One such lexicon is given below, for which we describe the rest of the experimental results.

$$\mathbf{W}_{normalized} = \{karaka, karakA, karAka, karAkA, kAraka, kArakA, kArAka \}$$

Here, k and r are placeholders for consonant (C), A is a placeholder for vowel (V) and a stands for schwa. We have chosen tri-syllabic words, where the syllables are of type CV or Ca. Therefore, we have 8 possibilities, out of which CVCVCV is uninteresting, since it does not have any schwa. The other 7 possibilities are considered in the normalized lexicon. Note that k and r and their order in the words are arbitrary choices, which do not affect the results described next.

Table 7: Time taken for simulation. The values reflect the real time and not the exact system time and are therefore dependent on system load. Machine specs: Pentium 4, 1.6 GHz

Vocabulary size	Games played	Time req. (in secs)	Time req. (in secs) per million games	Time req. (in secs) per million games per word
8	3M	290	96.67	12.08
8	2M	200	100	12.5
1800	55000	680	12363	6.87

Results for the Normalized Lexicon

4.5

The emergent pattern for the normalized lexicon for a specific run of the simulation has been presented in Table 8. Since the agents share the same lexicon **W**$_{normalized}$, the only aspect

where they vary is the realization of the lexicon. In other words, they may disagree only with respect to the duration of the schwas. Therefore, we list the duration of the schwas averaged over all the agents. We assume a schwa to be deleted if its duration is less than 0.67, and retained if it is greater than 1.33. We make this choice on the basis of the observation that schwas in Hindi can be long, short or deleted. Since the duration of the schwa can vary from 0 to 2, we divide the region into three equal length zones from 0 to 0.67, 0.67 to 1.33, and 1.33 to 2 representing the deleted, short and long schwas respectively. Based on this, we derive the emergent pronunciation of the population and compare it with the pronunciation in standard Hindi derived according to Ohala's rule.

Table 8: The observations averaged over the language model of all the agents for the normalized lexicon. The parameters for this typical experiment were: N = 4, $k_{sn} = k_{sm}$ = 0.7, p_{sn} = 0.6, p_{sm} =0.2, p_{fn} = 0.6, p_{fm} = 0.0, d = 0.01. Games required for convergence: 70 million

---

Words	Vowel duration (in order of occurrence in the word)	Emergent Pronunciation	Pronunciation in standard Hindi	Number of errors
karaka	1.99, 1.49, 0.00	*ka-rak*	*ka-rak*	0
karakA	2.00, 0.00, 2.00	*kar-kA*	*kar-ka*	0
karAka	2.00, 2.00, 0.00	*ka-rAk*	*ka-rAk*	0
karAkA	0.00, 2.00, 2.00	*krA-kA*	*ka-rA-kA*	1
kAraka	2.00, 1.99, 2.00	*kA-ra-ka*	*kA-rak*	1
kArakA	2.00, 0.50, 2.00	*kAr-kA*	*kAr-kA*	0
kArAka	2.00, 2.00, 0.00	*kA-rAk*	*kA-rAk*	0

4.6

There are two errors in the emerging pattern, one deletion error, where the schwa is actually retained, and one retention error, where the schwa is normally deleted. There were 12 schwas in the whole lexicon. Therefore, the emerging pronunciation shows 83.33% similarity to the actual pronunciation with respect to schwas and 71.4% similarity at the word level. Although the results clearly show that the model captures the evolution of the schwa deletion pattern in Hindi to a great extent, certain phenomena like the immunity to deletion of the schwa in the first syllable of the word have not been reflected in the emergent pattern (*karAkA* → *krA-kA*). We make two remarks on this issue: 1) there are languages like Punjabi, which feature deletion of schwas in the first syllable. This implies that the emergent pattern is not unnatural; and 2) Immunity to deletion in such cases might be a result of other features like stress patterns, which have not been captured in this model.

4.7

Figure 5 shows a plot of the duration of a particular schwa that was finally deleted (and correctly so) averaged over all the agents against the number of games played. The plot shows how over time the duration of the schwa reduced and finally dropped to zero. The transitions however are very sharp (spanning over less than 10000 games) spaced by significantly longer periods of stable intermediate pronunciations. Almost all the schwas that finally got deleted exhibit such a curve, which is also called the S–shaped curve (see [above](#)). Languages change over a very short period of time remaining stable for longer time scales. The fact that the MAS model also exhibits similar property is a further validation of its plausibility. A deeper scrutiny reveals that each of these drops correspond to the deletion of a specific schwa by a particular agent. In other words, there are four drops in Figure 5. The first one is observed when the first agent

dropped the schwa, the second one is observed when another agent drops the schwa and the average duration reduced from 1.5 to 1. Recall that in this experiment the population size was N = 4. However, it requires a much deeper analysis of the dynamic behavior of MAS in general and the present model in particular to explain why a particular agent drops the schwa by reducing its duration sharply over a thousand games and not gradually over a longer period overlapped with the deletion phases of the other agents. We omit any further discussion on this here.
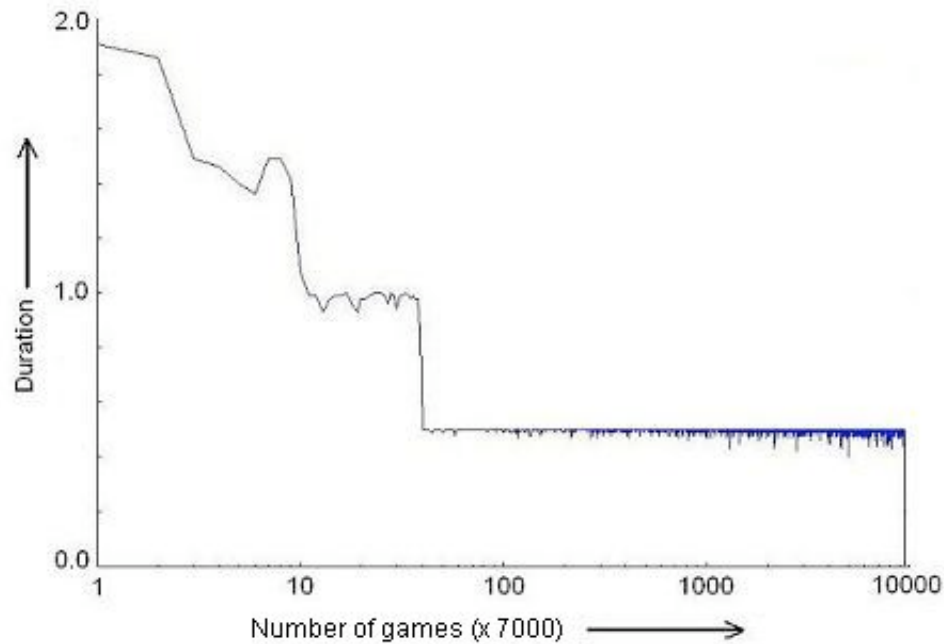

Figure 5. The average duration of a schwa vs. the number of games. The plot is for the final schwa of the word *karAka*. The number of games is 7000 times the number shown in the plot. The scale is logarithmic

Effects of other Parameters

4.8

The thresholds k_{sn} and k_{sm} that determine whether an agent will learn or not based on its average success rate in communication have a significant impact on the final communication success at convergence point as well as the emerging pattern. When these thresholds are set to 1.0, just after a few games, when all the agents have encountered some failure, they stop learning and therefore, the system stabilizes very early, and the system retains its initial pronunciation, i.e. no schwas are deleted. On the other hand if the threshold is set to 0.0 a successful game just by chance allows the agents to learn and hence almost all the schwas are deleted. The system takes a long time to stabilize, whereby the communication success falls drastically.

4.9

There are two parameters related to deletion — the duration reduction step d and the duration reduction probability p_r. The duration step parameter d has a strong influence over the emergent pattern. If it is very small, convergence is steady, but in such cases the deletion of successive schwas are often prohibited resulting in two short schwas. On the other hand very large d (>0.5) leads to proper schwa deletion patterns, but the population of agents seem to develop two distinct dialects, one following the left to right convention suggested by Ohala and another following the right to left convention. In fact, apart from the vocabulary, d and k_{sn} are the other two most influential parameters. Figure 6 and 7 illustrate how these two parameters govern the communication success rate and the average duration of the schwa in the simulation experiments. We observe that when k_{sn} is close to 1, the effect of d is negligible, but for smaller values of k_{sn} (less than or equal to 0.8), communicative success drops significantly for large d. This can be explained as follows. When the agents greedily reduce the duration (large d) without considering the communicative success (low k_{sn}), there is no global emergent pattern. In such a

case, every agent develops its own dialect (or more correctly *idiolect*), and the communicative success of the system falls. However, if the agents reduce the durations slowly (small *d*) or if they consider the communicative success while reducing the duration (high k_{sn}), a global pattern emerges leading to more successful communication. Thus, a non–greedy deletion strategy is a must for the emergence of a global pattern.
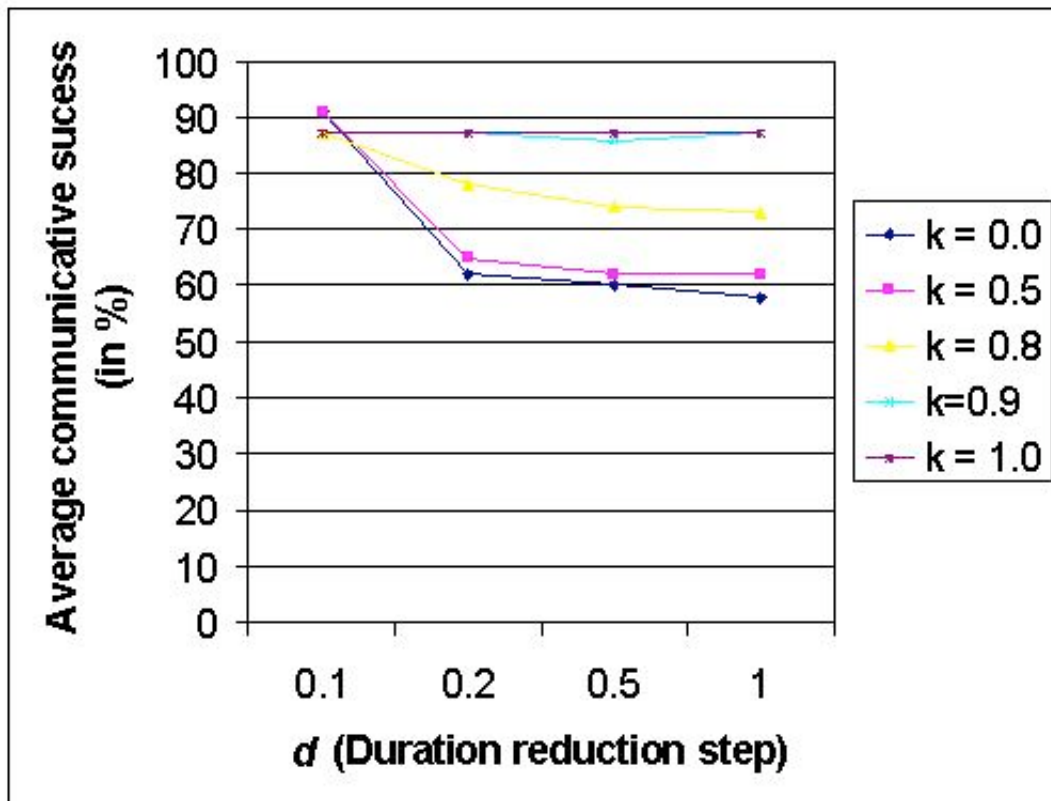

Figure 6. The dependence of average communication success rate on *d* (duration reduction step) and learning threshold k ($=k_{sn}= k_{sm}$). Other simulation parameters: vocabulary size = 7, N = 4, p_{sn} = 0.6, p_{sm} =0.2, p_{fn} = 0.6, p_{fm} = 0.0, number of games=300000

Figure 7. The dependence of average schwa duration on *d* (duration reduction step) and learning threshold k ($=k_{sn}= k_{sm}$). Other simulation parameters: vocabulary size = 7, N = 4, p_{sn} = 0.6, p_{sm} =0.2, p_{fn} = 0.6, p_{fm} = 0.0, number of games=300000. The expected duration
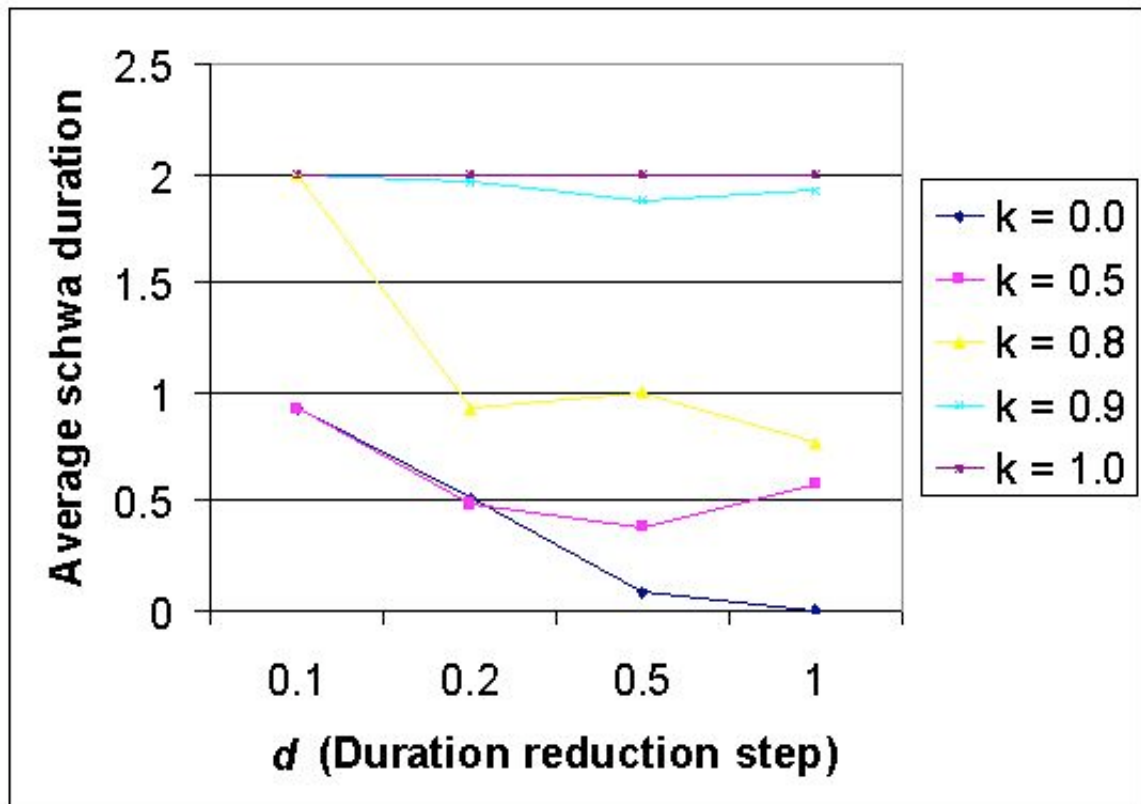
The d vs. average schwa duration curve (Figure 7) however presents a slightly different scenario. It is clear that when k is small (0.5 or below), all the schwas are deleted leading to complete communication failure (as reflected in Figure 6). However, when k is very close to 1, the system becomes too strict to allow schwa deletion and the original pronunciations are retained. Such a system has very high communicative success rate (as reflected in Figure 6), but fails to facilitate the emergence of schwa deletion. For moderate values of k (between 0.5 and 1), a schwa deletion pattern emerges that is closer to the one observed in Hindi.

Dialects and Synchronic Variation

4.10

The previous subsections discuss the average behavior of the MAS experiments, where the schwa durations were averaged over all the agents and/or all the schwas in the lexicon. A deeper look inside the mental states of individual agents reveals several other interesting facts. Although the observed mean schwa durations vary from 0 to 2, the schwa durations in the mental states of the agents are categorical in nature. A particular schwa has a duration of either 0 or 2. Very rarely an agent has a fractional duration for a schwa (2 out of 130 cases), but even when it does, the value is very close to one of the two extremes. Note that Figure 5 suggests something similar, where the agents show a sharp decline in the schwa duration over a very short period of time (measured in terms of games). Table 9 lists the different variants of the words that were observed in a particular simulation experiment. We make the following observations regarding the variants:

- The schwa is invariably deleted in the words with a single schwa. However, as discussed above, these results are also dependent on the vocabulary.
- All the emergent variants are unique. This is possibly the reason behind their coexistence, because the lexical distinctions are not affected by the presence of the different variants.
- The word *karaka* shows two variants – *karka* and *karak* pertaining to Ohala's rule applied from left to right and right to left respectively.
- Not all the variants that emerged are observed in Hindi or its dialects. For example, *krkA* is not a valid phoneme sequence in any of the Indo-Aryan languages, but it emerged as a variant of the word *karakA*. The realization of *krkA* and its perception will be limited to the monosyllabic word *kA*. Therefore in this case, the deletion of the two schwas implies the dropping of the first two syllables *ka* and *ra*.

Table 9: Different variants of a word that emerged during a simulation. The number of agents speaking that variant is given in the parentheses. Simulation parameters: N = 10, vocabulary size = 7, d = 0.1, p_{sn} = 0.6, p_{sm} =0.2, p_{fn} = 0.6, p_{fm} = 0.0, k_{sn}= k_{sn}= 0.9, number of games=3M

Words	Variants
kArakA	*kArkA* (10)
karakA	*krakA* (4), *krkA* (3), *karkA* (2), *karakA* (1)
kAraka	*kArk* (10)
karAkA	*krAkA* (10)
karaka	*karka* (6), *karak* (4)
karAka	*karAk* (5), *karAka* (3), *krAk* (1), *krAka* (1)
kArAka	*kArAk* (10)

Robustness and Convergence Issues

4.11

What happens when under the same parameter settings we run two different simulations with

different initial random seeds? Table 10 reports the average communicative success and the average schwa duration for 10 runs under the same simulation settings, except for different values of the initial random seed. We note that the average communicative success is nearly the same for the different runs, but the mean schwa duration is not and it takes certain specific values like 0.85 (2 runs), 0.92 (3 runs) etc. This is not surprising though. There were 13 schwas in the vocabulary and there were 4 agents. Therefore, the value by which the mean schwa duration will decrease (recall the sudden drops in Figure 5) should be a multiple of 2/52, i.e. 0.0385. Thus an average duration of 0.85 indicates that out of the 52 schwas, exactly 30 (= 52− 0.85/0.0385) were dropped. Similarly, 0.92 indicates that exactly 28 schwas were dropped. Therefore, the difference in the mean schwa duration implies the difference in the number of schwas deleted in the whole system.

4.12

We cannot however assume the schwas to be deleted in the same manner for every run. Observe that in Figure 5 the last agent to delete the schwa was much later than the other 3 agents (note that the x-axis has a logarithmic scale). In fact, theoretically it is impossible to predict the number of games after which a particular schwa will be deleted (see Bhat 2001 for general findings regarding predictability of sound changes). This leads us to an extremely difficult problem: how to decide whether a MAS experiment has stabilized or not? This question presumes the existence of a stable fixed point of a MAS. To the contrary, studies in language change have shown that there is no concept of absolute stabilization in the process of language change and languages often change along cyclical paths (Niyogi 2006). Thus, it seems that there is no method for deciding on the convergence of a simulation experiment and neither is there an upper bound on the games required for a particular change to take place. This is precisely the reason why we see a considerable variation in the average schwa durations, even though all the parameters were set to the same values.

Table 10: The results for 10 different runs under the same parameter settings. Simulation parameters: N = 4, vocabulary size = 7, d = 0.1, p_{sn} = 0.6, p_{sm} =0.2, p_{fn} = 0.6, p_{fm} = 0.0, k_{sn}= k_{sn}= 0.9, number of games=3M

Random seed	Average Communicative success (%)	Average schwa duration
1	78.19	0.85
2	76.77	0.92
3	76.78	0.85
4	78.73	0.73
5	76.39	0.92
6	78.63	0.92
7	78.00	1.27
8	76.77	1.15
9	77.16	0.81
10	75.18	0.73

4.13

Therefore, it is not possible to judge the robustness of the current model in terms of its average or asymptotic behavior. However, the following observations about the model provide us with reasons to rely on its plausibility:

- The emergent behavior is sensitive to the model parameters. In other words, the schwa deletion pattern is not ingrained in the model itself. This gives us a way to estimate the model parameters by trying to simulate the emergence of the pattern observed in real languages. These parameters can be further compared with their real counterparts (easier said than done!) for further validation of the model.
- Under certain parameter settings, the model behaves realistically, giving rise to the

pattern observed in Hindi as well as some other variants, but the variations are not random. They exhibit some patterns too.

- The communicative success, which is an important measure of the whole system is fairly stable over different runs and despite synchronic variations in the system, the system can maintain a high communicative success.

Conclusion

5.1

In this paper we described a new and emerging application area of multi-agent simulation in the field of historical linguistics. Multi-agent simulations can be used as a powerful investigatory tool for evaluation, exemplification and exploration of the theories and hypotheses of language change. This has been substantiated here by a MAS based model for explaining the emergence of schwa deletion pattern in Hindi. To summarize, this work has two major contributions — first, development of a computational framework for simulating phonological change, which is a special case of language change; and second, simulation of a real phonological change, namely the emergence of Hindi schwa deletion pattern, based on the developed framework. We discuss below the findings of this work from the perspectives of the general model as well as the specific simulation experiments.

5.2

The mental model, perceptor and articulator models, learning, and the interaction pattern between the agents are the four crucial components of the computational framework described here. The mental model is too simplistic, where the agents remember the pronunciation of each of the words separately, rather than remembering a set of general pronunciation rules. Although this model is able to capture the emergence of the schwa deletion pattern, the fact that human language acquisition as well as representation is based on phonological generalizations can hardly be doubted. An immediate extension of the model, therefore, is to replace the list of words and their realizations in the mind of the agents by a set of generalized phonological rules represented in terms of finite state transducers (FST). The FST can be learned and changed over time. This enhances the current model at least in two ways. First, generalization will obviate the learning of the individual word pronunciations, thereby reducing the simulation time and consequently allowing us to work with larger vocabulary. Second, it will reduce the amount of synchronic variation and certain characteristic dialects are expected to emerge rather than a random mixture of variants. It is expected that under suitable circumstances the agents will be able to arrive at Ohala's rule through phonological generalizations.

5.3

The articulator and perceptor models play a very crucial role in phonological changes. The articulator models the errors cropping up during casual speech leading to synchronic variation, which is the basic driving force behind any phonological change. This work models only one type of articulatory error — the reduction of schwa duration, because other types of errors seem unnecessary for explaining schwa deletion. In order to model more complex phonological changes, the articulator model can be enhanced to capture other types of errors such as metathesis and assimilation. Note that in the current framework the inherent tendency to reduce the schwa duration is a necessary precondition for the emergence of schwa deletion. Stated differently, the articulator model embeds the driving force for the phonological change, though it alone cannot explain the structure of the emergent pattern. The perceptor model however can explain the emergent pattern to some extent. The three basic assumptions used to build the perceptor model are

1. Consonants can be perceived only from the transition cues, where the perceptibility of a consonant-vowel transition depends on the duration of the nearest vowel.
2. Vowels can be perceived from their steady-state realizations as well as transitions and the perceptibility depends on the duration.
3. A string of phoneme is perceived as the word nearest to it according to the minimum edit-distance measure. However, the cost of deletion of a schwa is smaller when that particular schwa has a short duration in the agent's own vocabulary.

5.4

The first two assumptions loosely model the universal tendencies of human perception ([Jusczyk and Luce 2002](#)); "loosely" because some consonants like sibilants and liquids can also be perceived from their steady-state realizations and the duration of the transitions are not dependent on the duration of the steady-states of the nearest vowels. Nevertheless, the model succeeds to capture the fact that a consonant cannot be perceived when the duration of the adjacent vowel becomes sufficiently small. Thus, deletion of schwas reduces the perceptibility of the consonants, especially when there are no other vowels adjacent to a consonant. This decreases the likelihood of the deletion of schwas immediately after/before consonant clusters. Similarly, it prohibits the deletion of two successive schwas in words like *kAraka*, because the deletion of the two schwas will result in the string *kArk*, where the final *k* has a very low perceptibility according to the perceptor model. On the other hand, it enhances the probability of deletion of the word-initial and word-final schwas, because the peripheral schwas need to support the perception of only one consonant unlike the word-medial ones which need to support two. These facts are observable in the emergent pattern (Table 8).

5.5

The third assumption also has some important consequences on the emergent pattern. A word can be perceived if and only if there is sufficient information in its realization to distinguish it from the rest of the words in the vocabulary. Therefore, the sensitivity of the emergent pronunciation on the vocabulary is an outcome of this assumption. Moreover, since deletion of schwas is less costly than that of other phonemes, it is possible to map the pattern *kArak* to *kAraka* and not *kArakA*, though both the words could have generated the string *kArak* by deletion of a single vowel. The perception model, therefore, clearly have a strong influence on the emergent pattern. In general, we may hypothesize that one of the key factors shaping phonological change is the nature of human perception. This hypothesis can be further verified through computational modeling as well as cognitive experiments.

5.6

The present framework implements a trigger learning algorithm (TLA), where an agent learns from the last positive example that it has encountered. Niyogi ([2002](#)) discusses the different learning algorithms including TLA, and their consequences on language change. Although convergence is guaranteed by several learning algorithms, the number of examples required for convergence varies for the different strategies. The agent interaction has been modeled here through imitation games. It is a horizontal model, and therefore, there is no distinction between a learner and a teacher (or adults and children). Furthermore, the system is closed in the sense that new agents do not enter the system and old agents do not leave or die. We omit any further discussion on these issues here and leave them as open research problems for the future.

5.7

Despite the fact that the agent and agent interaction models have a significant impact on the emergent pattern, a complete description of any phonological change calls for the postulation of other influential parameters, because the agent and agent interaction models are universally identical and therefore, cannot explain the various ways in which languages have changed over time. For example, the schwa deletion pattern observed in the three languages Bengali, Oriya and Punjabi are quite different from that of Hindi, even though all of them are derived from the Vedic Sanskrit. Oriya does not exhibit schwa deletion; Bengali features word final schwa deletion, but does not allow deletion word medially; whereas unlike Hindi, Punjabi also allows deletion of schwas from the word-initial syllable. These differences have to be explained independent of the agent model. We have observed that there are certain preconditions involving the allowable rate of communication and the duration reduction step that are crucial to the emergence of the Hindi schwa deletion pattern. We propose that these parameters are some of the possible factors that govern the emergent pattern. When the allowable rate of communication is "held at a high value we observe that none of the schwas are deleted in the system (Figure 7). This is clearly the case with Oriya. We also observe that word-final schwas are deleted first. If the process of phonological change stops after this phase, or takes some other course, we can explain the schwa deletion pattern of Bengali. The case of deletion of the schwa from the word-initial syllable, as seen in Punjabi, has also been observed in the current model ([above](#)). Other factors that are known to be significant for the schwa deletion pattern observed in a language include stress and morphology. It would be interesting to extend the current model to encompass the stress pattern and the morphological features of a language, and study

their effect on the emergent pattern.

5.8

The sensitivity of the emergent pattern on the vocabulary is not surprising. It is a consequence of the perception model (assumption 3 above). The complete vocabulary of a language is expected to be optimally encoded with little redundancy, which ensures that deletion of a consonant or a vowel makes correct recognition of a word less likely. Note that in reality words are normally uttered and recognized in context, which provides extra clues for proper recognition. Moreover, a tendency towards phonological generalization will even out any discrepancy arising due to the effect of the vocabulary. Thus, we believe that the schwa deletion pattern observable in a language is independent of its vocabulary, even though in our model we do observe its sensitivity to the vocabulary owing to lack of phonological generalization and a small vocabulary size.

5.9

Some of the other important observations are as follows. 1) The simulation experiments predict the coexistence of several variants of the words. 2) The course of schwa deletion shows an S-shaped curve, which strengthens the plausibility of the model. 3) It is not possible to provide an upper bound on the time required (in terms of games) for a particular schwa to be deleted, which in turn implies that we cannot predict whether a schwa will be deleted at all. Some of these conclusions have been independently validated by other linguistic studies (discussed in the text), but it remains to be seen whether all the observations made on the MAS model are also true for the real world.

Notes

[1]Please note that this problem is not from the domain of language change; rather it pertains to the broader issue of language evolution. Nonetheless, the simplicity and the popularity of the problem and the models proposed thereby make it an excellent example for illustrating this fact.

[2]Although the generative model of language is the most popular view in synchronic linguistics, there are other views on language and language models (e.g. empiricism, memetics, functionalist views, optimality theory etc.)

[3]The "?" in the left context "VCC?" of the rule implies that the final C is optional.

[4]Once, the process of deletion or epenthesis etc. becomes a part of the regular phonology of a language, these are no longer articulatory errors, and rather learnt pronunciation rules. But the reason why we observe them initially is due to general tendencies of the human articulation process towards such effects (Ohala 1989).

Acknowledgements

References

ANDERSEN, H. (1973) "Abductive and deductive change," *Language* 49:765-793

BAILEY, C.-J. (1973). *Variation and linguistic theory*. Washington, DC: Center for Applied Linguistics.

BHAT, D.N.S. (2001) *Sound Change*. Motilal Banarsidass, New Delhi

BRISCOE, T. (2000). Macro and micro models of linguistic evolution. In *Proceedings of the 3rd*

International Conference on Language and Evolution.

BODIK P. (2003) "Language Change in Multi-generational Community", in *Proceedings of CALCI-03*, Elfa

CANGELOSI, A., and PARISI, D. (2001). How nouns and verbs differentially affect the behavior of artificial organisms. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 170–175). London: Erlbaum.

CANGELOSI, A. and PARISI, D. (Eds.) (2002) *Simulating the Evolution of Language*, London: Springer Verlag

CHEN, M., and WANG, W. (1975). Sound change: actuation and implementation. Language 51(2):255–281.

CHOMSKY, N. (1995) *The minimalist program*. MIT Press, Cambridge, MA

CHOPDE, A. (2001) "ITRANS version 5.30: A package for printing text in Indian languages using English-encoded input", Available at http://www.aczoom.com/itrans/

CHOUDHURY, M. and BASU, A. (2002) "A Rule Based Algorithm for Schwa Deletion in Hindi" *Proc Int Conf Knowledge-Based Computer Systems*, Navi Mumbai, pp. 343 — 353

CHOUDHURY M., BASU A. and SARKAR S. (2004) "A Diachronic Approach for Schwa Deletion in Indo-Aryan languages" In *proceedings of SIGPHON'04*, Barcelona, Spain, pp 20 — 26

CHRISTIANSEN, M. H. and DALE, R. (2003) "Language evolution and change". In M.A. Arbib (ed), *Handbook of brain theory and neural networks (2nd ed.)*, pp 604–606. Cambridge, MA: MIT Press.

CLARK, R. and ROBERTS, I. (1993) "A Computational Model of Language Learnability and Language Change." *Linguistic Inquiry*, 24:299–345.

CLEMENTS, G. N. (1990). "The role of the sonority cycle in core syllabification." In: J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech,* 283–333. Cambridge: Cambridge University Press

DE BOER, B. (2001) *The Origins of Vowel Systems*. Oxford University Press.

DRAS, M., HARRISON, D. and KAPICIOGLU, B. (2002) "Emergent Behavior in Phonological Pattern Change." *Proceedings of Artificial Life VIII*, 390--393. Sydney, Australia

DUTOIT T. (1997) *An Introduction to Text-To-Speech Synthesis.* Kluwer Academic Publishers

ELLEGARD, A. (1953). *The auxiliary do: the establishment and regulation of its use in English*. Stockholm: Almqvist & Wiksell.

FONTAINE, C. (1985) *Application de méthodes quantitatives en diachronie: l'inversion du sujet en francais*. Master's thesis, Universite du Québec `a Montreal.

FONTANA, J. M. (1993). *Phrase structure and the syntax of clitics in the history of Spanish*. PhD thesis, University of Pennsylvania.

HARE, M., and ELMAN, J. L. (1995). Learning and morphological change. *Cognition*, 56:61–98.

HARRISON, D., DRAS, M. and KAPICIOGLU, B. (2002) "Agent-Based Modeling of the Evolution of Vowel Harmony." in *Proceedings of North East Linguistic Society 32 (NELS32)*, 217--236. New York, NY, USA

HAUSER, M. D., CHOMSKY, N., and FITCH, W. T. (2002) "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science*, 298:1569 — 1579.

JUSCZYK, P. W. and Luce, P. A (2002) "Speech Perception and Spoken Word Recognition: Past and Present." *Ear & Hearing*. 23(1):2–40.

JURAFSKY, D. and MARTIN, J. H. (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall

KE, J., OGURA, M., and WANG, W. S-Y. (2003) "Modeling evolution of sound systems with genetic algorithm" *Computational Linguistics*, 29(1):1 — 18

KIRBY, S. (2001). "Spontaneous evolution of linguistic structure –an iterated learning model of the emergence of regularity and irregularity." *IEEE Transactions on Evolutionary Computation*, *5*(2):102–110.

KISHORE, S.P. and BLACK, A. (2003) "Unit size in unit selection speech synthesis" *Proc of Eurospeech*, Geneva, Switzerland.

KLEIN, S. (1966) "Historical Change in Language using Monte Carlo Techniques." *Mechanical Translation and Computational Linguistics* 9:67–82.

KLEIN, S. (1974). "Computer Simulation of Language Contact Models", in *Toward Tomorrows Linguistics*. R. Shuy and C-J. Bailey (eds), Georgetown: Georgetown University Press, pp 276–290.

KLEIN, S., KUPPIN, M. A., and KIRBY, A. (1969) Meives *Monte Carlo simulation of language change in Tikopia and Maori. In the Proceedings of the International Conference* on Computational Linguistics (COLING)

KROCH, A. S. (1989). "Reflexes of grammar in patterns of language change," *Language Variation and Change* 1:199–244

KROCH, A. S. (2001). "Syntactic change," in Mark Baltin and Cris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, Blackwell, Oxford, pp. 699–729.

KROCH, A. and TAYLOR, A. (1997) "Verb movement in Old and Middle English: dialect variation and language contact" in van Kemenade, A. and N. Vincent (ed.), *Parameters of Morphosyntactic Change*, Cambridge University Press, pp. 297– 325

LABOV W. (1972). *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press

LILJENCRANTS, J. and LINDBLOM, B. (1972). "Numerical simulation of vowel quality systems: the role of perceptual contrast". *Language,* 48:839 — 862.

LIGHTFOOT, D. (1991) *How to Set Parameters: Arguments from Language Change.* MIT Press/Bradford Books.

LIGHTFOOT, D. (1999) *The development of language: Acquisition, change and evolution.* Blackwell: Oxford.

LIVINGSTONE, D. and FYFE, C. (1999) "Modelling the Evolution of Linguistic Diversity". In D. Floreano, J. Nicoud and F. Mondada, (eds), *ECAL99*, pp 704–708. Berlin: Springer-Verlag.

LUPYAN, G. and MCCLELLAND, J.L. (2003). "Did, Made, Had, Said: Capturing Quasi-Regularity in Exceptions". Presented at the *25th Annual Conference of the Cognitive Science Society*. Available at http://www.cnbc.cmu.edu/~glupyan/LandM-cogsci2003.pdf

MISRA, B. G. (1967) "*Historical Phonology of Standard Hindi: Proto Indo European to the present*" PhD Dissertation, Cornell University

NARASIMHAN, B., SPROAT, R., and KIRAZ, G. (2004). "Schwa-deletion in Hindi Text-to-Speech Synthesis," *International Journal of Speech Technology*, 7(4):319–333

NIYOGI, P (2002) "The Computational Study of Diachronic Linguistics." In D. Lightfoot (ed), *Syntactic Effects of Morphological Change*. Cambridge University Press

NIYOGI, P. (2006) *The Computational Nature of Language Learning and Evolution.* Cambridge, MA: MIT Press.

NIYOGI, P. and BERWICK, R. C. (1998) The Logical Problem of Language Change: A Case Study of European Portuguese. *Syntax: A Journal of Theoretical, Experimental, and Interdisciplinary Research*, 1.

OHALA, M. (1983) *Aspects of Hindi Phonology*. MLBD Series in Linguistics, Motilal Banarsidass, New Delhi.

OHALA, J. (1989). "Sound change is drawn from a pool of synchronic variation". In L.E. Breivik & E.H. Jahr (Eds.), *Language change: Contributions to the study of its causes*. Berlin: Mouton de Gruyter

OUDEYER, P.-Y. (1999). "Self-organization of a lexicon in a structured society of agents." In D. Floreano, J.-D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life: The Fifth European Conference (ECAL '99)* 1674:725-729). Berlin: Springer.

OUDEYER, P-Y. (2005) "The Self-Organization of Speech Sounds", *Journal of Theoretical Biology,* 233(3): 435- 449

PARISI, D. and CANGELOSI, A. (2002) "A Unified Simulation Scenario for Language Development, Evolution, and Historical Change." In A. Cangelosi and D. Parisi (eds), *Simulating the Evolution of Language*, pp 255-276. London: Springer Verlag

PERFORS, A. (2002) "Simulated Evolution of Language: a Review of the Field" *Journal of Artificial Societies and Social Simulation*, 5(2).

ROBERTS, I. (2001) "Language change and learnability". In S. Bertolo (Ed.), *Parametric Linguistics and Learnability: a Self Contained Tutorial for Linguists*, Cambridge University Press

RUSSELL, S. J and NORVIG, P (2003) *Artifical Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, New Jersey, second edition.

SCHWARTZ, J. L., BOEK, L. J., VALLEHE, N. and ABRY, C. (1997) "The dispersion-focalization theory of vowel systems", *Journal of Phonetics*, 25, 255 — 286

SMITH A. D. M. (2005). "Mutual Exclusivity: Communicative Success Despite Conceptual Divergence". In Maggie Tallerman, editor, *Language Origins: Perspectives on Evolution*. Oxford University Press.

SMITH, K. and HURFORD, J. R. (2003) "Language Evolution in Populations: Extending the Iterated Learning Model" in W. Banzhaf, T. Christaller, J. Ziegler, P. Dittrich and J. T. Kim (eds.) Advances in Artificial Life: Proceedings of the 7th European Conference on Artificial Life, pp. 507-516.

STEELS, L. (1997) "The synthetic modeling of language origins" *Evolution of Communication*, 1(1):1 — 34

STEELS, L. (2001) "Language games for autonomous robots" *IEEE Intelligent systems*, pp 16 — 22

TURNER, H. (2002) "An Introduction to Methods for Simulating the Evolution of Language" In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, pp 29 — 50. London: Springer Verlag.

WANG, W. S-Y. and Minett, J. W. (2005) The invasion of language: emergence, change and death. *Trends in Ecology and Evolution*, 20(5):263-269

WAGNER, K., REGGIA, J. A., URIAGEREKA, J., and WILKINSON, G. S. (2003) Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37--69.

WEINREICH, U., LABOV, W. and HERZOG, M. (1968). "Empirical foundations for a theory of language change" In W. P. Lehmann & Y. Malkeil (eds.), *Directions for historical linguistics: A symposium*. Austin: University of Texas Press. 95-188