Computational Models of Real World Phonological Change

Thesis submitted in the partial fulfilment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

Monojit Choudhury

under the guidance of

Prof. Anupam Basu
Prof. Sudeshna Sarkar

Department of Computer Science and Engineering,
Indian Institute of Technology, Kharagpur, India
May 2007

Certificate

This is to certify that the thesis titled **Computational Models of Real World Phonological Change**, submitted by **Monojit Choudhury**, a research scholar in the *Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, India*, for the award of the degree of **Doctor of Philosophy**, is a bona fide record of an original research work carried out by him under our supervision and guidance. The thesis fulfills all the requirements as per the regulations of the institute and in our opinion, has reached the standard needed for submission. Neither this thesis nor any part of it has been submitted for any degree or any academic award elsewhere.

Anupam Basu
Professor
CSE, IIT Kharagpur

Sudeshna Sarkar
Professor
CSE, IIT Kharagpur

Acknowledgments

Nothing makes sense, unless in the context of its history. Therefore, in my opinion, the best way to acknowledge the people and places behind any work is to churn out the history behind it. Here goes the story of this thesis, as viewed from my perspective . . .

Five years back, in the summer of 2002, after I graduated with a B.Tech from the Department of Computer Science and Engineering, IIT Kharagpur, I was wondering what could be the next step that's best for my intended research career. And when I felt that I needed more time to search for the *optimal strategy* of my career, I joined the Communication Empowerment Laboratory (CEL) funded primarily by Media Lab Asia (MLA). The choice was obvious because, under the supervision of Prof Anupam Basu, Mohit and I had been working on a grapheme-to-phoneme converter for Hindi during my B.Tech final semester, and in CEL Prof Basu wanted me to continue with the same interesting stuff. It is during this period, when I met the first problem of this thesis – "schwa deletion in Hindi" – and fell in love with it!

Meanwhile, motivated from the discussions that I had with Prof S. Rajopadhyaye, Prof P.P. Chakrabarty, Prof P. Dasgupta and none other than Prof Sudeshna Sarkar and Prof Basu, in January 2003, I finally joined as a research scholar in IIT Kharagpur to continue my journey for four more years with the department and the then-born CEL. My research topic was – no, not language change, but "formal models of visual languages", though I had been simultaneously working on several aspects of Hindi and Bengali NLP.

Intrigued by the structural properties of natural languages, I wanted to delve deeper into them; in the process, I met Prof Shirajuddin Ahmed, Prof Jayshree

Chakrabarty, Prof Thakur Das and Prof B. N. Patnaik, who introduced me to the basic concepts in diachronic linguists, came up with suggestions such as schwa deletion could have resulted from syllable minimization, and encouraged me to pursue research in NLP as well as diachronic models. I met the second problem tackled in this thesis – morpho-phonological change affecting BVI – when we were working towards morphological generator and analyzer engines for Bengali.

As a few months passed by, seeing my natural instincts and interests for natural languages, my supervisors were kind and flexible enough to provide me the option of switching the research area despite the fact that the field was completely alien to them, as was to me. I was most happy to change the problem area of my thesis to the one where it stands now. Nevertheless, it was only during EvoLang 2004 (Leipzig) that I got a chance to get myself well acquainted with the tools and techniques used in the field. Prof J. Hurford and Prof B. Comrie – the prime organizers of the conference – readily agreed to partly support the trip (the rest was funded by MLA) and waive the registration fees. In EvoLang, I had the opportunity to interact with Prof J. Hurford, Prof B. de Boer, Prof P. Vogt, Prof R. Ferrer-i-Cancho and Prof D. Livingstone among many others, who enlightened and encouraged me to work in the area. I am especially grateful to Prof de Boer, who painstakingly replied to all my technical queries during the whole PhD tenure, and whose suggestions were of great help in shaping this thesis.

I must acknowledge the contributions of several persons who worked with me on parts of the work presented in the thesis: Mohit and Harish V. (implementation of Hindi grapheme-to-phoneme converter), Pradipta Ranjan Ray (implementation of Bengali grapheme-to-phoneme), Monjur Alam (collection of Bengali verb data and implementation of the rewrite rule based model), Arijit Sengupta (an alternative model for Bengali verb inflections) and Vaibhav Jalan (implementation of the MOGA model). I am also grateful to all my co-authors, with whom I have done some interesting research work (the list is long and I shall not try to enumerate).

It had been a great experience to work in CEL and I must thank my friends and fellow-researchers in the Lab. Apart from all the personal and professional helps extended by them at every time, it is for them that I never had a chance to feel low during the last four years. Staying at VSRC was fun, and special thanks to my roomies: Nilanjan, Anirban and Tirthankar.

During the last couple of years, I have been exploring statistical mechanical models of language and I am grateful to Prof Niloy Ganguly and Animesh Mukherjee – my co-explorers in this venture for making this journey exciting and successful. Very special thanks to Animesh, who along with my supervisors has painstakingly reviewed every chapter of this thesis and helped me improve the presentation. I am also grateful to Prof Ganguly, Mousumi-di and Pratyush, who reviewed some parts of the draft. Thanks again to Tirthankar and Santosh for helping me out with the printing and binding processes.

I express my sincere indebtedness to the staffs and faculty members of the department. Special thanks to Prof A. Gupta, Prof S. Ghosh, Prof A.K. Majumdar, Prof D. Sarkar and Prof P. Mitra for their suggestions and comments on the work. I am grateful to Elsevier and Cambridge University Press for granting the permission to reproduce Fig. 1.1 and Fig. 2.1 respectively.

Finally, I would not like to demean the role of my parents and supervisors by "thanking" them, for they are the invisible shapers of this story.

Monojit Choudhury
Date:

Abstract

Language change refers to the phenomenon of alterations in the linguistic structures of the language used by a community over time, which may subsequently give rise to one or more languages or dialects. Phonological change is a special type of language change that involves alteration in the pronunciation patterns. The aim of this thesis is to develop computational models for cases of phonological changes that have been observed for real languages and validate the same against real data.

Several theories have been proposed to explain the cause and nature of phonological change, which, nevertheless, cannot be validated and established firmly due to the lack of sufficient linguistic data and scope of experimentation. Computational models not only provide a formal description of the linguistic theories, but also serve as excellent virtual laboratories. The inherent complexities of our cognitive and social structures, however, make it hard or impossible to replicate reality within a computational framework. Consequently, most of the present computational models are developed for toy languages and are rarely compared with real linguistic data – an obvious weakness that has attracted criticisms from linguists and computer scientists alike.

In this thesis, we present computational models pertaining to the different explanations of phonological change for two cases of real world language change: *schwa deletion in Hindi* (SDH) and *change affecting the verb inflections of the dialects of Bengali*. Hindi and Bengali are two major languages of the Indo-Aryan family that are spoken in the Indian sub-continent. The models are developed in an incremental manner, where the findings of a more detailed, but computationally intensive model are used to abstract out the details in the sub-

sequent phases of the modeling. This, in turn, leads to extensive models that are not as detailed as their precursors, but computationally more tractable and linguistically equally plausible. This, which we call the *hierarchical abstraction methodology*, is the prime contribution of the thesis, apart from the novel formalisms, analysis techniques and the applications that have been developed in the context of the two specific problems tackled.

The models proposed in this work include: a constrained-optimization model for SDH that has been solved analytically to show that the pattern of SDH is optimal under certain assumptions; two multi-agent simulation models for SDH, where it has been shown that under suitable circumstances the pattern of SDH automatically emerges in these systems without any explicit optimization or global control; a multi-objective genetic algorithm based optimization model to explain the morpho-phonological change of Bengali verb inflections and the dialect diversity arising as a consequence of the same. While the optimization models put forward functional explanations of phonological change, the multi-agent simulation models provide an emergent explanation grounded in the principles of *phonetically-based phonology* and *evolutionary phonology*.

To summarize, this thesis shows that through the concept of hierarchical abstraction methodology, it is indeed possible to build computational models of real world language change. Thus, we believe that in future, computational techniques will serve as an important and inevitable tool for research in diachronic linguistics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As you are reading these words, millions of neurons are triggered in your brain; through a mysterious coordination and combination of electrical signals, they paint the meaning of the sentence on the canvas of the mind. Despite such a complex underlying mechanism, we utter and understand new sentences quite effortlessly. Indeed, human language is one of the greatest natural wonders of the universe, whose structure, function, evolution and dynamics are as elusive to modern science as these were to its earliest forefathers.

Language is a unique cognitive ability of the human species that facilitates storage, expression and sharing of unlimited ideas and information - the building blocks of our social, cultural and scientific endeavors. Animals do communicate with alarm calls, pheromones, and other types of signals, but their communication medium lacks the *recursive syntax* and *compositional semantics* (see Hauser (1997) and Hauser et al. (2002) for reviews). These two unique properties of human language allow the creation and comprehension of virtually an infinite number of sentences, an advantage clearly bereft in animal communication.

Like recursive syntax and compositional semantics, there are several other features, collectively termed as *language universals* (Greenberg 1963; Hawkins 1983), that are common to all human languages. Nevertheless, languages display a wide range of variation in not only the set of words or sounds they use, but also in the way sounds are combined to form words, words are combined to form sentences and so on. Even within a linguistic system (i.e. a language), one can observe a wide range of variation at different levels: variation among individuals (*idiolects*), variation among different communities (*dialects*), and variation over time. The first two kinds are examples of *synchronic variation*, whereas the third one is known as *diachronic variation* or *language change*.

As we shall see, in spite of a long, fruitful and incessant history of over two centuries of research, diachronic linguistics has given rise to far more controversies than universally accepted facts and theories. The reason, presumably, lies in the methodological difficulties associated with the field due to the paucity of data and impossibility of experimental validation. Like many other fields pertaining to social and natural sciences, diachronic linguistics has of late seen a rise in the use of computational techniques for exploring, evaluating and enhancing the existing theories. Nevertheless, modeling a case of a real world language change turns out to be an extremely difficult problem; this is due to the complexities associated with our physiological, cognitive and social structures – all of which are equally instrumental in shaping the course of language change.

What are the methodological difficulties associated with theorizing and validation in diachronic linguistics? Can computational techniques be of any help in this regard? How can one construct linguistically plausible, yet computationally tractable, models of language change? What kind of inferences can be drawn from such models about the problems being modeled, and language change, in general? The present work is an attempt to answer these fundamental questions regarding the usefulness and feasibility of computational modeling in diachronic linguistics.

In this chapter, we introduce the central problems and concepts in diachronic linguistics and enumerate the methodological difficulties associated with the field (Sec. 1.1). This is followed by a brief overview of the computational techniques used in diachronic linguistics, and their advantages and limitations (Sec. 1.2). A detailed survey of the same will follow in Chapter 3. The objectives of the thesis are presented in Sec. 1.3 and Sec. 1.4 summarizes the salient contributions of the work. The organization of the thesis is described in Sec. 1.5.

1.1 Diachronic Linguistics: An Overview

The issue of "the origin and past history of an individual's ability to speak a language" can be studied from three different perspectives (Parisi and Cangelosi 2002):

- **Language Acquisition:** The process of learning a language by an *individual* based on the linguistic inputs from the environment is termed as language development or language acquisition. The developmental history of an individual over a few decades determines the characteristics of the *idiolect* – the individual's language usage.
- **Language Change:** The linguistic inputs available to an individual that shape the idiolect, are a collection of utterances of the members of a *social group*, to which the individual belongs. The nature of these utterances, collectively called a dialect, is an outcome of the linguistic and socio-cultural history of the group spanning a few centuries. This phenomenon of transformation of dialects at the level of social groups is referred to as language change.
- **Language Evolution:** The cognitive faculty of language as well as the allied articulatory and perceptual mechanisms, which are essential for an individual to acquire a language, are the results of the biological evolution of our *species* over hundreds of thousands of years. This process has been termed as language evolution.

These three processes - acquisition, change and evolution, are active at three widely different time scales (Hurford 1991b; Wang 1991). Nevertheless, they are greatly interdependent, and as depicted in Figure 1.1, cannot be studied in isolation (see Parisi and Cangelosi (2002), Christiansen and Kirby (2003a) for further discussion on this).

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.

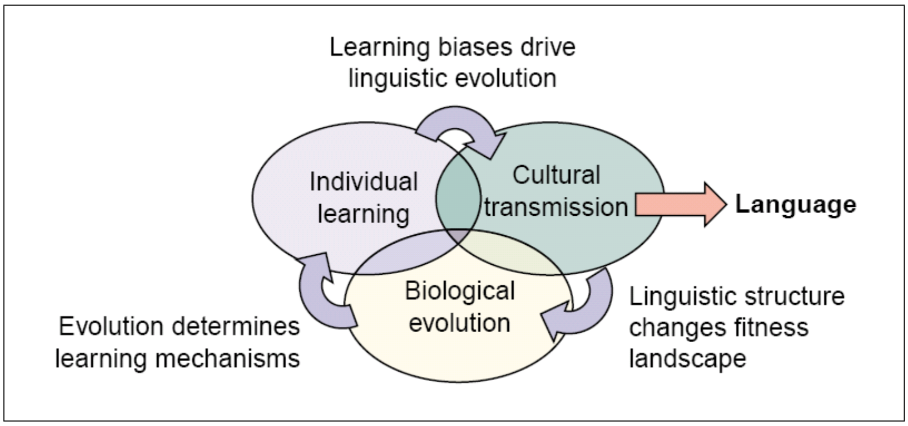


Figure 1.1: Language arises from the interactions of three adaptive systems: individual learning, cultural transmission, and biological evolution. Any comprehensive theory of language evolution is expected to explain the interaction of these systems on three different timescales: the lifetime of the individual (tens of years), the language (thousands of years), and the species (hundreds of thousands of years). Adapted from (Christiansen and Kirby 2003).

1.1.1 Different Perspectives on Language Change

The theories of language and its dynamics can be broadly classified into the *nativist* and *functionalist* paradigms (Hurford 1991b) (also see Perfors (2002) for a review on the theories of language evolution).

According to the *nativist position*, popularized by Noam Chomsky (Chomsky and Halle 1968; Chomsky 1981; Chomsky 1995), every individual has an internalized model or grammar of the language, called the *I-language*. The externalized language or *E-language*, which is the collection of utterances, is the realization of the I-language under specific conditions. The I-language is considered to be "an instantiation of the initial state of the cognitive system of the language faculty with options specified" (Chomsky and Halle 1968; Chomsky 1981; Chomsky 1995). The cognitive system of the language faculty is defined in terms of a fixed set of principles, the so-called *Universal Grammar* (UG), and each I-language is just a particular instantiation of these principles to certain parameters.

Thus, the nativists claim that language is an innate property of the humans, which is an outcome of the biological evolution. The linguistic universals are principles, i.e., the UG, whereas the variation is a consequence of difference in parameter values. In this framework, language acquisition is the process of setting the right parameters after observing a set of triggers, i.e. linguistic inputs; similarly, language change is the process of variation in the parameter values over time (Roberts 2001).

The *functionalist* perspective of language posit the brain as a general-purpose reasoning device; the properties of the languages, both universal and variable, are then explained in terms of language usage, rather than appealing to the structure of the language processing device. Although there is no unambiguous definition of the term "functionalism", we can consider all anti-nativist accounts of language as "functional" ones[1]. As an example, functionalists have argued that the structural properties of languages are explainable in terms of the diachronic processes (Sampson 1970; Bybee 1985; Bybee 1988; Bybee 2001; Blevins 2004). Also, see (Comrie 1981; Tomasello 2002; Tomasello 2003; Boersma 1998; Boersma 1997b) for other examples.

Since in the functionalist paradigm, language is not innate, the structure and dynamics of language are emergent properties of a linguistic system. Thus, language is viewed as a *complex adaptive system* (Steels 2000), whose dynamics is governed by the structure and the interactions of the language users. The process of language change is the evolutionary dynamics, or equivalently, the process of self-organization of the linguistic system. Several computational models of language evolution and change are built upon this idea, which will be discussed in Chapter 3.

There are also suggestions to combine the nativist and functionalist views in the same framework by positing, for example, the functional evolution of the UG through *natural selection* (Pinker and Bloom 1990; Pinker 1994; Pinker 2000), or the emergence of language universals through language usage, subjected to the constraints imposed by UG (Kirby 1999).

[1] In Chapter 2, in the context of phonology, we shall provide a more specific definition of "functional explanations".

1.1.2 The Paradox of Language Change

One of the holy grails of diachronic linguistics is to resolve the "paradox of language change"(Lightfoot 1991; Lightfoot 1999; Clark and Roberts 1993; Niyogi and Berwick 1998). Languages are stable over 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? For example, if we compare the syntax of Japanese and English, we find that the former has hardly changed over the past five hundred years, while the latter has undergone several significant changes; the English verb *do*, for instance, developed into an auxiliary from one of its main verb senses at some point in Middle English without any known exogenous cause (Kroch 2001; Ellegard 1953).

Apparently, we are faced with a logical paradox; if we say that the process of language acquisition is not perfect, then how can we explain the stability of a language like Japanese over several generations; and if we claim that language acquisition is perfect, then what triggers the language change as observed in the case of a language like English? This problem of why change occurs, when and where it does, has been termed as the *actuation problem* by Weinreich, Labov and Herzog (1968).

There have been several attempts to explain this paradox. One of the major approaches has been to deny the very possibility of endogenous language change by attributing all cases of change to some socio-cultural and/or other exogenous causes. See, for example, (Labov 1972; Milroy 1993; Chambers 1995; Nettle 1999) for social accounts of language change, and (Santorini 1989; Kroch and Taylor 1997a) for explanations based on language contact. Endogenous change at the phonological level is a well accepted fact, and changes at the higher strata of linguistic structure, like the morphology and syntax, have often been attributed to some underlying phonological change prior to the higher level change (Kroch 2001). The proponents of endogenous change have tried to explain the paradox by postulating functional forces of low perceptual complexity and ease of articulation (Vincent 1976; Slobin 1979; Kiparsky 1996), non-deterministic learning process coupled with the co-existence of multiple compatible grammars for a set of linguistic inputs (Andersen 1973a), drifts in the usage frequencies of the various types of linguistic structures (Lightfoot 1991; Lightfoot 1999), etc. Kroch (2001) provides a comprehensive and in-depth account of these theories and their criticisms analyzed in the light of real language change data.

1.1.3 Causal Forces in Language Change

"Language change is explained when its causal forces are identified and their interactions are made clear" (Yang 2000). To represent the various possible causes discussed above, that might trigger and/or govern the course of language change, we propose a schematic of a linguistic system shown in Fig. 1.2. The bubbles denote the I-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 conflicting explanations to co-exist that may explain the same historical data, but are hard to validate in general. We explain the different possibilities through Fig. 1.2.

In the figure, the vertical arrows represent learning and the horizontal (T-shaped) arrows represent communication between speakers through the E-language. The thick black colored arrows represent language acquisition by the children. It is possible that this process is imperfect, leading to a different G_{n+1} (Andersen 1973a). 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 $(n+1)$th generation learns the language, can be different from the E-language from which the nth generation learnt the language. Such a change in the E-language can be an outcome of any of the following cases.

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 con-

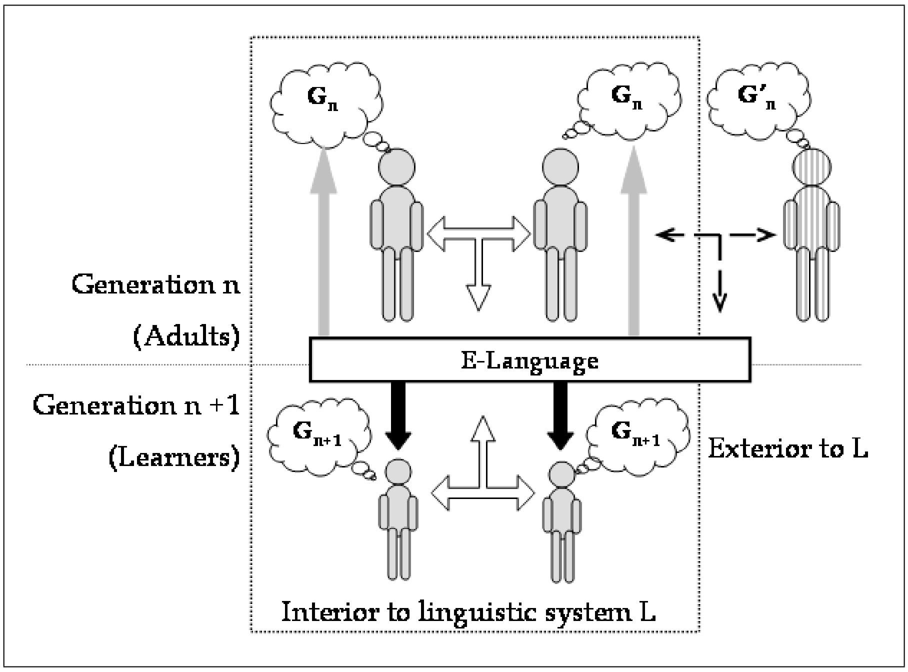


Figure 1.2: Language transmission in an open linguistic system. One or combination of several factors (shown by arrows in the diagram) can trigger an event of change – a case where the grammar G_{n+1} is different from G_n. The thick black and gray arrows represent learning in children and adults respectively. The thick white arrows represent communication between the speakers within a linguistic community, whereas the dashed black arrow represents communication between the speakers from two different language communities.

cerned (Lightfoot 1991); this means the thick white arrows initiate the process of 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. This is known as change due to language contact (Santorini 1989; Kroch and Taylor 1997a; Kroch and Taylor 1997b).

3. Adults themselves learn from the new E-language, which can be an out-

come of language contact or some other socio-cultural event (Labov 1972) and therefore, produces a different E-language. This is represented in the diagram by gray arrows.

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 complex, 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 1.2 the individuals of the same generation are shown to have the same I-language, synchronic variation is present in all linguistic systems (Ohala 1989).

1.1.4 Methodological Difficulties

Thus, we have seen that several theories have been proposed to explain the cause and the nature of language change (see e.g., Kroch (2001) and references therein for an overview of explanations of grammar change, Chapter 3 in Blevins (2004) for explanations of phonological change). However, various practical factors, such as those listed below, restrict us from accepting one of these numerous explanations as "the theory of language change" (Kroch 1989a).

- **Problem of evidence**: Languages do not fossilize. We do not have recorded speech of the past and the small amounts of written data that we have, hardly reflect the spoken form of the language or its dialects in the past. Therefore, in the absence of enough historical data, most of the theories remain untested. See Meillet (1967) for kinds of linguistic data, and the first chapters of Andersen (1973b) and Bhat (2001) for the pitfalls associated with different types of evidence used in historical linguistics.
- **Problem of validation**: Like other social sciences and unlike most of the natural sciences, it is not possible to experimentally verify the theories of language change. We cannot see languages changing in our laboratories, and have to embark on nature for providing the data.

- **Problem of theorizing**: In the absence of any comprehensive theory of human cognition and learning processes, we cannot predict the exact nature of the grammar and its representation in the brain. Consequently, our observations and postulations regarding language and language change are limited to the external language (speech and text data) only (Hauser et al. 2002).

The first two problems are the most severe methodological difficulties encountered in diachronic linguistics, due to which various assumptions become inevitable while inferencing from the little historical data that is available to a researcher (Labov 1975a; Labov 1975b). In the next section, we shall see how computational techniques can help us circumvent some of these problems.

1.2 Computational modeling techniques

In recent times, there has been a revival of interest in computational models of diachronic explanations, presumably due to the successful application of such techniques in the related field of language evolution. See (Steels 1997b; Perfors 2002; Christiansen and Dale 2003; Wagner et al. 2003; Wang et al. 2005; de Boer 2006) for surveys on these models. Computers can serve as excellent virtual laboratories for linguistic experiments, which are otherwise impossible. Thus, computational modeling can be used for evaluating, exploring and enhancing the theories in diachronic linguistics.

As described below, computational models can be classified into two broad categories: *analytical* and *synthetic*.

1.2.1 Analytical Models

In the analytical approach, the explanation for language change is encoded in a formal framework and solved analytically and/or computationally for obtaining the values of the model parameters. The parameters in turn reflect the necessary and sufficient conditions under which a particular or general case of language change might have taken place. Analytical techniques include models based on statistical physics (Itoh and Ueda 2004; Kosmidis et al. 2005; Kosmidis et al. 2006b), complex networks (Ferrer-i-Cancho 2003; Sole et al. 2005; Dall'Asta et al. 2006a; Ke and Yao 2006), dynamical systems (Niyogi and Berwick 1997c; Komarova et al. 2001; Niyogi 2002; Komarova and Nowak 2003; Abrams and Strogatz 2003; Niyogi 2006; Mira and Paredes 2005), constrained optimization (Liljencrants and Lindlom 1972; Ke et al. 2003) etc. These models have been successfully applied to explain various phenomena, such as the change in word ordering rules (Itoh and Ueda 2004), the interdependence of language change and language acquisition (Komarova et al. 2001; Komarova and Nowak 2003; Niyogi 2002; Niyogi 2006), and language competition and death (Abrams and Strogatz 2003; Mira and Paredes 2005).

1.2.2 Synthetic Models

In the synthetic approach, the linguistic system is usually modeled from the perspective of the language-users and it is simulated using a computer, whereby the system is closely observed. If several runs of the simulation under different parameter settings show the emergence of certain common characteristics, then one can conclude, with sufficient confidence, that the model is a plausible explanation for those observed characteristics. Thus, simulations, in this case, are analogous to virtual experiments. Multi-agent simulation (MAS) is the most popular computational approach to language evolution and change, and has been popularized by the works of Steels and his co-researchers (Steels 1996; Steels 1998; Steels and Kaplan 1998; Steels 2001; Belpaeme 2002) rather recently, even though the earliest examples date back to the 60s (Klein 1966; Klein et al. 1969).

The key issues in designing a MAS model for language change involve agent modeling (i.e., learning and other cognitive processes), agent interaction (usually modeled through *language games*) and language representation. MAS models have been constructed to explain the emergence of Turkic vowel harmony (Dras et al. 2003), change in word ordering rules (Minett et al. 2006), lexicon change (Steels and Kaplan 1998), and homophony and borrowing (Ke

2004). Other than MAS, connectionist models, such as artificial neural networks have also been used for synthetic explanations of language change (see, for example, Smith (2002) and Sugita and Tani (2005)).

Thus, if not completely, computational methods can help us overcome at least some of the methodological difficulties in diachronic linguistics (Cangelosi and Parisi 2002). For example, computer simulations can facilitate designing of virtual experiments for validation of linguistic theories of change. Similarly, mathematical models of diachronic variation can provide further insights into the necessary and sufficient conditions that lead to a particular type of change.

Nevertheless, due to the inherent complexities of our linguistic and social structures, modeling of real language change turns out to be extremely hard. Consequently, with the exception of a few (Klein 1966; Klein et al. 1969; Klein 1974; Hare and Elman 1995; Harrison et al. 2002; Dras et al. 2003; Ke 2004), all the mathematical and computational models developed for explaining language change are built for artificial toy languages. This has led several researchers to cast a doubt on the validity of the current computational models as well as the general applicability of computational techniques in diachronic explanations (Hauser et al. 2002; Poibeau 2006).

1.3 Objectives

The primary objective of this thesis is to develop computational models pertaining to the different explanations of *phonological change* for certain cases of "real language change", and thereby demonstrate the usability and effectiveness of computational techniques in diachronic explanations. We shall restrict ourselves to models of phonological changes because – (1) phonological changes are regular in nature (Bhat 2001), (2) in most of the cases, they are explainable in terms of endogenous causes only, (3) the nature of phonological change is functionally similar to other types of linguistic change and therefore, models developed for phonological change are also applicable to other cases of change.

There are several challenges in modeling real languages. The first problem is to identify instances of historical changes, for which we have enough historical data for validation. Furthermore, the cases chosen must be simple enough to facilitate tractable modeling and must be explainable in terms of a few causal forces.

Since it is not possible to model all the characteristics of the human articulatory, perceptual and learning processes, and the linguistic interaction patterns in the society, another major difficulty in modeling real changes turns out to be the determination of the *appropriate abstraction level*.

An equally important objective of this work is to compare the different explanations of language change, especially phonological change, in the light of computational modeling. Specifically, we would like to develop models for the various explanations of phonological change and compare their linguistic plausibility as well as computational tractability.

Besides the aforementioned generic objectives, this work also aims at making specific predictions regarding the real changes that have been modeled. Some of the typical questions that we would like to answer are - how these changes were actuated, what exactly was the course of the change, why did the different dialects emerge, and how to explain the exceptions, if any? Our aim here is to provide plausible answers to these questions through computational models and vis-á-vis investigate their linguistic possibility independent of the models.

Another objective of the work is to demonstrate the applicability of the diachronic models in practical systems such as text-to-speech synthesizer and natural language generator (see Jurfsky and Martin (2000) for an overview of these systems). We show that the insights gained from the diachronic models can be appropriately used for developing NLP systems, especially those catering to a multilingual or multi-dialectal environment.

1.4 Contributions

In this work, we have developed three models for the phenomenon of schwa deletion in Hindi and one model for explaining the phonological change affecting the verb inflections in Bengali. The models entail different explanations of phonological change and are built upon appropriate computational techniques.

Schwa deletion is a phonological phenomenon, in which schwas in unstressed syllables are deleted during speech, even though the morphological and etymological evidence assert the presence of the vowel in that particular context (Ohala 1983b). Schwa Deletion in Hindi (SDH) is a diachronic process (Mishra 1967; Ohala 1983b) and can be functionally explained in terms of syllable minimization[2]. The computational models developed for SDH during this work and the basic observations made from them are described below.

Constrained Optimization Model

We formulate SDH as a syllable minimization problem, subject to certain constraints that are representative of factors such as *perceptual contrast* and *phonotactics*, and theoretically prove that the optimal schwa deletion pattern under this formulation is equivalent to that observed for Hindi. An algorithm for SDH is proposed on the basis of this finding, which is used to develop a grapheme-to-phoneme converter for a Hindi text-to-speech system. Nevertheless, the model cannot explain the process, by which such an optimization might have taken place, despite the lack of any conscious effort on the part of the speakers.

MAS Model I

MAS Model I entails a *phonetically-based* explanation[3] of SDH, where the language users are modeled as linguistic agents that interact with each other through language games and learn from the outcome of the games. We observe that under suitable circumstances, the emergent schwa deletion pattern closely resembles that of Hindi. However, the emergent pattern is sensitive to the lexicon, and the high time complexity of the model does not allow us to run simulation experiments with a real lexicon.

[2]Functional explanations of phonological change state that the process of change tries to optimize the communicative function of a phonological structure by, for example, reducing the effort of articulation. See Sec. 2.2.1 for details.

[3]In phonetically-based explanations, a phonological change is attributed to the phonetic factors such as articulatory and perceptual errors. See Sec. 2.2.2 for details.

MAS Model II

In order to curtail the time complexity of MAS Model I, we make several assumptions, such as phonological rule generalization over words, and incorporate various sophisticated computational techniques to exploit those assumptions. The resulting model – MAS Model II – facilitates experimentation with real Hindi lexicon. Here also we observe that the emergent rule for schwa deletion is similar to its real world counterpart.

Phonological Change of Bengali Verb Inflections

Since SDH is a simple case of phonological change, we also choose to model a more complex incident of phonological change – Phonological Change of Bengali Verb Inflections – where many simpler phonological changes (similar to SDH) have taken place in sequence. We assume that the individual instances of simple or atomic phonological changes can be explained in terms of MAS-based models, and therefore, take them for granted. We develop a multi-objective genetic algorithm (MOGA) based model to explain the emergence of the verb inflections in the modern dialects of Bengali from their Classical Bengali counterparts (Chatterji 1926). Since the objectives and the constraints in the MOGA Model reflect the functional forces, the model offers a functional explanation for the phenomenon of change.

Hierarchical Abstraction Methodology

Modeling and validation of real language change are extremely hard. To render the problem tractable, one has to integrate appropriate assumptions at different levels, the assumptions being justifiable independently. We solve this problem through a *hierarchical abstraction methodology*, which can be stated as follows.

> *The findings of a more general, but computationally intensive model are used to abstract out the details in the subsequent phases of modeling. This, in turn, leads to models that are not as general as their precursors, but computationally more tractable and linguistically equally plausible.*

The Constrained Optimization Model provides us the insight that the schwa deletion pattern of Hindi can be explained in terms of syllable minimization. This fact has guided us in designing the signal representation scheme as well as the perception and articulatory processes in MAS Model I at the right level of abstraction. Nevertheless, the MAS Model I could not deal with a large lexicon. The results of MAS Model I enabled further simplifications of the articulatory and perceptual processes without the loss of generality, and consequently helped in overcoming the lexicon-size constraint in MAS Model II. These three models together established the fact that a simple phonological change like vowel deletion can be explained in terms of phonetic and evolutionary factors. Therefore, in the MOGA Model for Bengali verb inflections, we assume that similar explanations exist for the phonological phenomena of deletion, assimilation and metathesis, and treat them as atomic operations. This facilitates modeling of complex phonological changes involving a sequence of simpler atomic changes.

Some of the other significant observations made during this work are enumerated below.

1. **Explanatory power:** Functional and phonetic explanations for phonological change are often overlapping in nature, and a MAS based evolutionary model indirectly captures the phonetic and functional forces through agent modeling and language data. Nevertheless, evolutionary models can explain more complex changes than what functional and phonetic forces alone can.

2. **Predictability and convergence:** It is a well known fact that even though the direction of a sound change is predictable, its occurrence is not (Bhat 2001). This unpredictability of phonological changes has been observed in the MAS models, where the course and time of the change for different runs of the same experiment are different, despite the similarity of the globally emerging pattern. Since the time of a change is unpredictable, one cannot provide an upper bound on the number of games, after which a specific event of change will take place. Therefore, it is not possible to define a notion of "convergence" for the simulation experiments.

3. **Applicability:** Although the evolutionary model is more powerful than the functional or phonetic models, the latter are more efficient computationally, and thus, can cater to practical applications.

To summarize the contributions of this thesis in a single sentence, *we have shown that through the concept of hierarchical abstraction methodology, it is indeed possible to build computational models of real world language change. Thus, we believe that in future, computational techniques will serve as an important and inevitable tool for research in diachronic linguistics.*

1.5 Organization of the Thesis

The thesis is organized into eight chapters.

Chapter 2 provides the necessary linguistic background of the work by describing the different properties and explanations of phonological change. The two problems dealt with in the thesis, namely SDH and change in Bengali verb inflections, are also discussed at length.

Chapter 3 presents a survey of the existing computational models of language change. A taxonomy of the different models is proposed and the pros and cons of the modeling techniques are discussed. A few models from the domain of language evolution are also discussed at length due to their direct relevance to language change. Finally, possible applications, explanatory gaps and research possibilities are identified.

Chapter 4 is devoted to the constrained optimization model for schwa deletion in Hindi. A general framework for modeling functional explanations is introduced, and SDH is formulated as a constrained optimization problem within this framework. Next, an analytical solution to the optimization problem is derived and its equivalence to Ohala's rule is established. An algorithm for schwa deletion, its evaluation and application in Hindi text-to-speech synthesizer are also discussed.

Chapter 5 presents the design, experiments and inferences for the MAS Model I. The agent model, the structure of the language games and the simulation setup are outlined; simulation experiments and their results are presented;

and finally, the interpretations of the results are provided in the light of phonetically based accounts of phonological change.

Chapter 6 is dedicated to MAS Model II. The improvements made over MAS Model I (e.g. rule learning) and certain simplifications conceived as an outcome of the results of previous experiments (e.g. binary deletion pattern) are described, followed by an analysis and interpretation of the experimental results.

Chapter 7 deals with the MOGA-based model for the phonological change of Bengali verb inflections. The formulation of the problem as a MOGA and the motivations behind the choice of phenotype, genotype and genetic operations are clearly spelt out. This is followed by the descriptions of the various experiments and analysis of the results. A possible application of the model in natural language generation is also described.

Chapter 8 concludes the thesis by summarizing the basic findings. This chapter presents a comparative study of the role of different theories of phonological change in explaining the two problems studied during this work. The extents to which these theories can be computationally modeled and consequently verified are discussed, and the usability of the different computational techniques in modeling different types of explanations is analyzed. The chapter also outlines the possibilities of future research.

An annotated list of publications from this work, as well as the complete list of publications by the candidate are included at the end of the thesis. There are also a few appendices to the thesis that provide various supplementary materials related to the notational conventions and data from the experiments.

1.5.1 A Note on Notation

In this thesis, Indian language scripts are transcribed in Roman script following the ITRANS convention (Chopde 2001). Appendix C provides the tables for transcription of Hindi and Bengali graphemes. The graphemic forms are written in italics. Both Hindi and Bengali uses a phonemic orthography, where the letters usually correspond to a default sound, though some context-dependent variation is also observed. Therefore, we choose the same transcription scheme to represent the phonetic forms as well, where the letters correspond to their default pronunciations. Appendix C also lists the default sounds for each of the letters in International Phonetic Alphabet. The phonetic transcriptions in the text are also in italics and are usually presented within two slashes (//). Gloss for the Indian language words, if provided, are within parentheses. Two examples are shown below.

$$hindI \text{ (Hindi) is pronounced as } \quad /hindI/$$

$$bA.nlA \text{ (Bengali) is pronounced as } \quad /bA \sim NlA/$$

Chapter 2

Background

This chapter provides the linguistic background necessary for understanding and proper grounding of the work presented in the subsequent chapters. We begin with the definition of phonological change, illustrated through examples, followed by a brief overview of evidences used, properties documented and challenges in the study of diachronic phonology (Sec. 2.1). The different explanations of phonological change present in the literature are surveyed in Sec. 2.2. The two instances of real world phonological change, namely schwa deletion in Hindi and phonological change of Bengali verb inflections, are discussed in Sec. 2.3 and 2.4 respectively.

2.1 Phonological Change

Phonological change (also known as *sound change*[1]) is a special case of language change, which refers to "a change in the *pronunciation* of sounds" (Bhat 2001). The change can be conditioned or unconditioned, but only phonological

[1]Strictly speaking, sound change and phonological change refer to different things (Kiparsky 1988). While sound change is change in the pronunciation of a particular sound in a particular context, phonological change refers to a change in the phonological representation. Thus, sound changes are governed by phonetic factors and can be thought of as a source or cause of phonological change. However, many authors use the terms interchangeably, as we shall do here.

features, such as stress, intonation, and neighboring phonemes, can constrain its occurrence. Grammar (i.e. syntax), lexicon, meaning or function of a word has no effect on the change. Nevertheless, the effects of sound change may be altered at a later stage due to other types of changes that occur in the language (e.g., borrowing or syntactic changes).

2.1.1 Instances of Sound Change

It has been observed that sound changes are the first to occur, in almost all the cases of language change that have resulted in bifurcations in the phylogenetic history of languages (Kroch 2001). This makes sound changes prevalent in nature and one of the most well studied phenomena in diachronic linguistics (see Bhat(2001) and Blevins (2004) for review and examples). It has been also observed that, often, several phonological changes occur in sequence resulting in the birth of an altogether new language(s). Such changes are referred to as *shifts*.

Some of the most famous examples of sound changes and shifts include

- *The Great Vowel Shift* that affected the pronunciations of the modern English vowels was accomplished during 14^{th} to 16^{th} century and resulted in changes in the pronunciation of earlier long vowels (Baugh and Cable 1993). For instance, the long $/u/$ in middle English was diphthongized, as in "mouse".

- *The High Germanic Consonant Shift* took place between 3^{rd} and 5^{th} centuries altering the pronunciations of consonants in the Germanic languages (Waterman 1966). This explains the difference in pronunciations between English and German words such as "day" and "tag", "father" and "vater" ($d \rightarrow t$), and "ship" and "schiff" ($p \rightarrow f$).

- *The Loss of Consonant Clusters in Hindi* from their original forms in Sanskrit has resulted in words such as $/rAt/$ from $/rAtri/$ (night), $/Aga/$ from $/agni/$ (fire) and $/dUdh/$ from $/dugdha/$ (milk). Note that there are several other sound changes that took place during these transformations and the change took affect quite early, in Prakrit and Pali, which are precursors of modern Indo-Aryan languages (see Bhat (2001) for more examples of the same). Nevertheless, the Sanskrit word forms are also present in modern Hindi, which were borrowed from Sanskrit at a much later stage.

Sections 2.3 amd 2.4 discusses in detail two particular instances of phonological change in the case of Hindi and Bengali.

2.1.2 Characteristics of Sound Change

There are several interesting properties of sound change, which have been observed and documented in the literature. See Chapter 2 of Bhat (2001) for a review. We summarize below some of the important characteristics.

- **Regularity:** Sound changes are regular in the sense that all the words in the lexicon are affected by the change irrespective of their meaning, function, or social status. Similarly, syntax has no effect on the process of change and its context. Nevertheless, there are several exceptions to this regularity hypothesis. Words borrowed from another language may not reflect the effect of sound change. There are instances of sound change, known as *analogical change* that are conditioned by non-phonological factors. *Hypercorrection* and *sporadic changes* are other examples of irregularities.

- **Directionality:** Language change, in general, has a preferred direction. In other words, if X changed to Y due to some natural linguistic change, then it is unlikely that Y, in some other linguistic system, would change to X. Although the hypothesis of directionality is neither devoid of exceptions, nor criticisms, several observations and linguistic principles suggest that there are certain general tendencies of sound change. For instance, a change resulting in raising of the height of a vowel (say, $/E/$ to $/e/$) in the context of a high vowel (e.g., $/i/$ or $/u/$) is quite commonly observed, for example in Bengali (Chatterji 1926), whereas the reverse of this has never been reported. See Blevins (2004) for more examples of preferred sound changes.

- **S-shaped dynamics:** Several independent lines of research suggest that almost all cases of language change, including sound change, proceed along an S-shaped trajectory[2], also known as the logistic curve (Weinreich et al. 1968; Bailey 1973). The interpretation of this S-shaped dynamics is as follows. Initially, one of the linguistic 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 a few generations, at the end of which the older form is completely driven out by the new variant. Fig. 2.1 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 1989a), French (Fontaine 1985) and Spanish (Fontana 1985). See Kroch (2001) and Briscoe (2002) for further discussions on this.

2.1.3 Evidence of Sound Change

Sound change is not directly observable and therefore, it must be inferred from its effects visible on the sound patterns of a language or genetically related languages. The data used for investigating and subsequently, establishing a particular case of change can be broadly classified as *diachronic* and *synchronic* evidence.

Diachronic evidence is derived from a comparison of two sets of records from the same language belonging to two different periods. The temporal order of the records provides us with a basis (called *external criterion*) for determination of the direction of the sound change. Nevertheless, there are many possible pitfalls associated with the diachronic evidence (see e.g., Meillet (1967)). Firstly,

[2]S-shaped dynamics is also seen in many other physical, social and biological phenomena, and is considered to be a general property of evolution (Modis 2002). In physics, S-shaped dynamics refer to a case of phase transition or bifurcation in the system (Landau and Lifshitz 1994; Ott 2002).

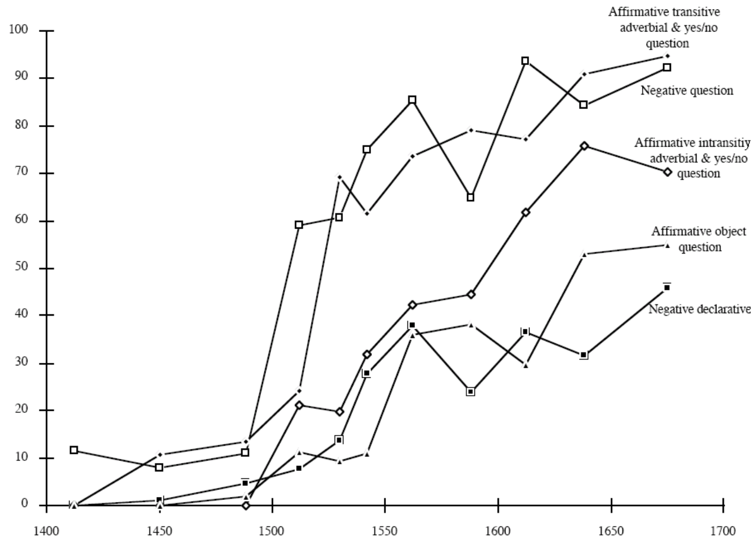


Figure 2.1: The rise of periphrastic *do*. The horizontal axis denotes the date (in year) and the vertical axis denotes the percentage of usage of the new variant found in historical data. (Adapted from Ellegard (1953) as cited in Kroch (Kroch 1989b))

written historical records may not reflect the spoken forms of the same period. For instance, the verb forms used in the Bengali literature of the 19^{th} and early 20^{th} centuries were similar to the spoken language of the Middle Bengali period (1200 – 1800 AD), which had been replaced by their modern counterparts by the beginning of the 19^{th} (Chatterji 1926). Thus, the possibility of shift in the dialect used for writing purposes calls into question the reliability of diachronic evidence. Secondly, gathering a good amount of historical data and assigning proper dates to the records are extremely challenging tasks.

Synchronic evidence, on the other hand, analyzes synchronic data from one or more languages to infer a case of sound change. The issue here, however, is to ascertain the direction of the sound change, which must be done on the basis of some *internal criterion*. Several plausible assumptions regarding phonological changes are used to solve this problem. The most commonly used technique relies on the fact that sound changes cannot introduce new phonological contrast. For instance, the fact that Hindi has long vowels for words where Punjabi fea-

Gloss	Sanskrit	Punjabi	Hindi
work	*kArya*	*kAj*	*kAj*
work	*karma*	*kamm*	*kAm*
thread	*sUtra*	*sUt*	*sUt*
son	*putra*	*putt*	*pUt*

Table 2.1: Synchronic and diachronic evidence for a sound change: the elongation of the vowel in closed syllables in Hindi. Adapted from example 10, chapter 1, Bhat (2001)

tures both short and long forms (see Table 2.1) indicates a change that resulted in elongation of vowel length in Hindi in the closed syllables. The analysis of the corresponding Sanskrit forms (i.e., the older forms) provides further evidence for the change. While the former evidence is synchronic in nature, the latter forms a diachronic evidence for the same case of change.

Synchronic evidence can be *descriptive* (data taken from a single language and grammatical factors used for reasoning) or *comparative* (data taken from two languages) in nature. Nevertheless, for all types of synchronic evidence, identification of the direction of a sound change is a controversial issue and is based on numerous assumptions.

To summarize, one of the challenges in diachronic phonology is the paucity of data, and this is primarily due to the facts that (1) languages do not fossilize, and (2) it is impossible to conduct experiments to gather more historical data. See Meillet (1967), Andersen (1973b), Labov (1975a; 1975b) and Bhat (2001) for a review and criticisms of the different types of evdience used in diachronic linguistics.

2.2 Explanations of Phonological Change

In the previous section, we have seen that phonological changes display a wide range of interesting properties. It is natural to ask why sound changes are regular, have a preferred direction or follow an S-shaped dynamics. Similarly, the actuation problem (Sec. 1.1.2) takes an extra dimension in the case of sound change: while one needs to explain the cause of a sound change, the sound change itself may serve as an explanation for more complex linguistic changes. Thus, if we can solve the actuation problem for sound change, it can provide us with further insights and facts to resolve the actuation problem for the more general case of language change.

There have been several attempts to explain the aforementioned questions by identifying the causes and dynamics of sound change. The explanations can be broadly classified into three categories as described below (also see Ohala (1987) and Chapter 3 of Blevins (2004) for reviews).

2.2.1 Functional Explanations

The functional principle in linguistics maintains that the primary function of language is communication and this fact should be reflected in the structure of the languages and the way it changes over time (see Boersma (1997a; 1997b; 1998) and references therein; also see Liljencrants and Lindblom (1972), Lindblom (1986) and Schwartz et al. (1997) for examples of functional explanations). The basic functional principles are enumerated below (Boersma 1998):

1. The speaker will minimize her articulatory and organizational effort; i.e., she will try to minimize the number and complexity of her gestures.

2. The speaker will minimize the perceptual confusion between utterances with different meanings.

3. The listener will minimize the effort needed for classification, i.e., she will use as few perceptual categories as possible.

4. The listener will minimize the number of mistakes in recognition, i.e., she will try to use the maximum amount of acoustic information.

5. The speaker and the listener will maximize the information flow.

Apart from the above, another functional principle, often referred to in the literature is that of learnability (Roberts 2001; Brighton et al. 2005; Oudeyer 2005a).

These principles are inherently conflicting; minimization of articulatory effort conflicts with minimization of perceptual contrast. Similarly, maximization of information flow seems to be in conflict with the first four principles. The inherently conflicting nature of these forces, therefore, dismisses the possibility of "the optimum language". Stated differently, a language can be always changed in a way to optimize one of the criteria, say articulatory ease, but at the expense of another, such as perceptual contrast. Therefore, lack of "the optimum language" gives rise to a situation where several (possibly an infinite number of) languages co-exist that achieve different trade offs between the functional forces. The languages continuously evolve in order to improve their functionality, but only to reach a different sub-optimal state. This leads to the possibility of circular changes (Boersma 1997b).

Criticisms

Functional explanations of phonology have been strongly criticized by several researchers for their goal-directedness (see for example, Ohala (1983a; 1990b) and Lass (1980; 1997)). Sound changes are the end results of gradual, minute and random changes in articulation, and therefore, are non-optimizing. The speakers and listeners never make a conscious effort to make the language optimized in terms of its functional benefits. There are several instances of phonological changes which are not explainable in terms of functional forces. For example, *metathesis* or swapping of adjacent phonemes is a commonly observed phenomenon. Nevertheless, the resultant form of a metathesis is neither easier to pronounce nor more distinctive. The advocates of functional principles argue that functional explanations are non-teleological, because phonological change does not lead towards a globally optimal goal. Rather, changes result in local optimization of the functional objectives. Lindblom (1998) suggests that although the initial changes are random as proposed by Ohala (1974; 1989), the changes that facilitate certain functional benefits are consciously preferred by the users over those which do not.

2.2.2 Phonetic Explanation

In *phonetic* or *phonetically based explanations*, phonological changes are attributed to the underlying phonetic principles grounded in human articulatory and perceptual mechanisms (see Hayes et al. (2004) and references therein). Like in functional explanations, articulation and perception play an major role here as well, however unlike the former, the changes here are not goal-directed.

Phonetic explanations of sound change based on perception errors have been extensively studied and argued for by John J. Ohala (1974; 1983a; 1987; 1989; 1990b; 1993). For example, in (Ohala 1990b), it has been shown that consonantal place assimilation is a result of "innocent misapprehensions" of the speech signal. Similarly, it has been argued that random articulatory errors lead to synchronic variation and in turn, to diachronic variations.

2.2.3 Evolutionary Explanation

Evolutionary explanations combine both functional and phonetic factors along with the concept of frequency drift (Blevins 2004). The aim of evolutionary phonology is to explain recurrent synchronic sound patterns based on historical, non-teleological and phonetic explanations. However, in the process, it also postulates a set of principles that can explain the nature of sound change.

At the level of a language-user, phonetic principles as well as functional pressures like ease of articulation and the need to be understood (that manifests as perceptual contrast) play an important role in shaping individual language usage. These factors are categorized as CHANGE, CHANCE and CHOICE. While CHANGE refers to a situation where a word/sound is misheard and consequently, (re)analyzed in a different way, CHANCE refers to a situation where a given pattern can be analyzed in multiple ways, and the individual usually selects the analysis that has the highest token frequency. These two factors are clearly reminiscent of the phonetic explanations. For instance, the form /*np*/ might change to /*mp*/ due to misinterpretation (as in Ohala's case for consonantal place assimilation) or it might be interpreted as both /*np*/ and /*mp*/, of which /*mp*/ is selected due to its high frequency. CHOICE models functional principles, where a particular form X is consciously favored by an individual over its variant Y if X facilitates fast speech or Y results in confusion.

Empirical observations make it evident that all sound changes do not reflect the presence of functional or phonetic biases; else how can we explain the co-existence of both the natural (read frequent) and unnatural (read infrequent) patterns of sound change? Evolutionary phonology provides an explanation to this in terms of random drifts. If in a language a particular sound pattern is more frequent than an alternative variant, the language users are expected to decide in favor of the frequently occurring pattern whenever a given input has two or more possible analyses. This triggers a cycle of change that makes the frequent form more frequent in later stages of the language. Thus, a small difference in the frequencies of two types can trigger a sound change in favor of the more frequent pattern irrespective of its functional fitness or phonetic structure. However, the initial difference in frequency can be a result of phonetic factors, and the change may be restrained by the functional forces.

Croft (2000) proposed the *utterance selection model* of language change, which is similar to the aforementioned evolutionary explanation for phonological change. According to the utterance selection model, utterances are analogous to DNA or genes and variation in utterances (i.e., linguistic variables) crop up from the differences in language usage between the speakers and the same speaker over time. Language change is the shift of the probability distribution of the different linguistic variables over time. This shift is postulated as a result of selection, which is a socio-cognitive process.

2.3 Schwa Deletion in Hindi

A major part of this thesis is devoted to the computational modeling of the phenomenon of schwa deletion in Hindi (SDH). In this section, we present a brief overview of the phenomenon, its diachronic roots and related computational works.

2.3.1 Hindi Orthography and Phonology

Hindi, a language of the modern Indo-Aryan family, is written from left to right using the *Devanagari* script. The script is an *alpha-syllabary* derived from Brahmi and consists of 35 consonants, 12 vowels, 15 diacritics and a few more symbols for the commonly used conjugates. The complete set of vowels and consonants of *Devanagari* and their usual pronunciations are listed in Appendix C.

The organizational unit of Hindi orthography is an alpha-syllable (called *akshara* in Sanskrit and Hindi), which is of the form C^*V (C and V denote the consonants and vowels respectively). An *akshara* of the form V is represented by a free vowel, but whenever the vowel follows one or many consonants (as in CV, CCV etc.), its presence is denoted through a diacritical mark drawn around the consonant(s). The first vowel of the *Devanagari* alphabet, *a*, has no associated diacritical mark; rather, it is the absence of any vowel-diacritic in an *akshara* that denotes the presence of *a*. There is also a special diacritic called *halant* that is used to denote the absence of *a*, and is usually used in between consonants to represent a consonant cluster.

Thus, *a* is the inherent vowel of *Devanagari*. In Sanskrit, which is also written using *Devanagari*, all the inherent *a*-s in a word are pronounced. However, Hindi and several other Indo-Aryan languages, such as Bengali and Punjabi, allow optional or mandatory deletion of *a* during pronunciation. Since in Hindi and Sanskrit the vowel *a* is usually pronounced as the *schwa*, the aforementioned phenomenon is referred to as *schwa deletion*.

Schwa is defined as the mid-central vowel that occurs in unstressed syllables. We shall denote this vowel as /*a*/. Schwa deletion is a phonological phenomenon, in which schwas in unstressed syllables are deleted during speech, even though the morphological and etymological evidence assert the presence of the vowel in that particular context (Ohala 1983b). Table 2.2 illustrates this phenomenon for Hindi.

The phenomenon of schwa deletion is not unique to Hindi and is found in languages like French, Dutch, Russian and English. In some languages like

Sanskrit	Hindi	Gloss
sAphalya	*/sAphalya/*	success
rachanA	*/rachnA/*	creation
gagana	*/gagan/*	sky

Table 2.2: Examples of schwa deletion in Hindi

English and French, schwa deletion can be optional and depends on semantics or socio-linguistic contexts (Hooper 1978; Tranel 1999).

2.3.2 The Schwa Deletion Pattern

SDH has been studied from its linguistic perspective by several researchers (Mishra 1967; Pray 1970; Kaira 1976), the most substantial of them being by Manjari Ohala (1977; 1983b). On the basis of several experiments and morphological evidence from the language, Ohala showed that schwa deletion is an important feature of Hindi phonology. The context for SDH has been summarized in (Ohala 1983b) as follows.

$$a \rightarrow \phi \,/\, \mathrm{VC}_1^2 \text{---}\ \mathrm{CV}/ \quad \left\{ \begin{array}{c} \left[\begin{array}{l} +\ \text{loan} \\ +\ \text{casual speech} \end{array} \right] \\ [\, +\ \text{normal tempo}\,] \end{array} \right\}$$

Condition 1 : There may be no morpheme boundary in the environment to the left.

Condition 2 : The output of the rule will not violate the sequential constraints of Hindi.

Convention : The rule applies from right to left.

The rule states that schwa is deleted in a context where it is preceded by a vowel and one or two consonants, and followed by a single consonant and vowel. *Sequential constraints* tell us which consonant clusters are considered valid in the words of a language. An alternative term for this is *phonotactic constraints.* Strangely, Ohala's rule does not capture the context for the deletion of word final schwas (eg. the case of *gagana* in Table 2.2). Probably, the very existence of the word final schwas in Hindi, which is clearly indicated in the orthographic form of the word, is denied by Ohala due to its psycholinguistic uncertainty. However, as illustrated in example 51, chapter 3 of Bhat (2001), Hindi lost the word final schwa during the course of its development, which was otherwise present in Middle Indo-Aryan.

To capture the word final schwa deletion we propose the following modification of the Ohala's rule, where $ represents the end of a word.

$$a \rightarrow \phi\,/\mathrm{VCC?}\text{—}\ \{\$, \mathrm{CV}\} \tag{2.1}$$

Note that the conditions 1 and 2 of Ohala's rule regarding morphology and phonotactic constraints also apply to the modified rule stated above. The following example illustrates the application of Ohala's rule on a few Hindi words.

Example 2.1 Let us consider the Hindi words *rachanA* "creation" and *gagana* "sky" (refer to Table 2.2). The corresponding C-V patterns for these words are CVCVCV$ and CVCVCV$. As per the modified schwa deletion rule stated in equation 2.1, only the second schwa in *rachanA* is delible, deletion of which results in the surface form /*rachnA*/. This indeed is the usual pronunciation of the word in standard Hindi. In case of *gagana*, the contexts of both the second and third schwas conform to the deletion rule. However, deletion of the third schwa gives rise to the pattern CVCVC$, which prevents further deletion of the second schwa. On the other hand, deletion of the second schwa results in the pattern CVCCV$, where further deletion of the last schwa is possible. Nevertheless, the deletion leads to the surface form /*gagn*/ which violates the phonotactic constraints of Hindi (/*gn*/ is not allowed word finally).

Thus, we are left with two possibilities /*gagan*/ and /*gagna*/, and the right to left convention stated in Ohala's rule decides in favor of the former. It should be noted that the two possible surface forms for the word *gagana* arise only due to the postulation of the word final schwa. Since Ohala's rule does not consider the presence of this word final schwa, the ambiguity between the two possible surface forms does not arise at all. However, in case of words like *ajanabi* (stranger), two possibilities arise even for Ohala's rule, which can be resolved using the convention. □□

2.3.3 Diachronic Nature of SDH

It is a well established fact that SDH has evolved through a diachronic process (Mishra 1967; Ohala 1983b). The middle vowel schwa that occurs mostly in unstressed syllables is often deleted or added (schwa-epenthesis) during pronunciation in almost all the languages that have this phoneme. This high susceptibility to deletion or addition is due to the neutral nature of the vowel. Deletion of schwas lead to minimization of syllables (Tranel 1999). For example, the word $saradAra$ "chief", if pronounced as $/sa - ra - dA - ra/$ has four syllables ('-' indicates syllable boundaries); but in Hindi, it is pronounced as $/sar - dAr/$ and thus have only two syllables. The lesser the number of syllables in a word, the smaller is the time to pronounce it, though the relation may not be linear. Therefore, schwa deletion enables faster speech through reduction in the number of syllables. In fact, it has been found that in Hindi, schwa deletion is much more prevalent in *faster* and *informal speech* than in formal style of speaking (Ohala 1983b).

Although a need for faster communication can explain why schwas are deleted, it alone cannot explain why schwas are deleted only in particular contexts. To explain the specific schwa deletion pattern observed in Hindi, it is necessary to identify the forces that act against syllable minimization and prevent the deletion of schwas in certain contexts. One such force, for example, is the pressure to maintain the lexical distinctions; deletion of schwas in any arbitrary context might remove the distinction between two words leading to a loss of communication. Nevertheless, we do not know of any work on explanations for SDH in Hindi, apart from the aforementioned functional account based on syllable-minimization.

2.3.4 Computational Models for Schwa Deletion in Hindi

Schwa deletion is one of the most important phenomena in Hindi phonology and an algorithmic solution to SDH is essential for developing G2P converter for Hindi TTS (Sen and Samudravijaya 2002; Kishore and Black 2003; Narasimhan et al. 2004). Narasimhan et al. (2004) describes an algorithm for schwa deletion in Hindi based on Ohala's rule. In this algorithm, word-final schwa deletion has been captured using a cost model, which otherwise is not dealt with by Ohala's rule. The algorithm also uses a morphological analyzer for identification of morpheme boundaries, because schwa deletion in Hindi respects morpheme boundaries (*condition 1*). The accuracy of the algorithm, measured in terms of the fraction of schwas correctly deleted/retained, has been reported to be 89%. Kishore and Black (2003) mentions the problem of SDH as the hardest part of Hindi G2P converter and describes certain heuristics to model the phenomenon. However, according to the authors, these heuristics do not solve the problem completely.

Thus, several aspects of this phenomenon make it an excellent example for computational modeling. Some of them are listed below.

- The diachronic roots as well as synchronic validity of this problem are well established.
- The context for SDH can be described by a simple rewrite rule (Ohala's rule), which has hardly any exception. This pattern can serve as the acid test for a model of SDH, because the model must be able to account for this pattern.
- There exists a functional explanation for SDH.
- Apart from Hindi, languages like Punjabi and Bengali also feature schwa deletion, but the pattern differs from language to language. Several dialects of these languages exhibit further variation in the deletion pattern. It would be interesting to explore how a single framework for change can capture the emergence of these variable patterns, and what are the circumstances, under which specific deletion patterns evolve.

2.4 Phonological Change of Bengali Verb System

While SDH makes an interesting study of sound change, it is quite simple in nature. In general, several simpler sound changes occur in sequence to give rise to a complex pattern of change or shift. These changes often give rise to a set of new dialects or languages (e.g., the great vowel shift affected several Germanic languages including English, Dutch and German). The second case of a real world phonological change, for which we attempt to develop a computational model in Chapter 7, is an example of a complex sequence of sound changes. This change, which affected the morpho-phonological structure of Bengali verb inflections[3] (BVI), was accomplished during 13^{th} to 18^{th} century. The change gave rise to more than 20 different dialects of Bengali, spoken over parts of India and Bangladesh. In this section, we present a brief overview of the verb inflections in Bengali and their history.

2.4.1 History of Bengali

Bengali, a language spoken in the eastern region of India and Bangladesh, is a member of the Indic group of the Indo-Iranian or Aryan branch of the Indo-European family of languages. In his book – *The Origin and Development of the Bengali Language* (Chatterji 1926), Suniti Kumar Chatterji provides a detailed description of the history of Bengali and evolution of the current phonological and morphological features of the language. The origin of modern Bengali can be traced back to Vedic Sanskrit (circa 1500 BC 600 BC), which during the middle Indo-Aryan period gave rise to the dialects like *Magadhi*, and *ArdhaMagadhi* (circa 600 BC 200 AD), followed by the *Magadhi-apabhramsha*, and finally crystallizing to Bengali (circa 10th century AD). The verbal inflections underwent significant phonological changes during the middle Bengali period (1200 - 1800 AD), which gave rise to the several dialectal forms of Bengali, including the present standard form of the language known as Standard Colloquial Bengali (SCB).

However, as stated earlier, the Bengali literature of the 19^{th} century was written in the Classical Bengali dialect or the *Sadhubhasha* that used the older verb forms and drew heavily from the Sanskrit vocabulary, even though the forms had disappeared from the spoken dialects by 17^{th} century. Here, we shall take the liberty to use the term "classical forms" to refer to the dialectal forms

[3]The change also affected other words, however we concentrate here on the verb inflections, mainly because of the regularity of the pattern

of middle Bengali (i.e., the spoken forms of 17^{th} century) and not Classical Bengali of the 19^{th} century literature.

2.4.2 Verbal Inflections in Bengali

Bengali has a rich and productive verb morphology, where suffixes denoting the tense, aspect, modality, person and honorific value are added to the verb root. Although the inflectional system is mostly *agglutinative* in nature, there are cases where the same suffix stands for different persons in different tenses (e.g., '*a*' in the future tense marks 1^{st} person, as in *karaba /korbo/*, classical form: *kariba*, but in the present tense it stands for the 2^{nd} person, as in *kara /karo/*, classical form: *karaha*). Therefore, we shall treat the suffixes as a single fused unit, rather than a concatenation of smaller suffixes marking for tense, aspect etc. Going by this, any verb in Bengali has 49 finite forms, two non-finite forms and one nominal form. The complete list of inflectional suffixes for SCB and Classical Bengali are presented in Appendices D.1 and D.2.

The verb roots of SCB can be classified into paradigms based on the syllable structure. The roots in the same class undergo similar orthographic changes during suffixation. A 12-class formal classification system for the verb roots of SCB is proposed in (Chatterji 1926). In order to develop morphological analyzer and generator for Bengali, we have extended this classification system into a 19-class system, which is shown in Appendix D.3.

2.4.3 Derivation of Verb Inflections

The verb inflections of the middle Bengali period have undergone a sequence of sound changes, including, but not limited to, deletion, assimilation and metathesis. The derivations of almost all the verb forms of SCB (and occasionally, of the other dialects) from their classical counterparts are documented in (Chatterji 1926). Sometimes the derivations from the Sanskrit (i.e. Old Indo-Aryan) or Prakrit (Middle Indo-Aryan) forms are also shown. These example derivations provide sufficient information for constructing the derivations of all the 52 forms of SCB for all the 19 morphological paradigms.

OIA 600 BC	MIA 200 AD	Classical 1200 AD	Intermediate 1200–1800 AD			SCB
karomi	*karimi*	*kari*				*kori*
		kariteChi	*kariChi*	*kairChi*	*koirChi*	*korChi*
karantahi	*karante*	*karite*	*kairte*	*koirte*		*korte*
		kariA	*kairA*	*koirA*		*kore*

Table 2.3: Derivations of some representative verb forms of SCB from older forms. Legends: OIA: Old Indo-Aryan, MIA: Middle Indo-Aryan, Classical: Middle Bengali Period. All the words are in phonetic forms.

Table 2.3 shows the derivations of a few representative forms of the verb /*kar*/. The corresponding derivations in another dialect can be constructed likewise, as illustrated below for /*kartAslo*/, the counterpart of /*korChilo*/ in a dialect spoken in Agartala (the capital of the state of Tripura, situated in north-eastern part of India). We shall refer to this dialect as Agartala Colloquial Bengali or ACB, for short.

$$karuthilA \gg kariteChila \gg kartesila \gg kartesla \gg kartAslo$$

We defer the discussion on the details of the phonological changes that affected the BVI to Chapter 7.

2.5 Summary

In this chapter, we have seen that sound change or phonological change is the most basic type of linguistic change that is frequently observed in nature. Sound changes are regular, have a preferred direction and follow an S-shaped dynamics. Sound changes have been extensively studied in diachronic linguistics, but there are several issues regarding the actuation (i.e., explanation) of sound change, which are not very well understood. Researchers have put forward functional, phonetic and evolutionary models to explain the origin and nature of sound change. However, none of the theories are devoid of limitations.

The prime reason behind the disagreement between the different schools of researchers in this area is the paucity of historical data and impossibility of experimentation. At the same time, being the foundational step of any diachronic process in linguistics, understanding the dynamics of sound changes is of prime importance to diachronic linguistics. Furthermore, if one accepts the cultural transmission theories of language evolution, then the nature of sound change can also provide us with insights regarding language evolution, which is arguably one of the hardest problems of modern science (Hauser et al. 2002; Christiansen and Kirby 2003b).

In the next chapter, we shall see that computational models can, at least partially, address the issues of data scarcity and empirical validation in diachronic linguistics. We survey the previous works that attempted to resolve the theoretical issues related to language change, using computational techniques. We also discuss their limitations, which we address in this thesis by constructing models for the two cases of real world sound changes discussed in the earlier sections.

Chapter 3

Related Work

Computational modeling techniques are being used in the field of diachronic linguistics for over the past fifty years. However, these techniques gained popularity, rather recently during the last two decades. The Language Evolution and Computation Bibliography website[1] hosted by the UIUC Agents and Multi-Agent Systems Group, contains around 1200 references (as on May 2007) pertaining to works on computational models of language evolution. Some of these works also address issues related to language change and many of them are indirectly relevant to diachronic linguistics. Therefore, the site provides important information and resources for the study of computational models of language change.

Fig. 3.1 shows the number of publications per year listed in the Language Evolution and Computation Bibliography site since 1960. Although there has been some isolated works in the past, the field has gained momentum from the 90s. The number of publications in this area has almost doubled during 2005-2006 as compared to pre-2000 publications, and the curve is exhibiting an exponential growth pattern. All these facts point towards the growing popularity of the computational models as well as their impact in the study of language dynamics.

In this chapter, we present a brief survey of the computational models of language change. Since there has been several surveys on the computational

[1] *http://www.isrl.uiuc.edu/∼amag/langev/*

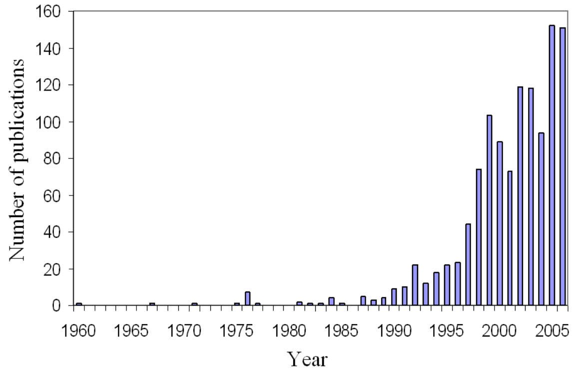


Figure 3.1: Bar chart showing the number of publications per year listed in the Language Evolution and Computation Bibliography site

models of language evolution (Steels 1997b; Perfors 2002; Christiansen and Dale 2003; Wagner et al. 2003; Wang et al. 2005; de Boer 2006), here we do not discuss at depth the models of evolution. Nevertheless, some of the works that are directly relevant to the thesis are covered in this survey.

In the context of diachronic linguistics, computational techniques can be employed to construct the phylogenetic tree of languages and infer the genetic relationships between languages and language families (also known as reconstruction; see Nakhleh (2005) for a review; also see Lowe and Mazaudon (1994) and the references therein), or to explain the phenomenon of language change. The topic of the thesis, and consequently the scope of this survey, is restricted to the models pertaining to the second problem.

In the next section, we discuss some of the previous classifications of computational models of language evolution (or change) and propose a new classification system that covers a wider range of computational as well as mathematical models in this area. The rest of the survey is organized according to this proposed taxonomy.

3.1 A Three-dimensional Classification System

The inter-disciplinary nature of the models demands a multi-dimensional classification system. Since the models, techniques, and experiments in this field have both linguistic and computational aspects, they fit well in the conventional taxonomy of these disciplines. At the same time, the issues involved in language change and the nature of explanations provided by a model calls for its own classification system. Thus, given a computational model of language change, one can ask the following questions regarding the nature of the model:

- What kind of *linguistic explanation* the model intends to validate? This we shall call the *objective of modeling*.
- What kind of computational or mathematical tools are used to construct the model? This we shall call the *technique of modeling*.
- Which level of linguistic organization (e.g., phonology, morphology, lexicon and syntax) the model is based on? This we shall call the *domain of modeling*.

The survey by Steels (1997b) addresses all of the aforementioned dimensions in the context of language evolution. Nevertheless, the scope of the survey is limited to only the MAS based models. Similarly, Niyogi (2006) provides a classification of the dynamical system models of language change following the nature of language transmission that has been incorporated in the models. Thus, the works are classified on the basis of the learning algorithm, and the cultural and spatial transmission factors that have been considered. However, in this book MAS models are not discussed at depth. In fact, as far as our knowledge goes, there is no comprehensive survey that covers the simulation-based models vis-à-vis the analytical and mathematical ones.

3.1.1 Objective of Modeling

In chapter 2 we have discussed the various linguistic theories of language change with special emphasis on phonological change. Computational models can be constructed to (in)validate any or all of these theories. To illustrate the concept, let us take the case of the *vowel inventories*. Linguists have documented several strikingly similar cross-lingual patterns observed across the vowel inventories of the world's langauges (Maddieson 1984; Ladefoged and Maddieson 1996). For example, the three vowel inventories are usually formed from the vowels $/A/$, $/i/$ and $/u/$. A functional explanation for the observed regularities in the vowel inventories beg the concept of perceptual contrast (Liljencrants and Lindlom 1972), whereby it is claimed that the vowels try to spread themselves maximally over the perceptual space. Several computational models have been proposed to validate this functional explanation. Liljencrant and Lindblom (1972) used numerical simulations to show that the optimal configuration of vowel inventories as per the distinctiveness criterion is similar to those observed in the real world. Schwartz et al. (1997) further refined the model (i.e., the precise formulation of the optimization criterion), so as to obtain vowel systems that are closer to the real ones.

de Boer (1997; 1998; 1999b; 2000a; 2000b; 1999a; 2001) proposed a MAS model for explaining the structure of vowel inventories, which does not encode any global optimization strategy. Thus, the model is not a functional one, but puts forward a self-organization based or emergent explanation of the vowel systems. Since in the model the articulatory and perceptual mechanisms of the agents have been modeled with some detail, the explanation can also be considered to have a phonetic basis. Joanisse and Seidenberg (1999) uses neural nets to show that the so-called optimal vowel systems are easier to learn and Ke et al. (2003) uses multi-objective genetic algorithms to explain the same.

In the specific context of language change, the objectives of a model could be to

- provide a functional explanation for a specific case of change by identification and formulation of the functional forces and solving the resulting optimization problem
- provide an emergent explanation by showing that the change could be a naturally emergent behavior of the linguistic-system under specific circumstances

- explain the problem of actuation

3.1.2 Technique of Modeling

Given an objective (i.e., explanation) and domain (i.e., problem) for modeling, a variety of computational approaches can be adopted to realize the modeling process. Broadly, the approaches to modeling can be classified into two classes – the *analytical* and the *synthetic* techniques.

In the analytical approach, the hypothesis to be tested is first modeled as a formal problem. Depending on the technique followed, at the end of formulation step one might arrive at a difference or differential equation, an optimization problem, or a more complex mathematical structure such as a graph or network. Note that all these models have some associated free parameters (e.g., coefficients in equations, relative weights of the objective functions in an optimization problem, the number of nodes, edges etc. in a graph like structure and so on). The next step is to use standard mathematical techniques or computational method to analyze the model and find out the set of values for the parameters that makes the model behave like its natural counterpart. In the case of equations the process thus boils down to solving them through numerical or analytical methods, whereas optimization problems are typically solved using soft computing techniques such as genetic algorithms, simulated annealing and numerical simulation.

If it is possible to show that for certain configurations of the model parameters the behavior of the model resembles the natural systems, then those particular parameter configurations are considered as the necessary or sufficient conditions for the hypothesis to be valid. The analysis can then be taken one step further by identifying the physical correlates of the model parameters, so that one can precisely point out under what natural circumstances the case of change and its explanation being studied are plausible.

Synthetic techniques, on the other hand, model the situation as an interaction of entities (i.e., agents), where the design of the agents, their interaction patterns depend on the hypothesis to be modeled as well as other simplifications made to maintain computational tractability of the model. Usually there are a large number of model parameters governing the cognitive and social processes of the agent society. The model is explored systematically by running simulation experiments under controlled conditions (i.e., desired parameter settings) and thereby, observing the emergent behavior of the model. If it is observed that for several runs of the simulation experiments under the same parameter settings the behavior of the model resembles the natural system, then one can conclude that the particular parameter settings are sufficient or necessary for the emergence of the behavior. As in the case of analytical models, the physical significance of these parameter settings can then be further investigated.

The conceptual difference between the synthetic and analytical approaches is similar to that between the inductive and the deductive techniques. This is because, in the analytical approaches, if one assumes that the formulation of the model is correct, then one must accept the conclusions arrived at from the analysis of the model. To the contrary, the synthetic approach is similar to an inductive technique, because even if the model is correct, the results cannot be taken for granted, because for different runs of the simulation the results are different. If several runs of the simulation show the emergence of one particular behavior, then only under the inductive generalization it can be concluded that the model always entails that behavior.

The major difference between these two approaches, however, springs up from a practical reason; while it might be possible to encode an agent based stochastic model into a differential equation, the equations are unsolvable unless one makes several assumptions (e.g., the number of agents tends to infinity, generations of agents do not overlap). Some researchers have shown that the model obtained after such simplified assumptions may yield significantly different results from the original one. This casts serious doubts on the conclusions of analytical models. The synthetic models retain their stochastic and finiteness properties during experimentation and thus, are more trustworthy. However, a large number of model parameters that are usually associated with the synthetic models along with a large simulation time make it hard to explore completely.

An alternative way to view the analytical and synthetic approaches is through the macro and micro distinction. A macroscopic description of the system views the system as a black-box with some general global properties, whereas a microscopic view of the system considers the system as a conglomeration of interacting entities, where the general global characteristics at the macroscopic level are emergent properties of the microscopic description. This distinction is quite commonly made in thermodynamics and other branches of physics, where temperature, pressure etc. are considered as macroscopic properties of a system, which emerge through the microscopic interactions of molecules and atoms.

Clearly, synthetic approaches subscribe to the microscopic view of the language, whereby the properties of the language emerge or self-organize due to the interactions between the agents (speakers). The analytical approaches, on the other hand, view the linguistic system as a whole (i.e., at a macroscopic level) because the microscopic view does not yield to mathematical analysis. Language or any physical system can also be viewed from an intermediate mesoscopic level, where the entities and their interactions are modeled (usually as nodes and edges respectively of a network), but the precise nature of the agents/interactions are abstracted out.

3.1.3 Domain of Modeling

Computational models have been proposed to explain cases of linguistic change at different levels of linguistic structure. The inherent complexities of languages do not allow us to model a language and its dynamics over time to entirety. Consequently, the successful models are those that address a phenomenon at a particular level of linguistic structure, such as phonetics, phonology, morphology, syntax, lexicon and semantics. A domain-wise description of the models are presented in Sec. 3.4.

Since the objective of this thesis is to show that it is possible to build computational models of real world language change for the different types of linguistic explanations, this survey has been organized primarily along the dimension of "technique of modeling". Thus, Sec. 3.2 and 3.3 describe the various analytical and synthetic techniques of modeling. We also give some examples of computational models from the different domains of linguistics in Sec. 3.4. Since the different problems in diachronic linguistics and their explanations are already discussed in Chapter 2, we do not elaborate here along the dimension of "objective of modeling". Nevertheless, while discussing the particular models, we touch upon the objective wherever appropriate.

3.2 Analytical Techniques

Analytical techniques can be classified as optimization based models and other mathematical models. The mathematical models are usually inspired by similar treatments in statistical physics and information theory. Here, we have categorized the analytical models into three classes:

- Dynamical system based models, where the goal is to formulate language as a dynamical system and obtain the rate equation for temporal evolution of the system.
- Statistical physical models, which includes models inspired by statistical physics. While most of them are directly based on a particular model of statistical physics (say the Ising model), the dynamical system models do not maintain any such analogy or assumptions. Note that this differentiation has no theoretical consequences, and only reflects the modelers perspective.
- Optimization models, where the problem of linguistic change is stated as an optimization problem. Note that these models are theoretically different from the two aforementioned types.

We discuss each of these modeling techniques briefly and explain the basic concepts with some examples.

3.2.1 Dynamical Systems

We discuss three examples of dynamical system models of language change to illustrate the modeling principles. References to other similar works are also cited along with, even though we do not elaborate on them. All these models are inspired, to various extents, by similar treatments in population genetics; see (Ewens 2004) and references therein, also see (Cavalli-Sforza 2000).

Abrams and Strogatz model of language death

This is a simple model of language competition and death proposed by Abrams and Strogatz (2003). The objective of the model is to explain the pattern of decline in the number of speakers in the minority languages finally leading to the extinction of the language.

The authors consider a system of two competing languages X and Y. Let x be the fraction of speakers speaking X and $y = 1 - x$ be the fraction of speakers speaking Y. The attractiveness of a language depends on the number of speakers of the language and its relative status in the society. Let s be the relative status of the language X and $P_{yx}(x, s)$ be the probability per unit time that a speaker of language Y will start speaking the language X. Under the aforementioned assumptions, the dynamics of the linguistic system can be expressed by the following rate equation.

$$\frac{dx}{dt} = (1 - x)P_{yx}(x, s) - xP_{xy}(x, s) \tag{3.1}$$

In words, the rate of change of the number of X-speakers is the difference between the rate of the number of Y-speakers turning to X and X-speakers turning to Y. Note that $P_{xy}(x, s) = P_{yx}(1-x, 1-s)$. A few simple assumptions regarding the boundary conditions as well as the form $cx^a s$ for $P_{yx}(x, s)$ helps one to solve the value of x as a function of time, t. The authors show that for several examples of near-extinct real languages – Scottish Gaelic of Sutherland, Quecha of Peru and Welsh in Wales – the predictions made by the model about the number of speakers matches the real data for suitable choice of the model parameters s, a, c and $x(0)$ (the initial condition).

This is a simple example of a dynamical system model of language change, where languages compete for speakers; thus, the model subscribes to a macroscopic view of language. Interestingly, according to the Abrams and Strogatz model there are two stable fixed points at $x = 0$ and $x = 1$, which implies that there is no possibility of an equilibrium with two coexisting languages. Mira and Paredes (2005) extended the Abrams and Strogatz model by introducing the concept of similarity between the two competing languages, thereby showing that a stable bilingual configuration is indeed possible. The authors validated their model against real data from two coexisting languages – Castillian

Spanish and Galician, both spoken in northwest Spain. Some other models investigating language competition in a similar framework are Patriarca and Leppanen (2004), Tesileanu and Meyer (2006) and Pinasco and Romanelli (2006).

Niyogi and Berwick model of language change

A slightly more detailed model of language change as a dynamical system is offered by Niyogi and Berwick (1996; 1997c; 1997a; 1997b; 1998). In this model, a linguistic system is defined as a 3-tuple: $\langle \mathcal{G}, \mathcal{A}, \mathcal{P} \rangle$, where $\mathcal{G}$ is the set of possible grammars or languages, $\mathcal{A}$ is the learning algorithm that takes as input a set of sentences produced accoring to a grammar and outputs a grammar $g \in \mathcal{G}$, and $\mathcal{P}$ is the probability distribution with which the input sentences are presented to a learner.

From the description of the language at the level of individual, the authors propose a formulation of the language dynamics at the population level under three simplified assumptions of (1) non-overlapping generations, (2) same input distribution to all children and (3) fixed grammars for adults (i.e., adults do not change their grammars). Let $s_p(i)[g]$ be the fraction of population in generation i that speak according to grammar g, and let $s_p(i)$ be the corresponding distribution function over $\mathcal{G}$. Then clearly, $s_p(i+1)$ is a function of $s_p(i)$, $\mathcal{A}$ and n – the number of input examples presented to a child.

Language change in this model is change of the distribution $s_p(i)$ over time i. Note that language acquisition is considered as the only factor affecting language structure, and therefore, the objective of the Niyogi and Berwick models is to validate the acquisition based accounts of language change as proposed in (Lightfoot 1991; Lightfoot 1999). The models have been solved for various learning algorithms such as the trigger learning algorithm (Gibson and Wexler 1994), Galves batch learning (Niyogi and Berwick 1998) and the dynamics obtained has been compared with real data. For example, in (Niyogi and Berwick 1998) the authors consider the case of change from Classical Portuguese to European Portuguese and show that with a batch learning algorithm, the dynamics is given by the following rate equation.

$$\alpha_{i+1} = 1 - (1 - \alpha_i p)^n \tag{3.2}$$

Here, α_i represents the fraction of peole speaking Classical Portuguese in the i^{th} generation and p is the probability with which proclitics are used in Classical Portuguese. Similar models have been used to explain the phonological merger in Wenzhou province of China and syntactic change in French (Niyogi 2006).

Some of the other works in this framework pertaining to language evolution are: Nowak et al. (2001), which models the evolution of the Universal Grammar; Komarova et al. (2001), which formulates an equation for grammar acquisition and analyzes the same to find out necessary constraints on grammar acquisition process (such as error tolerance limit); Komarova and Nowak (2003), which extends the (Nowak et al. 2001) model for stochastic processes and finite population; Komarova and Niyogi (2004), where authors explore the conditions for convergence to a shared vocabulary in a group of agents. The last work is particularly important from the perspective of language change, because authors show that two languages can be mutually intelligible, even though they might be quite different. This implies that a language can change over a few generations without breaking down the communication flow between them, and therefore, can be a plausible answer to the problem of actuation.

Utterance selection model of language change

Baxter et al. (2006) offers a mathematical treatment of the utterance selection model (see Croft (2000) and Sec. 2.2.3), whereby the discrete rate equation is approximated by a Fokker-Plank equation under the continuous time assumption. The equation is exactly solvable for a single speaker and has been analytically investigated for the case of multiple speakers. One of the interesting observations made from the model is that extinction of a linguistic form in case of multiple speakers take place in two stages. In the first stage, all the speakers quickly converge to a common marginal distribution. The ultimate extinction of the form takes place in the second stage after a long time due to some random fluctuations.

3.2.2 Statistical Physics Models

The view of language as a complex adaptive system immediately allows one to apply the theories and techniques of physics to analyze the dynamics of a linguistic system. This line of research has been pioneered by the works of M. Gell-Mann (1992; 2005), who is better known for his works on physics of complexity (Gell-Mann 1995). The dynamical system based analyses of languages that have been discussed in the previous subsection employ tools and techniques from physics. In this subsection, we look at a few more models of language change inspired by the current theories of statistical mechanics such as the *Ising model* and *complex networks*.

Ising model of language change

"The Ising model is a mathematical model in statistical mechanics. It can be represented on a graph where its configuration space is the set of all possible assignments of $+1$ or -1 to each vertex of the graph. A function, $E(e)$ is defined, giving the difference between the energy of the bond associated with the edge when the spins ($+1$ or -1) on both ends of the bond are opposite and the energy when they are aligned." (Wikipedia 2007). The Ising model finds several applications in statistical physics ranging from explanations in magentization and superconductivity to models of simple liquids.

Itoh and Ueda (2004) proposes an Ising model for change of word-ordering rules. Typological studies on word-ordering rules reveal that the languages of the world can be classified into two groups: prepositional and other (postpositional or adpositionless) languages. The word order parameters depend on whether a language is prepositional (Greenberg 1963) and it seems that the languages in the world fluctuate between the two stable structure of word ordering rules. In the Itoh and Ueda model, each word order parameter is assigned a value between -1 and $+1$ such that the parameter values for Japanese are assigned $+1$ (an arbitrary choice) and those opposite to them are assigned a value of $+1$. A ternary interaction model with 8 particles with a small mutation rate is found to display characteristics similar to that of real languages.

The authors also construct graph, where a node represents a language and the weight of an edge is defined as the Manhattann distance between the vectors (a 66-dimensional vector is constructed for each language for based on the word-ordering parameters) of the two nodes connected by the edge. The change of word ordering rules is simulated by a random walk on the graph, where the next node in the walk is chosen randomly from the set of nodes that are within a predetermined Manhattann distance from the current node. The ternary interaction model as well as the random walk suggests that languages for which all the word-ordering parameters have the same sign are stable, and there are occasional fluctuations between these two stable states.

Equilibrium statistical mechanics model

Kosmidis et al. (2006b) suggests the use of equilibrium statistical mechanics to study the properties of natural languages such as Zipf law, growth of vocabulary in children and reduced communication abilities by the schizophrenics. According to this model, if an individual has a vocabulary of N words, then the associated language faculty in the brain is assumed to be in one of the N possible states. The brain is in the ith state when the individual utters the word i. Temperature of the system is a measure of the willingness or ability of the users to communicate. Based on a particular *ansatz* for the Hamiltonian of the system, the authors derive several universal properties of languages. Although not directly related to language change, the work makes interesting predictions about language acquisition and vocabulary growth, and demonstrates the application of techniques from physics in linguistics. Some other models that directly employ tools of physics are (Ferrer-i-Cancho 2005a; Kosmidis et al. 2005; Kosmidis et al. 2006a; de Oliveira et al. 2006; de Oliveira et al. 2006).

Complex network based models

In statistical mechanics, physical systems are often modeled as a collection of large number of interacting entities, where the entities are the nodes and their interactions are represented as edges of a large network, commonly referred to as a complex network. In recent times, complex networks have been successfully employed to model and explain the structure and organization of several natural and social phenomena, such as the foodweb, protein interaction, WWW, social collaboration, scientific citations and many more (see Barabasi (2002) and Newman (2003) and references therein for a review of complex networks and their applications).

In the context of languages, complex networks have been used to model the relationship between the words and concepts (Ferrer-i-Cancho and Sole 2001; Dorogovtsev and Mendes 2001; Sigman and Cecchi 2002; Motter et al. 2002; Holanda et al. 2004), the mental lexicon (Luce and Pisoni 1998; Tamariz 2005; Kapatsinski 2006), the structure of syntactic dependencies between words (Ferrer-i-Cancho and Sole 2004; Ferrer-i-Cancho 2005b) and numerous other applications to language modeling (Hudson 2006) and NLP (see, e.g., Biemann (2006), Pardo et al. (2006), Antiquiera et al. (2007)). Although none of these works are directly relevant to diachronic linguistics, they provide useful insights into language evolution. The topological properties of the linguistic networks (e.g., degree distribution, clustering and mixing patterns) reflect linguistic universals and variations, whereas the dynamical processes acting on the network as well as the evolutionary dynamics of the system that lead to the observed topologies correspond to the processes of language evolution and change.

Some of the more direct applications of network models in the field of diachronic linguistics pertain to the simulation of language competition or variation on social networks (Nettle 1999; Ke 2004; Gong et al. 2004; Dall'Asta et al. 2006a; Dall'Asta et al. 2006b) and modeling of language acquisition (Ke and Yao 2006). It is interesting to note that while a majority of the multi-agent simulation models assume that the probability of interaction between any pair of agents is the same (e.g., Steels 1996; Boer 2000; Harrison et al. 2002), variations are typically observed in agent-based systems when this probability is skewed. In other words, agents are spatially grounded and communicate more frequently with their neighbors. The interaction pattern in the former case can be modeled as a clique (agents are nodes, edges denote interaction), while the latter cases are modeled as single or many dimensional lattices (Livingstone 2002; Patriarca and Leppanen 2004) or more complex small world networks (Ke

2004; Gong et al. 2004; Dall'Asta et al. 2006a; Dall'Asta et al. 2006b). Lee et al. (2005) systematically explores the role of the population structure on language evolution and variation.

3.2.3 Optimization Models

Constrained optimization is a natural choice for modeling functional explanations of language change because language is viewed as an optimizing system under the functionalist view (Boersma 1998; also see Sec. 2.2.1 of this thesis). As an example, let us consider the problem of explaining the universal principles observed across the vowel inventories of the languages all over the globe (Ladefoged and Maddieson 1996). 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 principle of maximal perceptual contrast, which states that higher the perceptual distance among the vowels of an inventory, the easier it is to perceive and learn the vowels. Liljencrants and Lindblom (1972) uses numerical simulations for solving the optimization problem and the vowel systems obtained through the process closely resemble the naturally occurring ones. The model has been refined by later researchers (Crothers 1978; Lindblom 1986; Schwartz et al. 1997), which also employ numerical simulations for carrying out the optimization process.

All the aforementioned models of vowel systems, however, are based on a single optimization criterion. In some of the models (e.g., Lindblom 1986) multiple optimization criteria have been proposed, though they are then combined in certain ways to yield a single optimization criterion. The problem with a single objective optimization model is that it yields a unique optimal solution – clearly a disadvantage for modeling linguistic systems, which display a wide range of variation. Ke et al. (2003) proposes a multi-objective optimization model for explaining the structure of vowel and tonal systems. In this model, Genetic Algorithm (GA) (originally proposed by Holland (1975) as a *natural selection* technique mimicking the biological evolution), or more specifically Multi-objective GA (MOGA) (Goldberg 1989) has been used as the optimization tool. In MOGA, the concept of optimality is replaced by Pareto-optimality[2] and one typically obtains a large number of Pareto-optimal solutions rather than a single optimum candidate.

Redford et al. (1998; 2001) proposes a model for emergence of syllable systems, where several functional pressures related to the articulatory and perceptual processes acting over the words and vocabulary as a whole are formulated as objective functions and/or constraints. A single optimization function is then formed through the weighted linear combination of the different objectives. The optimization process is carried out using GA. The model shows the emergence of universal syllable typologies.

3.2.4 Criticisms

The analytical models have the following two important advantages over the synthetic models:

- It is often easy to identify the causal relationship between the model parameters and the observed properties of the linguistic systems. This is due to the fact that once the exact analytical solution for the models are known (esp. for the dynamical system and statistical mechanics models), the dependence of the solution (observed pattern) on each of the variables are also known.
- These techniques usually yield models that are computationally tractable and can be employed to develop NLP applications. For example, the closed form solution obtained from a dynamical system model can be readily implemented as a program that outputs the pattern for a given set of initial conditions.

Nevertheless, there are also several disadvantages associated with the analytical techniques. These are listed below.

[2]See Sec. 4.1.1 for definition of Pareto-optimality and other notions related to Multi-objective optimization

- In most of the cases, the dynamical system models (e.g. Abrams and Strogatz 2003; Niyogi and Berwick 1996; Niyogi and Berwick 1998) as well as the statistical mechanical models are analytically solvable only in the thermodynamic limit (i.e., when the number of entities/speakers in the system tends to infinity). Moreover, these models also assume that the generations are non-overlapping in nature. Clearly, these assumptions do not hold good in reality. Several lines of research show that the predictions of analytical models (also known as the macroscopic models[3]) under the above assumptions do not match those of synthetic or microscopic models (see Stauffer and Schulze (2005) for a comparison of the macroscopic and microscopic models in the context of language competition; also see Briscoe (2000c; 2000a)). This is a serious drawback, as this calls into question the validity of the predictions made by the analytical models.

- Although the optimization models do not suffer from the aforementioned limitation, the problem with them is that these models are silent about how the optimization process might have taken place in reality. There are at least three reasons for which the issue of optimization deserves a non-trivial explanation (Oudeyer 2005b): 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.

[3]The distinction between macroscopic and microscopic descriptions of a system is prevalent in disciplines such as Physics, esp. Thermodynamics, Economics, Population Biology and Linguistics. Macroscopic models abstract out the details of a system and deal with the average behavior, whereas the microscopic models take into account the details of the individual entities of a system. In the context of language, a microscopic model considers a language as a collection of utterances by a population of speakers, whereas a macroscopic model abstracts out those details and views a language as an atomic system, that is, the average of the language of the individuals.

In the next section, we discuss the synthetic techniques based on the microscopic view of language, which circumvents some of the aforementioned limitations of the macroscopic models.

3.3 Synthetic Techniques

As discussed previously, Multi-Agent Simulation (MAS) is the most popular synthetic technique in modeling of language evolution and change. The earliest examples of MAS modeling for language change goes back to the 1960s, when Klein and his colleagues developed a general framework for Monte Carlo simulation of language change (Klein 1966; Klein et al. 1969; Klein 1974) and demonstrated it on Tikopia and Maori languages. 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.

A MAS model has three basic components: agent representation, agent interaction, and the world in which the agents are situated (Turner 2002). The existing surveys on MAS models of language evolution classify the models according to one or more of these three basic dimensions. Wagner et al (2003), for example, suggests a classification of the simulation models based on the features related to the agent's world – *situatedness* and the agent's linguistic model – *structuredness*. A model is said to be *situated* if the agents interact with the "artificial world", in which they are situated, in non-communicative ways. A simulation is *structured* if the "utterances are composed of smaller units, such as the words forming a phrase." Thus, according to this classification system, there are four basic types of simulations:

- Situated and structured (Steels 1998; Cangelosi and Parisi 2001; Gong et al. 2004; Gong and Wang 2005)
- Situated and unstructured (Steels 1995; Oudeyer 1999; Smith 2005)

- Nonsituated and structured (Hare and Elman 1995; Kirby 1999; Kirby 2002; Oudeyer 2005a; Oudeyer 2005b)
- Nonsituated and unstructured (de Boer 1999b; de Boer 2000a; de Boer 2000b; Fyfe and Livingstone 1997; Livingstone 2003)

The linguistic issues addressed by the simulations, however, have been expressed in terms of another classification system that was proposed by Hockett and Altmann (1968) as a suggestion to the problem of evolution of communication and complexities of communication systems of the biological world.

Steels (1997b), on the other hand, places the simulation models in a two dimensional framework reflecting the computational issues and the problem at hand. The problems addressed are arranged on the conventional organizational hierarchy of languages, such as phonology, morphology and syntax.

In the following three subsections we briefly discuss the different modeling paradigms within MAS in terms of the structure of the agents, their interaction patterns and the simulation world.

3.3.1 The Agent Model

A *linguistic agent* has three basic components. The articulation and perception mechanisms help the agent to utter and understand linguistic messages. The mental model or the grammar is the abstract description of the language in an agent's brain. Learning is the process of updation of the mental model or the grammar based on the inputs from the environment.

Articulator and Perceptor

The details of the articulator and perceptor models in any MAS depend on the complexity of the language being modeled. For example, in de Boer (1999b; 2000a; 2000b) the scope of modeling is restricted to vowel systems. Therefore, agents are designed to articulate vowels signals through formant-synthesis and perceive the same through formant-analysis. Similarly, Oudeyer (2005b) describes an extremely detailed articulator and perceptor models based on the formant-synthesis and analysis approach (de Boer 2000b). In this model, the movement of the vocal tract is simulated to generate the formant frequencies.

Such detailed articulatory or perceptual processes, though very accurate and realistic, are time-consuming, and therefore, are not suitable for models that address questions related to morphology and syntax. Consequently, in (Kirby 2001; Kirby 2002; Briscoe 2000b) that attempts to model the structure and dynamics of syntactic constructs, the articulators and perceptors are kept exceedingly simple that can generate symbolic strings of words or phones.

Mental Model

Mental model or the representation of the grammar of the language (i.e., the I-language) within an agent is a very crucial issue in MAS because this alone determines the set of possible linguistic structures that can eventually emerge in the system. If the mental model is so defined that it is tuned towards (i.e., has a strong bias towards) the final results obtained through simulation, then there is every reason to doubt its plausibility as well as the validity of the results. On the other hand, extremely general mental models are computationally expensive and usually intractable unless the problem being investigated is very simple.

The agent representation schemes can be broadly classified into the *symbolic* and the *connectionist* paradigms. The symbolic models use *rules* for representing and processing of the language (Briscoe 2000b; de Boer 2000a; Kirby 2001; Kirby 2002), whereas connectionist models are typically implemented using Artificial Neural Networks (Hare and Elman 1995; Batali 1998; Cangelosi and Parisi 1998; Smith 2002; Cangelosi 2003; Cangelosi 2005; Oudeyer 2005b). See Turner (2002) for a detailed survey on use of neural nets and rules in MAS. While connectionist models provide a strong framework for representation of agents and do not require any unnecessary assumptions to be made on the part of the modeller about the internal representation of the agents, they are not transparent in the sense that one cannot precisely point out why and how a system (read neural network) works the way it does. Thus, it is more difficult to extrapolate from the model to the real world.

Learning

Language acquisition is one of the key factors affecting language change (see Sec. 1.1.1 for a discussion). Consequently, the proper modeling of the learning mechanism is of utmost importance to any MAS based model of language change. Formally, the process of learning is an updation of the mental (or language) model of the agent or equivalently, transition from one mental state to another. Some of the different learning schemes that have been described in the literature are (Turner 2002)

- Rule generalization (Kirby 2001): Given a set of non-compositional rules, they are replaced by an equally expressive, but more compositional rules.
- Obverter (Oliphant and Batali 1997): The rules for signal generation are chosen so as to maximize the probability of correct decoding and similarly, the rules for signal decoding are adopted such that the probability of correct generation is maximized. Correct generation and decoding are defined as the average generation and decoding rules of the population at a particular point of time, and thus the obverter procedure requires that an agent has access to the mental states of all the other agents. Batali (1998) describes a self-understanding based learning scheme, which is similar to the obverter procedure but differs from the former in the fact that during rule adoption for signal generation, the agents try to maximize the probability based on their own decoding rules. Thus, in the self-understanding method agents need not access the internal states of other agents for learning the generation rules, even though they need to do so for learning the decoding rules.
- Imitation (de Boer 2000a; Livingstone 2002; de Boer 2000b): Agents update the rules so that they can imitate the other agents as closely as possible. In contrast to the above two learning schemes, imitation does not presume the knowledge of the internal states of other agents by an agent, and thus, is more realistic.

The details of the implementation of the learning scheme in a MAS depends on the representation of the mental model. In neural networks, learning is carried out using the standard techniques such as back propagation (Rumelhart et al. 1986) and genetic algorithms (Montana and Davies 1989). In rule-based agents, learning is implemented through statistical techniques such as *expectation maximization*, Bayesian learning, example based learning etc. See Mitchell (1997) for an accessible introduction to these machine learning techniques.

It might be worthwhile to mention here that Niyogi (2006) compares three basic learning schemes – *memoryless learner* (trigger learning algorithm), *batch learner* and *asymmetric cue based learner*, and other variations of the same in the context of dynamical system based models of language change. Interestingly, different theoretical predictions are made for the different learning algorithms. For instance, in the case of language competition between two languages (Niyogi 2006: Ch 5), the memoryless learner gave rise to a single stable attractor (i.e., only one of the languages dominate), the batch learner gave rise to two stable attractors (i.e., both the language can co-exist) and depending on a model parameter p the cue based learner could give rise to both a single stable attractor, or two stable attractors.

3.3.2 Agent Interaction

In MAS the agents form a linguistic community and they interact with other agents linguistically and sometimes also extra-linguistically or through the shared environment. The interaction pattern between the agents has crucial impact on the nature of the emergent linguistic system. Usually agent-interactions are modeled using *language games*.

The concept of language game was introduced by Steels (1995; 1998). Any linguistic communication among a group of individuals is viewed as a series of language games. The basic structure of a language game is as follows. An *initiator* (speaker) initiates a language game by identifying a concept to be communicated; the initiator then generates a linguistic signal corresponding to the concept; the signal is received by another agent – the *receiver*, who then tries to identify the concept from the linguistic signal perceived; finally, the receiver conveys the information regarding the identified concept through the generation of a linguistic or extra-linguistic signal, or through some action on the shared environment. In some cases an extra step is added, where the initiator, based on its interpretation of receiver's linguistic/extra-linguistic signal or action, informs the receiver whether the communication was successful or not (e.g., Steels 1995; de Boer 2001). Nevertheless, recent works (Oudeyer 2005b; Smith 2005) have shown that high communication success rate and convergence to a set of shared linguistic conventions can be arrived at in a multi-agent system even without the assumption of extra-linguistic communication.

There are numerous MAS models based on language games. The exact nature of the game depends on the objective of modeling. Naming games (Steels 1995), spatial language games (Bodik 2003a), imitation games[4] (de Boer 2000a) and advertising games (Avesani and Agostini 2003) are examples of commonly used language games. We do not elaborate any further on the structure of the language games; rather we discuss at some depth the different strategies for choosing the initiators and receivers in a language game and their physical significance.

Recall that for explaining the causal forces behind language change one must explain the interaction patterns of the speakers and the language acquisition process (Fig. 1.2). Different strategies for selecting the players of a language game can be correlated to the different arrows shown in Fig. 1.2. Based on this, we propose the following classification of the MAS models.

- *Vertical model* or *Iterated learning model*: Normally, there is one teacher and one learner. The information flows vertically as the learner acquires the teachers language. Thus, only the thick black vertical arrows are modeled (Hare and Elman 1995; Kirby 2001). These models try to explore the effect of language acquisition on language change.

- *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 and gray horizontal arrows are modeled (de Boer 2000a). These models try to

[4]In Chapter 5 and 6 of this thesis we use imitation games to develop MAS models for language change. Sec. 5.1.1 describes the concept of imitation game in details.

explore the self-organization of the language through cultural transmission. The effect of learning is not ignored, but more importance is laid on synchronic variation and social communication patterns (Gong et al. 2004).

- *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 2003b; Smith and Hurford 2003). Clearly, the hybrid model subsumes the vertical and horizontal models as far as the explanatory power is considered. However, owing to their simplicity and computational tractability, the later models are more popular than the former one.

3.3.3 World Models

In the context of language evolution the environment or world model, where the agents are grounded, is usually quite complex because it is important to model the benefits of a verbal communication system, if any, through the interactions with the world, such as the presence of predators (Jim and Giles 2000) or poisonous foods (Cangelosi and Parisi 1998). However, in the context of language change, where there already exists a language, the role of the environment is limited to social issues such as the structure of the society, the social status associated with a language and migration of population. Some of these issues are discussed below.

Society Structure: The interaction pattern between the agents plays an important role in the nature of linguistic variation observed in an artificial system. This has been discussed in Sec. 3.2.2.

Aging: Many MAS models consider the aging of the agents because the adults usually stick to the already learnt conventions, though the children during the learning phase acquire languages quite fast. It has also been observed that languages usually change when one generation of speakers are acquiring the language. See e.g., Hurford (1991a); Kirby and Hurford (1997) and de Boer (2005).

Migration and Language Contact: Migration is modeled in MAS through an influx or outflux of agents. Migration of agents leads to language contact and language change. Several MAS models (see e.g., de Boer (1999a; 2000a; 2000b)) implement small scale migration through random introduction or removal of agents. Steels (1997a) studies the effect of language learning and language contact on language dynamics.

We do not know of any MAS model of language change, where the social status associated with a language is explicitly modeled. However, as discussed earlier there are some analytical models (see e.g., Abrams and Strogatz (2003)) that take into account the effect of social status associated with a language.

3.4 Research Issues

The previous two sections primarily focus on the modeling techniques and the linguistic problems being modeled have been excluded from the discussion. Furthermore, due to the similar nature of the techniques used in modeling language evolution and change, a large number of examples included in those sections have been borrowed from the area of language evolution. In this section, we concentrate on the different domains and problems in diachronic linguistics that have been addressed using computational techniques. We shall briefly mention the field of investigation and cite relevant examples within the board areas of linguistics.

3.4.1 Phonetics and Phonology

Structure of sound inventories: Explanations of the emergence of vowel systems through optimization (Liljencrants and Lindlom 1972; Lindblom 1986; Schwartz et al. 1997; Lindblom 1998), neural networks (Joanisse and Seidenberg 1997), MAS (de Boer 1997; de Boer 1999a; de Boer 2000a; de Boer 2000b; de Boer 2001) and MOGA (Ke et al. 2003).

Complex utterances and syllables: Explaining universals and other phonological constraints through MAS (Oudeyer 2005a; Oudeyer 2005b); explaining the emergence of syllables typologies through GA-based constrained optimization (Redford et al. 1998; Redford 1999; Redford and Diehl 1999; Redford et al. 2001); interactions between syllables as complex networks (Soares et al. 2005).

Vowel harmony: Modeling the emergence of vowel harmony in the Turkic languages through MAS (Harrison et al. 2002; Dras et al. 2003).

3.4.2 Lexicon

There has been several works on emergence of a shared lexicon in a group of communicating agents (see e.g., Steels (1995; 1996); Oudeyer (1999)) and universal properties of the lexicon and word usage (see e.g., Dorogovtsev and Mendes (2001); Sigman and Cecchi (2002); Ferrer-i-Cancho (2005a); Tamariz (2005); Kosmidis et al. (2006b)). However, we shall discuss here only the issues related to lexicon change.

Spontaneous lexicon change: Models explaining the spontaneous change of a lexicon through MAS (Steels and Kaplan 1998; Dircks and Stoness 1999; Bodik and Takac 2003).

Homophony and borrowing: See Wang et al. (2005) for a survey on models of homophony and borrowing. Also see (Steels and Kaplan 1998; Ke and Coupé 2002).

Lexical diffusion: See Wang et al. (2005) for a survey on models of lexical diffusion.

3.4.3 Morphology and Irregularity

Hare and Elman (1995) describes a neural network based modeling of the historical change of English verb morphology. The input to the system is the highly complex past tense forms of Old English. The neural nets are trained to learn the past tense forms from this data. The learnt structure of the neural net is then set as the target for the next generation. Thus, it is similar to an Iterated learning model, where each generation is represented as a neural network. After several generations of training, the past tense forms get regularized and the modern forms emerge in the system. Certain verbs, nevertheless, remain irregular (i.e., retain their older forms), which is observed in modern English as well. This immunity towards regularization is explained through usage frequency of the verbs.

Kirby (1999; 2001; 2002) describe an Iterative learning based MAS model for emergence of regular and irregular patterns in morphology. The findings are similar to that of (Hare and Elman 1995). However, unlike (Hare and Elman 1995) the experiments are conducted with artificial string languages, rather than on real linguistic data.

3.4.4 Syntax

Diachronic syntax is a well studied and challenging research area in historical linguistics (see Kroch (2001) for a review). Several models have been put forward to explain the emergence of recursive syntax in a population of agents that initially had a language without recursive syntax. See for example, Steels (1998)and Kirby (2002) for MAS based accounts; Christiansen and Devlin (1997) for neural network based explanation; Ferrer-i-Cancho and Sole (2004) for complex network based approach; Niyogi (2006) for dynamical system model. Here we discuss a few models that have direct relevance to diachronic syntax.

Learning and syntactic change: Briscoe (2000b; 2002) discuss the effect of learning on diachronic syntax both in microscopic (i.e., MAS) and macroscopic (i.e., dynamical system) models. Effect of learning on language change (specifically syntactic change) is also discussed in Niyogi (2006), where the analysis is purely from the macroscopic perspective.

Word order change: Itoh and Ueda (2004) describes an Ising model for word order change, which has been described in Sec. 3.2.2. Minett et al. (2006)

proposes a MAS model for predicting word order bias. Also see Christiansen and Devlin (1997).

Loss of V2 (verb second property): Niyogi and Berwick (1997c; 1997; 1998) put forward dynamical system models of language change to explain the loss of V2 in languages such as Portuguese.

3.4.5 Other Areas

We have already discussed the models that explore the effect of interaction pattern between the individuals on language variation (synchronic and diachronic). Another common research problem in this field is to explain the emergence of S-shaped dynamics during language change (Briscoe 2000a; Harrison et al. 2002).

3.5 Concluding Remarks

In this chapter we have seen that computational and formal modeling of language evolution and change is a burgeoning research area. A variety of modeling techniques are being used and the results obtained from them are convergent as well as complementary. As is evident from the publication trend shown in Fig. 3.1, the field, however, is still in its initial phase; it is expected to flourish as a fruitful research program in the coming years with a larger number of contributions from the various disciplines involved.

At the same time, one cannot deny the fact that almost all the models pertaining to the field are weak on rigorous empirical validation. In the context of the models of language evolution, this limitation is understandable because we do not have any data from the past that show the different stages of evolution of human language. However, despite several methodological problems (see Sec. 1.1.4 for a discussion), we do have some amount of historical data to study and validate the models of language change.

Rather than simulating the change of real languages, a majority of the models of language change, e.g., (Steels and Kaplan 1998; Kirby 2001; Redford et al. 2001; Briscoe 2002), study the general properties of the model and compare them against the empirically observed linguistic universals. On the other hand, several works that claim to have modeled real languages typically validate the models against the statistical distribution of the linguistic forms or speakers, rather than the structure of the linguistic forms. For example, the dynamical system models described in Niyogi and Berwick (1997c; 1997a; 1998) that intend to model the change affecting European Portuguese, end up measuring the number of two competing linguistic varieties in a corpus of historical Portuguese, and compare this statistic with that predicted by the models. Similarly, the work on Turkic vowel harmony (Harrison et al. 2002; Dras et al. 2003) try to match the S-shaped dynamics observed in the real world, and not the phonological structure of the words obtained after application of harmony.

In fact, there are very few works that compare the real world linguistic forms with the emergent ones. The works on vowel systems (see Sec. 3.4.1 for references) and change in English verb morphology (Hare and Elman 1995) are few such examples. Despite the fact that a large number of cases of language change have been documented by the historical linguists, this extreme paucity of models of real world language change is presumably an outcome of the hardness of modeling vis-à-vis the lack of practical applications of such models. Else how can one explain the existence of exceedingly complex and extremely detailed "synchronic models" of natural languges that are used for the purpose of NLP applications? See Jurafsky and Martin (2000) for an accessible introduction to the synchronic language models used in NLP.

Given the fact that computational models of diachronic linguistics have been criticized for their aforementioned drawback (Hauser et al. 2002; Poibeau 2006), development of models of real language change seems to be an extremely important as well as challenging research direction that can provide further insights into the phenomena being investigated, giving more reasons to believe not only the plausibility of that particular model, but also the computational models as such. This is precisely the aim of the current work, which we explore in the subsequent chapters of this thesis.

Chapter 4

Constrained Optimization Model for Schwa Deletion in Hindi

Functional explanations of phonological change argue that the sound patterns of a language evolve through a constant process of optimization under the pressure of functional forces acting over a language (Boersma (1998) and Sec. 2.2.1 of this thesis). *Schwa deletion* in Hindi (SDH) can be explained in terms of syllable minimization within the framework of functional phonology. The objective of this chapter is to construct a functional model for explaining the schwa deletion pattern observed in Hindi. For this purpose, we propose a constrained-optimization framework, where the different functional forces are formally encoded as constraints and/or optimizing criteria, and the model is solved analytically to obtain the schwa deletion pattern of Hindi. This in turn helps us formulate a linear time algorithm for predicting the deletion pattern, which has been used to develop a grapheme-to-phoneme converter (G2P) for Hindi.

This chapter is organized as follows. In Sec. 4.1, we propose a general mathematical framework for constrained-optimization in functional phonology. Some basic definitions and notations specific to SDH are also explained in this section. Sec. 4.2 describes a way to specify the syllable structure of Hindi as well as the syllabification procedure in this framework. Sec. 4.3 discusses a formal encoding of the acoustic distinctiveness constraints in the context of SDH. In Sec. 4.4 we redefine SDH as a syllable minimization problem and analytically show that the optimal SDH pattern according to the proposed model is almost identical to Ohala's rule (Ohala 1983b). Sec. 4.5 presents an efficient algorithm for generating the SDH pattern based on the syllable-minimization technique. The effect of morphology on the deletion pattern is described subsequently, for which some modifications to the algorithm are suggested. The evaluation of the algorithms and their applicability to Hindi text-to-speech system are also reported in Sec. 4.5. Sec. 4.6 summarizes the salient features of the constrained-optimization model, its advantages and limitations with special reference to SDH.

4.1 The Framework

In this section we propose a general mathematical framework for modeling functional explanations. We also introduce certain basic concepts and notations in the context of schwa deletion. These will be used to develop a functional model for SDH in the subsequent sections.

4.1.1 A General Framework for Constrained Optimization

Let $\mathcal{L}$ be the universal set of natural languages consisting of all the languages of the past, present and the future (possibilities). Let $\mathcal{F} = \{f_1, f_2, \ldots, f_n\}$ be the universal set of objective functions, each measuring some functional benefit of a language $l \in \mathcal{L}$ under consideration. Note that as discussed in Sec. 2.2.1, the proponents of functional phonology (Boersma 1998) as well as optimality theory (Prince and Smolensky 1993) claim that the structure of all natural languages are shaped by the interaction of a universal set of functional objectives and/or constraints. Therefore, in this framework, we assume $\mathcal{F}$ to be universal (i.e. language-independent) and finite. Also, without loss of generality, we define each of the objective functions $f : \mathcal{L} \rightarrow \mathbb{R}$, to be maximizing in nature.

Definition 4.1 *For two languages $l_i, l_j \in \mathcal{L}$, and an objective function $f \in \mathcal{F}$, l_i is said to be better than l_j with respect to f if and only if $f(l_i) > f(l_j)$.*

Definition 4.2 *For two languages $l_i, l_j \in \mathcal{L}$, l_i is said to be better than l_j or l_i dominates l_j with respect to $\mathcal{F}' \subseteq \mathcal{F}$, if and only if*

1. *l_i is better than l_j with respect to at least one of the objective functions $f \in \mathcal{F}'$, and*
2. *there does not exist any objective function $f' \in \mathcal{F}'$, such that l_j is better than l_i with respect to f'.*

The inherently conflicting nature of the objective functions eliminates the possibility of any globally optimum language $l^* \in \mathcal{L}$ such that l^* dominates all the languages in $\mathcal{L}$ (see Sec. 2.2.1 for discussion on the same). Nevertheless, we can have languages which are dominated by no other language.

Definition 4.3 *A language $l_i \in \mathcal{L}$ is said to be Pareto-optimal or non-dominated if and only if there does not exist a language $l_{j \neq i} \in \mathcal{L}$, such that l_j dominates l_i.*

The concepts of domination and Pareto-optimality are illustrated in Fig. 4.1 with the help of two objective functions f_1 and f_2. Note that in general, it is not possible to achieve arbitrary high values for one of the functions, say f_2, by fixing the other function f_1 at a constant k_1. For example, if we restrict the maximum length of the words of a lexicon to some value, say 5, then the maximum acoustic distinctiveness, measured as the edit distance between the two words, can never exceed 5. Let $\phi_{12}(x)$ be the maximum value of f_2, when f_1 is held at x. As illustrated in Fig. 4.1, ϕ_{12} is expected to be a decreasing function of x and it bounds the set of possible languages over the $f_1 - f_2$ plane. It also defines the set of Pareto-optimal languages. Moreover, according to the functional principles (Boersma 1997b), all the languages observed in nature must belong to this set; if not, then the language will undergo a series of change and will finally converge to a language on the ϕ_{12} curve. In other words, all languages on the left of the ϕ_{12} curve are unstable, and languages on the ϕ_{12}

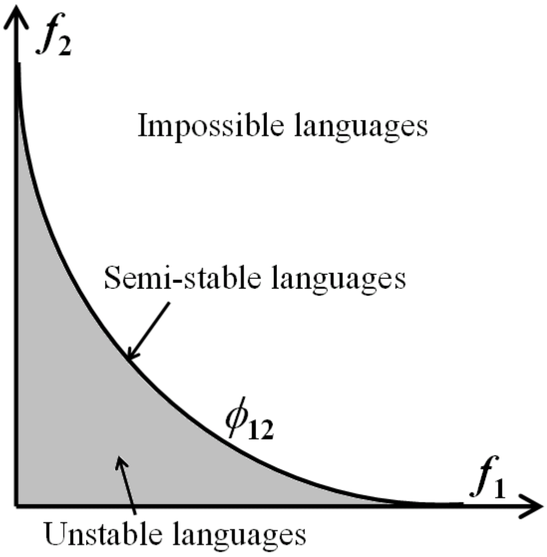


Figure 4.1: The concept of Pareto-optimality illustrated through two objective functions f_1 and f_2. The ϕ_{12} curve represents the Pareto-optimal languages, which are semi-stable. The points (languages) in the shaded region represents the *dominated set*, i.e. the languages that are unstable, whereas the points above the ϕ_{12} curve represents impossible languages.

curve are semi-stable. The curve acts as a *stable attractor* (Ott 2002) for the state space $\mathfrak{L}$, where the dynamics are governed by language change.

One possible way of defining language change in this framework is as follows. Let there be a language l, whose functional fitness across the n dimensions are given by $\{f_i(l)\}$, for $1 \leq i \leq n$. Suppose that out of these n objectives, the value of the first $m < n$ are fixed at constant values $f_i(l)$, for $1 \leq i \leq m$, and the rest are allowed to change. The process of *language change* under these circumstances will try to search for a language l^* such that

- $f_i(l^*) = f_i(l) \quad \text{when } 1 \leq i \leq m$
- l^* is non-dominated in $\mathfrak{L}$ with respect to the objective functions f_{m+1} to f_n

Note that depending on the situation, the process of change might define l^* uniquely. Alternatively, the restrictions might define a subset of $\mathcal{L}$, of which l^* is a member.

The process of language change as defined above is similar to a *constrained optimization problem* (Bertsekas 1996; Chong and Żak 2001), where the first m objective functions act as constraints and the rest $n - m$ as optimization criteria. The scope and complexity of the process of change is determined by the mathematical nature of the constraints and objective functions (e.g. linear vs. non-linear, discrete vs. continuous, etc.). Functional Optimality Theory (Boersma 1997a), for instance, proposes a hierarchical organization of constraints (through ranking) as the underlying interaction scheme between the objectives. Here we refrain from making any comments on the functional forms as it is a hard and debatable issue; rather, in the rest of this chapter we try to develop a very specific mathematical model for SDH under this general constrained optimization framework.

4.1.2 Notations and Definitions

In order to model SDH as a constrained optimization problem, we consider the orthographic representation as the underlying form, and the corresponding phonemic representation (i.e. the pronunciation of the word) as its surface form.

Let Σ_G (Σ_P) be the finite set of graphemes (phonemes) for a language l. Each of these sets can be partitioned into two subsets of vowels V_G (V_P) and consonants C_G (C_P), such that

$$\Sigma_G = V_G \cup C_G, \qquad \Sigma_P = V_P \cup C_P$$

There is a special symbol $a_G \in V_G$, which is the schwa. V_P also has a symbol a_P, which represents the default pronunciation of the grapheme a_G in l. An *orthographic word* w_G is a string over Σ_G, i.e. $w_G \in \Sigma_G^*$. It should be noted that not all strings in Σ_G^* are valid orthographic words of l. Similarly, a *phonetic word* w_P is a string over Σ_P, i.e. $w_P \in \Sigma_P^*$. Let Λ be the set of all valid orthographic words of l. We define a function $F_{G2P} : \Lambda \rightarrow \Sigma_P^*$ that maps a valid orthographic word in l to a phonetic string, such that $F_{G2P}(w_G)$ is the pronunciation of w_G. The subscript G2P stands for *grapheme-to-phoneme* conversion. It should be mentioned here that in most of the languages, one does encounter situations where the same orthographic form can have more than one valid pronunciation. Such words are called *homographs*, meaning different words with the same orthographic representation. However, such instances are rare in any language and almost absent in Hindi. This leads us to define F_{G2P} as a single-valued function, which in turn helps us avoid several unnecessary complications in the notations as well as analysis of the model.

To further simplify our analysis, we conceive F_{G2P} as a composition of two functions – f_{G2P} and f_{DS}, such that $f_{G2P} : \Lambda \rightarrow \Sigma_P^*$ maps an orthographic word to its proper pronunciation except for the case of schwas, which are kept unchanged. The function $f_{DS} : \Sigma_P^* \rightarrow \Sigma_P^*$ takes a phonetic string as input and maps it to another phonetic string by appropriately deleting the schwas. The subscript *DS* in f_{DS} stands for *Delete Schwa*. Mathematically, this fact can be stated as follows.

$$F_{G2P}(w_G) = f_{DS}(f_{G2P}(w_G)) \tag{4.1}$$

The assumption behind this decomposition is that all other phonological processes are independent of schwa deletion. Although this assumption may not be true for a language in general, it holds good for Hindi because of its phonemic orthography. In fact, f_{G2P} for Hindi involves mapping of each grapheme to a corresponding phoneme irrespective of the context[1], but f_{DS} happens to be a non-trivial function (Narasimhan et al. 2004; Kishore et al. 2002; Ohala 1983b). In the rest of this chapter, we focus only on the function f_{DS}.

4.2 Syllable and Syllabification

Our objective here is to show that the schwa deletion pattern of Hindi is an outcome of syllable minimization. Consequently, in this section, we define the concepts of syllable and syllabification following the framework and notations

[1]There are a few cases, however, where the mapping f_{g2p} requires context information. For instance the nasal marker M is replaced by an appropriate homorganic nasal: *chaMdA* /*chandA*/ (moon), *chaMpA* /*champA*/ (a flower), *chaMgA* /*cha∼NgA*/ (fit and fine)

described above, on which we build up the model for SDH in the subsequent sections.

In linguistics, a *syllable* is defined as the unit of processing and representation in the recognition of spoken words (Kahn 1976; Selkirk 1982), and is widely accepted as a psycholinguistically meaningful and morpho-phonologically influential concept (Hooper 1972; Marslen-Wilson et al. 1994). The internal structure of a syllable is composed of a *nucleus*, which is optionally preceded by an *onset* and followed by a *coda*. The process of segmenting a string of phonemes into syllables is known as *syllabification*. Discussions on syllable structure and principles of syllabification can be found in (Goslin and Frauenfelder 2000) and the references therein.

Based on the notations introduced in Sec. 4.1.2, we define a *syllable* σ as a string $\alpha\vartheta\beta$, where $\alpha \in C_P^*$, $\vartheta \in V_P$ and $\beta \in C_P^*$. Both α and β can be null strings. Here, α, β and ϑ represent the onset, coda and nucleus respectively. We also define the projection functions *onset* and *coda*, such that the function *onset* returns the onset and the function *coda* returns the coda of a syllable. In other words,

$$onset(\sigma) = onset(\alpha\vartheta\beta) = \alpha$$
$$coda(\sigma) = coda(\alpha\vartheta\beta) = \beta$$

4.2.1 Phonotactic Constraints

One of the most widely accepted principles of syllabification is that of the *Legality principle* (Pulgram 1970; Hooper 1972; Kahn 1976; Vennemann 1988), which states that the syllable onsets and codas are restricted to only a small subset of strings in C_P^*. The constraints on allowable consonant clusters (i.e. strings over C_P^*) are called *morpheme-sequential constraints* (Ohala 1983b) or *phonotactic constraints* (PC). According to the Ohala's rule, SDH should not violate the PCs of a language. There is some debate on whether PCs are to be defined with respect to a word (Ohala 1983b; Goldsmith 1990) or a syllable (Kahn 1976; Vennemann 1988). For the purpose of this work, we shall model the PCs at the level of syllables. However, as it will be evident soon, modeling of word-level PCs are similar to that of syllable level PCs.

Several techniques have been suggested in the literature for the representation of PCs in a language (Carson-Berndsen 1993; Belz 1998). As defined below, here we conceptualize PCs as constraint functions; nevertheless, these constraint functions can be implemented for computational purposes by following any of the previously proposed schemes for representation of the PCs.

Definition 4.4 *The function $C_{onset} : C_P^* \rightarrow \{0, 1\}$ maps a consonant cluster to 1 if and only if it is an allowable onset in the language l.*

Definition 4.5 *The function $C_{coda} : C_P^* \rightarrow \{0, 1\}$ maps a consonant cluster to 1 if and only if it is an allowable coda in the language l.*

An alternative approach to syllable structure is known as the *Sonority Cycle* (Clements 1990), according to which the phonemes or segments can be placed on a universal *sonority scale* (de Saussure 1916; Vennemann 1988); the sonority is maximum for the nucleus, and decreases as one moves away from it. In this theory, PCs in all the languages reflect this universal principle of sonority cycle, and consequently, the PCs as well as the syllabification procedure can be modeled independent of the language. Although it is tempting to adopt such an elegant generalization for syllabification, there also exist several counter-examples to this rule and the absolute validity of this principle is debatable (see Goslin and Frauenfelder (2000) for a discussion in the context of French). However, we note an interesting corollary of the above sonority cycle principle:

> For any language l, if $s \in C_P^*$ is an allowable onset/coda, then so is every substring of s.

We conjecture that even if the basic principle of sonority cycle is violated in a language, the above corollary is always true. We do not know of any previous work stating or investigating this result, but through empirical analysis of the Hindi PCs as enlisted in (Ohala 1983b), we have found this corollary to be valid for Hindi.

For the current purpose we maintain a list of allowable consonant clusters in Hindi, which is used during the process of syllabification as described below.

4.2.2 Syllabification

Syllabification is a language specific issue and often the syllable boundaries are not clearly evident from the acoustic signal. These issues have even lead some of the researchers in the past to cast a doubt on the existence and usefulness of the concept of syllables (Lebrun 1966; Kohler 1966). The specific case of Hindi is not an exception either. Several theories have been proposed to describe the syllable structure of Hindi, but there is hardly any consensus among the researchers. However, it is interesting to note that, unlike the case of most of the world's languages, Hindi does not seem to follow the *Onset maximization* principle[2] (Pulgram 1970; Hooper 1972; Selkirk 1982).

In fact, some of the studies claim that Hindi syllabification follows the law of *Coda maximization*[3]. As we shall see shortly, it is possible to explain the schwa deletion pattern of Hindi through syllable minimization, if one adopts a syllabification based on coda maximization and not onset maximization principle. However, syllabifications that tend to strictly maximize the length of the coda often lead to very counter-intuitive results. For example, according to this rule, $jantu$ (animal) is syllabified as $/jant - u/$, which is unacceptable. The acceptable syllabification in this case is $/jan - tu/$. Here we propose a syllabification principle for Hindi, which neither maximizes the onset nor the coda; rather it tries to strike a balance between the two principles. Since, our primary objective here is to present a plausible account of SDH within the constrained optimization framework, we make no attempt to empirically verify this principle as a part of this work. Nevertheless, it is worth mentioning that a text-to-speech system developed based on the proposed syllabification principle produces intelligible speech outut. Furthermore, we also show that apart from the proposed syllabification, certain other classes of syllabification rules for Hindi (which do not include the onset maximization principle) are capable of explaining the Ohala's rule.

We begin our discussion on syllabification with a very general definition of

[2]The onset maximization principle states that the syllable boundaries are so chosen in a word that every resulting syllable has the maximum length onset as allowable by the PCs

[3]Through personal communication with Prof. Thakur Das, Department of Linguistics, University of Agra, India

the term, and subsequently introduce several constraints and rules specific to Hindi. A syllable is *valid* in a particular language l if it satisfies the PCs of the languages. A *valid syllabification* of a phonetic word w_P is one in which all the syllables are valid and each phoneme in the word belongs to one and only one syllable[4]. These facts can be formally stated as follows.

Definition 4.6 *A syllable σ is valid in a particular language if and only if the following conditions hold.*

$$\begin{aligned} C_{onset}(onset(\sigma)) &= 1 \\ C_{coda}(coda(\sigma)) &= 1 \end{aligned}$$

Definition 4.7 *Let $w_P \in \Sigma_P^*$ be a phonetic word. A valid syllabification of w_P is a string of syllables $\sigma_1\sigma_2\cdots\sigma_m$, where*

- $\forall i, 1 \leq i \leq m, \sigma_i$ *is a valid syllable.*
- *Concatenation of the syllables in that order (from 1 to m) gives the string w_P.*

These concepts are illustrated through the following example.

Example 4.1 Let w_P be $/samprati/$ and the language under consideration be Hindi. The syllabification $\sigma_1 = /sa/$, $\sigma_2 = /mpra/$ and $\sigma_3 = /ti/$ is not valid because, $C_{onset}(onset(\sigma_2)) = C_{onset}(/mpr/) = 0$, i.e. $/mpr/$ is not an allowable valid onset in Hindi. Similarly, the syllabification $\sigma_1 = /samp/$ and $\sigma_2 = /rati/$ is invalid because $/rati/$ is not a valid syllable as it contains two vowels. But the syllabifications $/sam - pra - ti/$, $/samp - ra - ti/$, $/samp - rat - i/$ and $/sam - prat - i/$ are valid, where '$-$' denotes syllable break. □□

A valid syllabification is one that is pronounceable, however, not all valid syllabifications of a word are pronounced by the native speakers. In other words, there is always a preferred syllabification for every word. We shall refer to this

[4]Note that this may not be strictly true in real life, because there are cases where a segment may belong to more than one syllable; such segments are known as *ambisyllabic* (Kahn 1976).

preferred syllabification as the *optimal syllabification* of a word. According to the functional view, the optimal syllabification is the one which maximizes the perceptibility and ease of articulation of a word. Thus, our aim here is to define a function that measures the "goodness" of a syllabification and the optimal syllabification is the one, for which the value of this goodness function is maximum. Alternatively, we can define a hardness function for a syllabification, the value of which is minimum for the optimal syllabification.

Intuitively, this hardness function should measure the cumulative hardness of the syllables in a syllabification, where the hardness of a syllable depends on the structure of the syllable. Thus, in effect, we want to associate a hardness scale with the syllable structure. The empirical analysis of Hindi syllabification tells us that the hardness of the syllables must increase in the following order: *V, CV, VC, CVC, CCV, CCVC, VCC* and so on. One possible encoding of such a hardness scale, in the lines of OT, is through constraint-ranking of syllable structures, such that the harder a syllable is, the higher is its rank. The optimal syllabification in this scheme is defined as the one which violates the least number of constraints (i.e. possess as few hard syllables as possible). Moreover, the violation of a higher ranked constraint is considered costlier than the violation of any number of lower ranked constraints.

Nevertheless, here we take a different approach that can not only model the principles of Hindi syllabification elegantly within the framework of constrained-optimization, but also yield to simple algebraic analysis that helps us to prove several properties of the syllabification pattern. As far as our knowledge goes, this is a novel approach to syllabification and is based on the *reductionist* principle, where the hardness of a syllabification is defined in terms of the hardness of its constituent syllables, which in turn is defined in terms of the hardness of the constituent segments. We define the hardness function in a bottom-up fashion.

Definition 4.8 *The structure of a syllable σ, denoted by $CVMap(\sigma)$ is a string over $C^*V^*C^*$, where each phoneme x of σ is mapped to C if $x \in C_P$, or V if $x \in V_P$.*

Definition 4.9 *Let $\sigma = x_1x_2\cdots x_s$ be a syllable of length s, where $x_v \in V_P$ is* *the vowel, for some integer v satisfying $1 \leq v \leq s$. We define the hardness of a phoneme (segment) x_i, denoted by $h(x_i)$ as*

$$h(x_i) = \begin{cases} 2^{2(v-i-1)} & \text{if } i < v \\ 0 & \text{if } i = v \\ 2^{2(i-v)-1} & \text{if } i > v \end{cases}$$

Thus, the farther a phoneme is from the nucleus, the harder it is to perceive and/or articulate. For two phonemes at the same distance from the nucleus, the one which is in the coda is harder than the one in the onset. See (Ohala 1990a; Redford 1999; Redford and Diehl 1999) and references therein for empirical evidence in favor of the aforementioned principles; also see (Redford et al. 1998; Joanisse 1999; Redford et al. 2001) for computational models of emergence of syllable systems based on these principles.

Definition 4.10 *The hardness of a syllable $\sigma = x_1 x_2 \cdots x_s$, denoted by $H_\sigma(\sigma)$, where $H_\sigma : C_P^* V_P C_P^* \rightarrow \mathbb{N}$ is the sum of the hardness of its phonemes.*

$$H_\sigma(\sigma) = \sum_{i=1}^{s} h(x_i)$$

Definition 4.11 *The hardness of a syllabification $\psi = \sigma_1 \sigma_2 \cdots \sigma_m$, denoted by $H_\psi(\psi)$, is the sum of the hardness of its syllables.*

$$H_\psi(\psi) = \sum_{i=1}^{m} H_\sigma(\sigma_i)$$

This metric of hardness is just a hypothetical measure, which allows us to compare two syllables in terms of their functional hardness and is by no means an absolute measure of the degree of hardness based on the syllable structure of a syllable. Note that for two distinct syllables σ and σ', if $CVMap(\sigma) = CVMap(\sigma')$, then $H_\sigma(\sigma) = H_\sigma(\sigma')$. It is also easy to see that H_σ defines a total ordering over the set all syllable structures (i.e. C*VC*). Table 4.1 enlists the first few syllable structures in increasing order of the value of H_σ.

Syllable Structure	H_σ	Syllable Structure	H_σ
V	0	CV	1
VC	2	CVC	3
CCV	5	CCVC	7
VCC	10	CVCC	11

Table 4.1: H_σ defines a total ordering on syllable structures. The first few syllable structures in ascending order of their hardness

Definition 4.12 *Let $w_P \in \Sigma_P^*$ be a phonetic word and let $\Psi(w_P)$ be the set of all valid syllabifications of w_P. A valid syllabification $\psi_0 \in \Psi(w_P)$ is also an optimal syllabification of w_P if and only if* $\nexists \psi \in \Psi(w_P), H_\psi(\psi) < H_\psi(\psi_0)$

Thus, the optimal syllabification for a word is the valid syllabification which has the least hardness. This concept is explained through the following example.

Example 4.2 For the phonetic word $w_P = samprati$, the set of valid syllabifications is $\Psi(w_p) = \{samp{-}ra{-}ti, samp{-}rat{-}i, sam{-}pra{-}ti, sam{-}prat{-}i\}$. The hardness of the syllabifications are as follows.

$$\begin{aligned}
H_\psi(samp - ra - ti) &= H_\sigma(samp) + H_\sigma(ra) + H_\sigma(ti) = 11 + 1 + 1 = 13 \\
H_\psi(samp - rat - i) &= H_\sigma(samp) + H_\sigma(rat) + H_\sigma(i) = 11 + 3 + 0 = 14 \\
H_\psi(sam - pra - ti) &= H_\sigma(sam) + H_\sigma(pra) + H_\sigma(ti) = 3 + 5 + 1 = 9 \\
H_\psi(sam - prat - i) &= H_\sigma(sam) + H_\sigma(prat) + H_\sigma(i) = 3 + 7 + 0 = 10
\end{aligned}$$

Since, the hardness is lowest for $/sam - pra - ti/$, this is the optimal syllabification. □□

One of the interesting and useful consequences of defining the optimal syllabification in this way is the fact that for any string w_P, the optimal syllabification is unique. This fact is formally stated in theorem 4.1 and the proof is presented subsequently.

Theorem 4.1 *For a phonetic word $w_P \in \Sigma_P^*$, if the set of valid syllabifications $\Psi(w_P)$ is non-empty, then there exists one and only one optimal syllabification $\psi_0 \in \Psi(w_P)$. We shall denote the optimal syllabification of w_P as $OPT_\psi(w_P)$.*

Proof: That there exists at least one optimal syllabification of $w_P = x_1 x_2 \cdots x_n$ $\psi_0 \in \Psi(w_P)$ is obvious. We prove the uniqueness of the optimal syllabification by providing a construction methodology for ψ_0.

Note that every syllabification can also be uniquely specified by the sequence of the indices of the characters of w_P after which there is a syllable break. Let us denote such sequences by $\xi_0 \xi_1 \xi_2 \cdots \xi_{m-1} \xi_m$, where ξ_0 is always 0, since we can assume a hypothetical syllable boundary before the word beginning. Similarly, $\xi_m = \mid w_P \mid$, because we can assume a hypothetical syllable boundary after the last phoneme of the word. Here, m is the number of syllables in w_P, and is equal to the number of vowels in w_P.

Let $\psi = \xi_0 \xi_1 \cdots \xi_m$ be a valid syllabification of w_P and let v_i be the index of the i^{th} vowel in w_P, $1 \leq i \leq m$. Therefore, v_i satisfies the following relation.

$$\xi_{i-1} < v_i \quad \leq \quad \xi_i, \quad \forall i,\ 1 \leq i \leq m \tag{4.2}$$

Also, the i^{th} syllable of ψ, σ_i, is $x_{\xi_{i-1}+1} \cdots x_{\xi_i}$. Thus, we have

$$\begin{aligned}
H_\psi(\psi) &= \sum_{i=1}^{m} H_\sigma(\sigma_i) \\
&= \sum_{i=1}^{m} H_\sigma(x_{\xi_{i-1}+1} \cdots x_{\xi_i}) \\
&= \sum_{i=1}^{m} \sum_{k=\xi_{i-1}+1}^{\xi_i} h(x_k) \\
&= \sum_{i=1}^{m} \left[\sum_{k=0}^{v_i-\xi_{i-1}-2} 2^{2k} + \sum_{k=0}^{\xi_i - v_i} 2^{2k+1} \right] \\
&= \frac{1}{3} \sum_{i=1}^{m} \left[4^{(v_i-\xi_{i-1}-1)} - 1 + 2 \times (4^{\xi_i - v_i} - 1) \right] \\
&= \frac{1}{3} \left[4^{v_0-1} + 2 \times 4^{n-v_m} - 3m \right] + \frac{1}{3} \sum_{i=1}^{m-1} \left[4^{v_{i+1}-\xi_i-1} + 2 \times 4^{\xi_i - v_i} \right]
\end{aligned}$$

$$\therefore H_\psi(\psi) = K_c + \frac{1}{3}\sum_{i=1}^{m-1} f_i(\xi_i) \tag{4.3}$$

where,

$$\begin{aligned} K_c &= \frac{1}{3}\left[4^{v_0-1} + 2 \times 4^{n-v_m} - 3m\right], \text{ is independent of } \xi_i s \\ f_i(\xi_i) &= 4^{v_{i+1}-\xi_i-1} + 2 \times 4^{\xi_i - v_i} \end{aligned}$$

Note that $v_i \leq \xi_i < v_{i+1}$ for all the ξ_is. In other words, the range of ξ_is are non-overlapping, and therefore, their values can be chosen independently. This independence condition along with Eqn. 4.3 implies that $H_\psi(\psi)$ is minimized, when the terms $f_i(\xi_i)$ is minimized for each i. $f_i(\xi_i)$ is the sum of two terms, whose product

$$4^{v_{i+1}-\xi_i-1} \times 2 \times 4^{\xi_i - v_i} = 2 \times 4^{v_{i+1}-v_i-1}$$

is a constant for a given w_P and i. Stated differently, the product of the terms are independent of ξ_i. Therefore, by the AM-GM inequality (Rudin 1987) $f_i(\xi_i)$ is minimum when the two terms, i.e., $4^{v_{i+1}-\xi_i-1}$ and $2 \times 4^{\xi_i - v_i}$, are equal or as close as possible. This happens for

$$\xi_i = \left\lfloor \frac{v_{i+1} - v_i}{2} \right\rfloor$$

However, assigning ξ_i a value as described above may violate the PCs of the language. If the resulting coda is illegal, that is

$$C_{coda}(x_{v_i+1} \cdots x_{\xi_i}) = 0$$

Then ξ_i is to be redefined in such a way that for the new value of ξ_i

$$C_{coda}(x_{v_i+1} \cdots x_{\xi_i}) = 1, \text{ but } C_{coda}(x_{v_i+1} \cdots x_{\xi_i+1}) = 0$$

Intuitively, it refers to the process of minimally shifting the syllable boundary to the left, so that the resulting coda to the left of the syllable boundary is phonotactically valid. Similarly, one can optimally choose ξ_i if the C_{onset} constraints are not satisfied. However, for any given choice of ξ_i, it is not possible that both the onset and coda constraints are simultaneously violated. This is due to the substring validity property of PCs, which in this case implies that shifting boundary to the left would always violate the resulting onset to its right and shifting the boundary to the right would always violate the resulting coda to its left.

In conclusion, there is always a unique choice of the values of ξ_i for $1 \leq i \leq m-1$, that minimizes $H_\psi(\psi)$. This choice uniquely defines the optimal syllabification ψ_0.

$\Box\Box$

In essence, the optimal syllabification is obtained by placing the syllable boundary in the middle of the word-medial consonant clusters. In case the number of consonants in the cluster is odd, the boundary is so chosen that the length of the resulting coda of the syllable to the left is one less than the length of the onset of the syllable to the right. Thus, for VCCV, the resulting syllabification is VC-CV, whereas for VCCCV, it is VC-CCV.

Thus, we have defined a syllabification for Hindi within the constrained optimization framework, where the constraints are 1) phonotactic legality of the onset and coda clusters and 2) the well-formedness or validity of the syllabification, and the optimization criterion is minimization of the hardness function H_ψ.

4.3 Acoustic Distinctiveness

Theoretical and computational models of language change have emphasized on the property of acoustic distinctiveness, which states that the linguistic entities like phonemes, syllables or words must be maximally distinct, so that the probability of confusion between pairs remains small resulting in a high rate of successful communication (Liljencrants and Lindlom 1972; de Boer 2001). In the case of schwa deletion, acoustic distinctiveness restricts the deletion of schwas in the contexts, where the deletion might result in an incorrect interpretation of the word. This is illustrated in Fig. 4.2. Suppose that there are n words $w_{P1}, w_{P2}, \cdots, w_{Pn}$, represented as points in a d dimensional acoustic

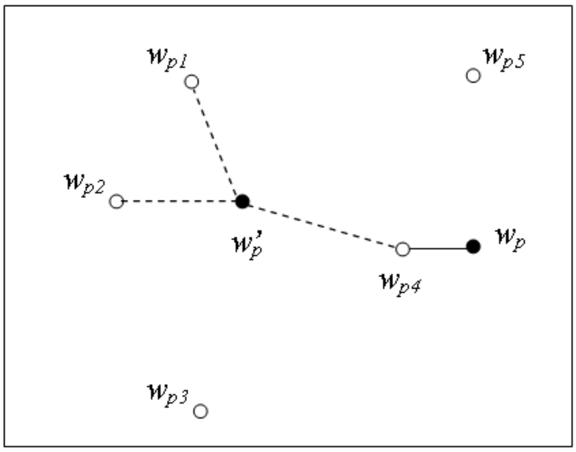


Figure 4.2: Words in a two dimensional acoustic space. The black dots represent the realization of the word w_{P4}. The lines indicate proximal words. The solid line shows that there is only one word that is close to w_p. The broken line shows that there are several words close to w_p and consequently a potential case for misunderstanding.

space. Deletion or shortening of the schwa leads to a realization of an acoustic signal w'_P that is distinct from the original word (w_{P4} in the figure). If the distance of this signal from the intended word (w_{P4}) in the acoustic space is comparable to other words (such as w_{P2} and w_{P1}), then the probability of incorrect interpretation goes up. On the other hand, if the acoustic signal is distinctively closer to the intended word (such as w_P), then there will be very little scope for wrong interpretation, making the communication successful. Thus, functional forces will work in favor of w_P and not w'_P, in course of language change.

A complete specification of the acoustic distinctiveness features between words would require an appropriate definition of the acoustic space and a realistic distance metric. Steriade (2001) suggests a hierarchical representation of the acoustic distinctiveness between syllables and segments through the use of P-maps. In this work, we adopt a simplified version of the concept and appropriately modify the same to model the pattern of SDH.

It is a well known fact that perceptibility of phonemes at the onset of a syllable is much higher than that in the coda (Ohala 1990a; Redford 1999; Redford et al. 2001). Moreover, stressed syllables are better perceived than unstressed one. Therefore, if two syllables have different phonemes/phoneme clusters at onset, the probability that they are perceived differently is much higher. In other words, they are acoustically distinct. Similarly, if two stressed syllables differ at any position, the difference is pronounced. On the other hand, if two syllables differ in their coda positions and both are unstressed, they are acoustically closer or less distinct. Thus, we can define an acoustic distance metric $D_\sigma : C_P^* V^* C_P^* \times C_P^* V^* C_P^* \to \Re$, such that

$$\begin{aligned} & & D_\sigma(\alpha\vartheta\beta, \alpha\vartheta\beta) &= 0 && (4.4)\\ 0 &< & D_\sigma(\alpha\vartheta\beta, \alpha\vartheta\beta') &\leq d_c && (4.5)\\ d_c &< & D_\sigma(\alpha\vartheta\beta, \alpha'\vartheta\beta'') &\leq d_o && (4.6)\\ d_o &< & D_\sigma(\alpha\vartheta\beta, \alpha''\vartheta'\beta'') & && (4.7)\end{aligned}$$

where,

$$\begin{array}{llll} \alpha, \alpha', \alpha'' \in C_P^* & \beta, \beta', \beta'' \in C_P^* & \vartheta, \vartheta' \in V_P & d_c, d_o \in \Re^+ \\ \alpha \neq \alpha' & \beta \neq \beta' & \vartheta \neq \vartheta' & d_c < d_o \end{array}$$

Although we have defined D_σ as a real valued function, it is essentially a ranking over the different syllable types. In the case of SDH, it suffices to define two syllables σ and σ' as distinct if $D_\sigma(\sigma, \sigma') > d_c$. Given two phonetic words w_P and w'_P, they can be considered to be acoustically non-distinct, if the corresponding syllables in their optimal syllabifications are acoustically non-distinct. However, since the deletion of schwas leads to a reduction in the number of syllables, we allow insertion of null syllables in the optimal syllabifications, so that the sequences can be properly aligned and compared. Moreover, an unstressed syllable in which schwa is the nucleus can be deleted altogether; therefore, a syllable with schwa can be considered to be acoustically non-distinct from a null syllable, if the former is unstressed. This is formally defined below.

Definition 4.13 *We define two syllables σ and σ', both of which can be null, as non-distinct if any one of the following conditions hold.*

1. *If both of them are null.*

2. *If none of them are null and* $D_\sigma(\sigma, \sigma') \leq d_c$

3. *If both of them are unstressed and one of them is null and the other one has a schwa in the nucleus.*

Definition 4.14 *Let* w_P *and* w'_P *be two phonetic words. Acoustic distinctiveness constraint* $C_{AD} : \Sigma_P^* \times \Sigma_P^* \rightarrow \{0, 1\}$ *maps the tuple* $\langle w_p, w'_p \rangle$ *to 1 (i.e. the constraint is satisfied) if and only if there exists an alignment* $\sigma_1\sigma_2 \cdots \sigma_m$ *and* $\sigma'_1\sigma'_2 \cdots \sigma'_m$ *between the sequence of syllables* $OPT_\psi(w_P)$ *and* $OPT_\psi(w'_P)$ *such that* σ_i *and* σ'_i *is non-distinct for all* i, $1 \leq i \leq m$.

The following example illustrates the concept.

Example 4.3 Let w_P be the Hindi word /*amara*/. There are three schwas, deleting each we get three phonetic words, whose optimal syllabifications, along with that of w_P are given below.

$$
\begin{aligned}
w_P &= /amara/ & OPT_\psi(w_P) &= /a - ma - ra/ \\
w_{P1} &= /mara/ & OPT_\psi(w_{P1}) &= /ma - ra/ \\
w_{P2} &= /amra/ & OPT_\psi(w_{P2}) &= /am - ra/ \\
w_{P3} &= /amar/ & OPT_\psi(w_{P3}) &= /a - mar/
\end{aligned}
$$

Being a bound stress language, in Hindi, the stress is always on the first syllable. Therefore, we get the following alignments between w_P and the other words.

w_P	a	ma	ra
w_{P1}	ϕ	ma	ra
w_{P2}	am	ϕ	ra
w_{P3}	a	mar	ϕ

Here, ϕ represents the null syllable. There exists no alignment between w_P and w_{P1} that respect the acoustic distinctiveness constraint. Although, both w_{P2} and w_{P3} are allowable by the acoustic distinctiveness constraints, the latter is the accepted pattern in Hindi. □□

4.4 SDH as Constrained Optimization Problem

In this section, we formulate SDH as a constrained-optimization problem and analytically solve the model to obtain the schwa deletion pattern of Hindi. The formulation of the problem is straight forward.

Formulation of SDH as a Constrained Optimization Problem: Given a phonetic word $w_P \in \Sigma_P^*$, $f_{DS}(w_P)$ is a subsequence of w_P obtained by deleting zero or more schwas from w_P, such that

- *Constraint*: $C_{AD}(w_P, f_{DS}(w_P)) = 1$
- *Optimization Criterion*: $f_{DS}(w_P)$ has minimum number of syllables (or equivalently phonemes).

Thus, SDH has been posed here as a syllable minimization problem, which is equivalent to maximization of the number of schwas being deleted, subjected to the acoustic distinctiveness constraint. Note that since the acoustic distinctiveness constraint is defined in terms of the optimal syllabification, syllabification is essentially a subproblem of SDH, which again has been modeled as a constrained-optimization problem.

We state and prove below the most important property of the problem of SDH as formulated above.

Theorem 4.2 *$f_{DS}(w_P)$ is obtained from w_P by deleting all the schwas in w_P that conform to the context specified by Eq 2.1.*

Proof: We prove the result in two steps. First we show that the deletion of a schwa in the context specified by Eq 2.1 does not violate C_{AD}, and then we show that deletion of a schwa in any other context violates the constraint.

Case Ia: $V_1C_1aC_2V_2 \rightarrow V_1C_1C_2V_2$; The optimal syllabification of w_P is $V_1 - C_1a - C_2V_2$, whereas the optimal syllabification of the string after deletion of the schwa is $V_1C_1 - C_2V_2$. The alignment of the syllables are shown below. It is easy to see that all the alingments respect the non-disctinctive property and therefore, the C_{AD} constraint holds good.

$$\begin{array}{ccc} V_1 & C_1a & C_2V_2 \\ V_1C_1 & \phi & C_2V_2 \end{array}$$

The non-distinctive alignments for the other three cases are shown below.

Case Ib: $V_1C_1C_2aC_3V_2 \rightarrow V_1C_1C_2C_3V_2$

$$\begin{array}{ccc} V_1C_1 & C_2a & C_3V_2 \\ V_1C_1C_2 & \phi & C_3V_2 \end{array}$$

Case Ic: $V_1C_1a \rightarrow V_1C_1$

$$\begin{array}{cc} V_1 & C_1a \\ V_1C_1 & \phi \end{array}$$

Case Id: $V_1C_1C_2a \rightarrow V_1C_1C_2$

$$\begin{array}{cc} V_1C_1 & C_2a \\ V_1C_1C_2 & \phi \end{array}$$

This completes the first part of the proof. To prove that deletion is possible in no other context, we classify the remaining contexts into four cases.

Case IIa Word initial schwa. Since the word initial syllable is stressed, the word initial schwa is also stressed and cannot be aligned to the null syllable. Consequently, a schwa in this position cannot be deleted.

Case IIb Schwa followed by consonant cluster: We show the case for two consonants; $V_1C_1aC_2C_3V_2 \rightarrow V_1C_1C_2C_3V_2$. There is no non-distinctive alignment between the two strings, and therefore, the acoustic distinctiveness constraint is violated (the last column is distinctive).

$$\begin{array}{ccc} V_1 & C_1aC_2 & C_3V_2 \\ V_1C_1 & \phi & C_2C_3V_2 \end{array}$$

The proof for more than two consonants in the following cluster is similar.

Case IIc Schwa preceded by more than two consonant: We show the case for three consonants; $V_1C_1C_2C_3aC_4V_2 \rightarrow V_1C_1C_2C_3C_4V_2$. Here also, there is no non-distinctive alignment between the two strings, and therefore, the acoustic distinctiveness constraint is violated (the last column is distinctive).

$$\begin{array}{ccc} V_1C_1 & C_2C_3a & C_4V_2 \\ V_1C_1C_2 & \phi & C_3C_4V_2 \end{array}$$

Moreover, longer sequences of consonants are rare and usually violate the PCs if the intermediate schwa is deleted.

Case IId schwa is immediately followed by vowel: The case is illustrated below. Here also A_{DC} is violated. $V_1C_1aV_2 \rightarrow V_1C_1V_2$.

$$\begin{array}{ccc} V_1 & C_1a & V_2 \\ V_1 & \phi & C_1V_2 \end{array}$$

The only case where it is possible to delete the schwa even though the rule does not allow the deletion is $V_1aC_1V_2 \rightarrow V_1C_1V_2$. Nevertheless, words featuring this pattern are rare, and the apparent discrepancy may be due to incorrect syllabification. Instead of $V_1 - a - C_1V_2$, the preferred syllabification in such cases may be $V_1a - C_1V_2$, which is impossible to incorporate in our model due to the very definition of syllable. □□

Thus, we have shown that the context of schwa deletion described by Eq 2.1 is necessary as well as sufficient to describe the schwa deletion pattern emerging according to the constrained-optimization based formulation of the problem. Also note that the optimization criterion regarding syllable minimization states that all the schwas whose contexts permit deletion must always be deleted. However, deletion of a schwa might destroy the context of deletion for another schwa. In such a case, our model suggests that the deletion operation must be carried out from the direction (left-to-right or right-to-left), which facilitates deletion of a larger number of schwas. There are cases, where deletion from left-to-right produces a different pattern than deletion from right-to-left, even though both the patterns have same number of syllables. In such cases, the model does not provide a unique schwa deletion pattern for an input string.

4.4.1 Constraints on the properties of syllabification

In Sec. 4.2.2 we have proposed a definition of optimal syllabification in Hindi, and we have shown that under the assumptions of the optimal syllabification and the acoustic distinctiveness constraint, one can derive the context for SDH as is suggested by Ohala's rule. Nevertheless, an issue worth investigating is what should be the formal properties of the syllabification of Hindi so as to yield the context of schwa deletion under the acoustic distinctiveness assumption.

It can be shown that a syllabification based on the onset maximization principle is incapable of giving rise to the desired pattern. This is because, for the string $V_1C_1aC_2V_2$, the syllable boundaries according to the onset maximization principle will be placed as $V_1 - C_1a - C_2V_2$. Schwa deletion in this context will lead to a resyllabification of the form $V_1 - C_1C_2V_2$, violating the acoustic distinctiveness constraint. Therefore, schwa deletion should not be permitted in such a context. Similarly, for $V_1 - C_1a - C_2C_3V_2$, deletion almost always leads to $V_1C_1 - C_2C_3V_2$, because $V_1 - C_1C_2C_3V_2$ is expected to violate the PCs of Hindi (i.e. there are very few consonant clusters of size three that are allowable at the onset). Hence, deletion of schwa in this context will not violate the acoustic distinctiveness constraint, and therefore, should be permissible. However, in reality, the context of SDH is just the reverse of it and therefore, onset maximization-based syllabification cannot explain SDH.

In fact, it is possible to derive the desired context for SDH for any syllabification scheme that predicts a syllable boundary of the pattern VCC^*V as $VC - C^*V$. This fact provides strong evidence against onset maximization-based syllabification in Hindi. Nonetheless, one must be aware that under a different definition of the acoustic distinctiveness constraint, it might be possible to show otherwise.

4.5 Algorithm for SDH

Apart from providing a functional explanation for SDH, the constrained optimization model also provides us with a framework to algorithmically compute the schwa deletion pattern. In other words, the formulation of f_{DS} as a constrained optimization problem implicitly captures a computational definition of the same. This is of practical importance, because schwa deletion is a challenging issue in Hindi G2P conversion. In this section, we present a rule based algorithm for SDH and empirical evaluation of the same.

The simplest way to implement f_{DS} is through exhaustive search. Given a word $w_G \in \Sigma_G^*$, we generate all possible schwa deleted forms of w_G. If there are k schwas in w_G, then there are 2^k such forms. For each of these forms, we check whether the acoustic distinctiveness constraint holds good. This in turn requires syllabification of w_G as well as the schwa deleted form. It may be mentioned at this point that the formulation of syllabification as an optimization problem also provides us with a method to compute the syllable boundaries. In fact, the constructive proof for Theorem 4.1 entails the method. Out of the word forms that conform to the distinctiveness constraint, the one(s) with the minimum number of syllables is/are finally chosen as the output(s).

The exhaustive search technique is correct, but inefficient. However, we have seen that the context of schwa deletion specified by Eq 2.1 is identical to that predicted by the constrained optimization model. This fact can be used to design an efficient implementation of f_{DS}. The basic idea is summarized below.

1. Scan the word from left-to-right. For every schwa encountered check whether
   a. its local context satisfies the condition specified by Eq 2.1?
   b. the deletion of this schwa does not violate the PCs?
2. If both the conditions are satisfied, then delete the schwa.
3. Continue till the end of the word. Let the transformed word so obtained be w_l.
4. Repeat steps 1-3, but now scanning from right-to-left. Let the transformed word so obtained be w_r.
5. Among w_l and w_r, choose the one with smaller number of schwas.

The schwa deleted form obtained by following the above algorithm does not violate the acoustic distinctiveness constraint. The optimality of the output is ensured by running the algorithm from both the directions. However, more often than not w_l is identical to w_r, and therefore, it suffices to carry out the scan only in one direction. In fact, Ohala's rule claims that the rule always applies from right-to-left.

We describe below an algorithm – DeleteSchwa – for SDH based on the aforementioned idea. Rather than modeling the PCs and contextual constraints separately, the algorithm combines these two types of constraints and checks them through a small look-ahead. The word is scanned twice from left-to-right. Thus, the complexity of the algorithm is O(n), where n is the length of the input word. The algorithm DeleteSchwa has a high accuracy for monomorphemic words.

4.5.1 Algorithm for monomorphemic words

For the description of the algorithm, we shall take the help of a notation called *half* – $\mathcal{H}$ and *full* – $\mathcal{F}$ sounds. We define a *full* sound as a consonant-vowel pair or a vowel alone, whereas *half* sound as a pure consonant sound, without any immediately following vowel. Therefore, any vowel or a consonant followed by a vowel is a *full* sound, whereas a consonant followed by *halant* (i.e. the consonants of a cluster, except the last one) are half sounds. Hence, when a schwa following a consonant is deleted, it becomes half, but if it is retained, the consonant is full. Since the nature of the consonants followed by schwa might not be known beforehand, we shall call such consonants as *unknown* - $\mathcal{U}$.

To illustrate this point, consider the example of *bachapana* cited before. Here, b is $\mathcal{F}$, n is $\mathcal{H}$ but ch and p are $\mathcal{U}$. In the algorithm, only the consonants and full vowels will be marked $\mathcal{H}$, $\mathcal{F}$ or $\mathcal{U}$, but the *mAtrAs* (i.e. the vowels) will not be marked.

Note that after marking the consonants of the word according to the rules stated above, only the consonants immediately followed by schwas can be marked as $\mathcal{U}$. The algorithm scans the marked word from left to right replacing each of the $\mathcal{U}$ s by either $\mathcal{F}$ or $\mathcal{H}$, depending on the two adjacent syllables of that particular $\mathcal{U}$-marked consonant. At the end of the algorithm, schwas following the consonants marked as $\mathcal{H}$ are deleted.

The formal steps of the algorithm DeleteSchwa are shown in below.

DeleteSchwa(w_G)

1 Mark all the full vowels and consonants followed by vowels other than the inherent schwas in the word as $\mathcal{F}$.
2 Mark all the h in the word as $\mathcal{F}$.
3 Mark all the consonants immediately followed by consonants or halants (i.e. consonants of conjugate syllables) as $\mathcal{H}$.
4 Mark all the remaining consonants, which are followed by implicit schwas as $\mathcal{U}$.
5 If in the word, y is marked $\mathcal{U}$ and preceded by i, I, rri, u or U, mark it $\mathcal{F}$.
6 If y, r, l or v are marked $\mathcal{U}$ and preceded by consonants marked $\mathcal{H}$, then mark them $\mathcal{F}$.
7 If a consonant marked $\mathcal{U}$ is followed by a free vowel, then mark that consonant as $\mathcal{F}$.
8 While traversing the word from left-to-right, if a consonant marked $\mathcal{U}$ is encountered before any consonant or vowel marked $\mathcal{F}$, then mark that consonant as $\mathcal{F}$.
9 If the last consonant is marked $\mathcal{U}$, mark it $\mathcal{H}$.
10 If any consonant marked $\mathcal{U}$ is immediately followed by a consonant marked $\mathcal{H}$, mark it $\mathcal{F}$.
11 While traversing the word from left-to-right, for every consonant marked $\mathcal{U}$, mark it $\mathcal{H}$ if it is preceded by $\mathcal{F}$ and followed by $\mathcal{F}$ or $\mathcal{U}$; otherwise mark it $\mathcal{F}$.
12 For all consonants marked $\mathcal{H}$, if it is immediately followed by (an implicit) schwa in the input word, then delete the schwa from the word.
13 **return** the resulting word after the above transformations.

4.5.2 Modification for Polymorphemic Words

Ohala's rule states that SDH respects the morpheme boundaries. This is illustrated by the word pairs *dha.Dakane* and $dha.DakaneM$, where the former is pronounced as $dha - .Dak - ne$ and the latter $dha.D - ka - neM$. The reason for these is that while *dha.Dakane* is derived from the root verb *dha.Daka* by affixing the case-ending *ne*, the latter is obtained by adding the plural marker *eM* to the noun root *dha.Dakana*. Ohala's rule states that except for the case of suffixes beginning with vowels, the algorithm DELETESCHWA must be separately applied to the individual morphemes and the resulting pronunciations must be concatenated in order to obtain the correct pronunciation of the polymorphemic word. The details of the rules in the case of polymorphemic words are enumerated below.

- Compound words are formed by concatenation of two or more words. Each of the words retain their original pronunciation, so DELETESCHWA is applied separately on the words and the results are simply concatenated to get the pronunciation of the compound word. For example, *charaNakamala* → (after morphological analysis, + represents morpheme boundary) *charaNa+kamala* → (after individual schwa deletion) *charaN* and *kamal* → (after concatenation) *charaNkamal*. On the other hand, without morphological analysis, the result would have been *charNakmal*, which is incorrect.

- For prefixes, the rule is identical to that above. E.g. *pra+gati* → *pragati* (and not *pragti*) or *a* + *samaya* → *asamay* (and not *asmay*).

- Suffixes that begin with a consonant are simply juxtaposed at the end of the stem as in rule 1 above. However, if the suffix begins with a vowel, DELETESCHWA is applicable to the whole word instead of the stem and the suffix individually. For example, *arab* + *I* → *arbi* (and not *arabi*, as would be the case if the morphemes were treated separately). Similarly, *namak* + *Ina* → *namkIn* (and not *namakin*).

- Stems, which have a conjugate syllable in the second last position, are exceptions to rule 3. For words derived from such stems, schwa deletion is separately applicable to the stem and the affix. For example, *nindak* + *oM* → *nindakoM* (and not *nindkoM*, as would be the case if rule 3 was followed).

Based on these rules, we extend the algorithm DELETESCHWA appropriately to handle the case of polymorphemic words. This algorithm, called

ModifiedDeleteSchwa, is described below.

ModifiedDeleteSchwa(w_G)

1 Analyze the morpheme boundaries of the word using morphological analyzer.
2 **if** the word is monomorphemic, **return** DeleteSchwa(w_G).
3 **else** apply algorithm DeleteSchwa(w_{Gi}) to the individual morphemes w_{Gi} as suggested by the rules above and concatenate the outputs accordingly.
4 **return** the resulting word after the above transformations.

Note that ModifiedDeleteSchwa presumes the knowledge of morpheme boundaries, which can be automatically obtained with the help of a morphological analyzer.

4.5.3 Evaluation

The algorithms DeleteSchwa and ModifiedDeleteSchwa have been evaluated on a set of around 11095 commonly used Hindi words obtained from a standard lexicon[5]. The output of the algorithms was manually judged by a native speaker of Hindi. The word level accuracy of DeleteSchwa and ModifiedDeleteSchwa has been found to be 96% and 99% respectively. Table 4.5.3 shows the break-up of the fraction of errors due to various morphological processes for DeleteSchwa. We observe that the majority of the errors are due to inflections and compound words. In this context it may be mentioned that it is easier to develop an inflectional morphological analyzer for Hindi than a compound word analyzer.

When the morpheme boundaries are correctly known, the algorithm ModifiedDeleteSchwa achieves a near perfect accuracy. In fact, the output of ModifiedDeleteSchwa is erroneous only for the loans and very infrequently used words of Hindi.

Previously, there has been some work on SDH from a computational perspective (Kishore et al. 2002; Kishore and Black 2003; Narasimhan et al. 2004;

[5] *Hindi Bangla English Tribhasa Abhidhaan*, Sandhya Publication, 1^{st} Edition, March 2001

Morphological process	Fraction of error
Derivational suffix	10.2%
Derivational prefix	13.7%
Inflectional suffix	30.2%
Compound word	43.1%
Others	2.8%

Table 4.3: Break up of error due to morphological processes for DeleteSchwa

Bali et al. 2004). However, only (Narasimhan et al. 2004) contains a detailed treatment of the problem. The work combines Ohalas rule (Ohala 1983b) and morphological analysis with finite state transducers (Kaplan and Kay 1994) and cost models. Initially the algorithm generates all possible output candidates for a given input, following Ohalas rule on possible contexts for schwa deletion. Then certain candidates, which violate phonotactic constraints, are filtered out. Among the remaining candidates, the one with the minimum cost according to the cost model is selected as the final output, where the cost mainly takes care of the deletion of the word final schwa. The authors report an accuracy of 89%. This figure, however, is not comparable to that of ours, because in (Narasimhan et al. 2004) the accuracy is measured in terms of schwas rather than words.

4.6 Conclusion

In this chapter, we have proposed a general framework for modeling language change from the perspective of functional phonology and described a model for SDH within the framework. Since the present orthography of Hindi faithfully reflects the historical pronunciations, the diachronic model proposed here for SDH can also be considered as a synchronic model. This in turn allows us to develop an algorithm for SDH that is useful from the perspective of G2P converter and speech synthesis systems. The algorithm for SDH proposed here takes time linear in the length of the input word, and achieves a word-level accuracy of 96% that can be further boosted up to 99% by using a morphological analyzer for the language.

There are a few important issues regarding the work presented in this chapter that are worth discussing. First, *in what sense the model for SDH proposed here reflects language change?* Second, *since the constraints and optimization criteria are to a large extent reverse-engineered to fit our needs, what is the linguistic plausibility of the explanation presented?* and thirdly, *how does this work further our knowledge regarding SDH, when facts like Ohala's rule and principles of syllable economy are already known?*

The answer to the first question can be sought for in *evolutionary phonology* (Blevins 2004), which states that the recurrent synchronic patterns in phonology are those that are outcomes of common sound changes. Stated differently, if a linguistic form X is transformed to a form Y through some regular phonological change, then the synchronic pattern Y can be explained by positing an underlying form X. The regularity of the phonological change in this context can then be viewed as the part of the synchronic grammar that maps the underlying form X to the surface form Y. Indeed, this alternative view to language change (as proposed in evolutionary phonology) can be used to design several NLP applications including machine translation systems (discussed in Chapter 7), and therefore, can be an important and useful paradigm for NLP. In the specific context of SDH, the proposed syllable minimization-based model can be considered as a synchronic as well as diachronic explanation. However, we argue in favor of the diachronic view as it provides a causal explanation making any further synchronic propositions unnecessary and extraneous (see Sampson (1970) and Chapter 1 in Blevins (2004) for a discussion).

To answer the second question, it is true that the precise formulation of the objective functions is primarily guided by the schwa deletion pattern of Hindi. Stated differently, there are several other possible formulation of SDH that are equally grounded in linguistics, but would predict an altogether different deletion pattern. Nevertheless, we emphasize the fact that the objective functions and constraints in our analysis have been kept as simple and realistic (i.e. independently grounded in linguistics) as possible, and nowhere anything has been assumed that overtly captures the deletion pattern. We have also attempted to introduce a general form for each of the functions, after which language specific parameters (like thresholds) are set. Furthermore, we also identify general properties of the functions (e.g. optimal syllabification) necessary and/or sufficient to give rise to Ohala's rule, which have been independently verified by other researchers in other contexts (such as the coda maximization principle for Hindi). Therefore, we feel that the present formulation of the problem is linguistically plausible as well as insightful.

Some points regarding the contribution of the current work, as raised by the third question, have already been discussed above. Besides those, an important aspect of the present work is its mathematical elegance. SDH has been analyzed and modeled within a single framework and this has allowed us to formally prove several properties of the problem through simple algebraic techniques. We believe that for a wide range of linguistic phenomena, similar formulations and proofs are possible within the proposed framework.

The general framework proposed here is based on constrained optimization, and is comparable to some of the previous models (Liljencrants and Lindlom 1972; Prince and Smolensky 1993; Prince and Smolensky 1997; Boersma 1997a; Boersma 1997b; Boersma 1998) in synchronic and diachronic linguistics that try to represent a linguistic phenomenon as an optimization problem. Nevertheless, unlike the previous models, the framework proposed here does not try to combine the objective functions into a single optimization criterion by, for example, ranking of constraints (Prince and Smolensky 1993); rather the problem of language change has been proposed as a multi-objective and multi-constrained problem, where the constraints and objectives operate on the linguistic forms, independent of each other. This, we believe, is the major strength of the current framework that enables one to model a wide range of linguistic phenomena through this framework. In Chapter 7 we shall revisit this multi-objective multi-constraint optimization framework and demonstrate its power and effectiveness in the context of the morpho-phonological change of Bengali verb inflections.

One of the strongest criticisms against functional explanations is that they are goal-directed (Ohala 1974; Ohala 1987; Ohala 1989). Indeed, most of the functional models, such as the one presented in this chapter or (Liljencrants and Lindlom 1972) proposed to explain the universals of vowel inventories are clearly goal-directed. According to these models, the linguistic systems try to attain an optimal state, dubbed as the *goal*. A major drawback of these models is that they do not provide any insight on how the optimization process takes place in the real world, despite the absence of any explicit drive towards optimization either in the speakers or in the system as such. Nevertheless, in many cases (such as Boer (2001) in the context of vowel inventories) it can be shown that the optimal state is an emergent property of the linguistic system. In other words, even though the linguistic system or the speakers are not goal directed, the global optimization process emerges in the system as an outcome of simpler non-global processes. In the next chapter, we describe a model based on multi-agent simulation, which show that the schwa deletion pattern of Hindi can be explained as an emergent behavior of simpler interactions between speakers, obviating the need for a functional explanation. Nevertheless, as we shall see in the subsequent chapters, it is much easier to develop an efficient computational model and practically useful algorithms based on the functional explanation, but not for the emergent models.

Chapter 5

MAS Model I: Emergence of the Schwa Deletion Pattern

In the previous chapter, we have presented a constrained optimization model for SDH, which provides a functional explanation for the schwa deletion pattern of Hindi. We have also argued in the lines of evolutionary phonology to establish the fact that the syllable minimization-based account of SDH entails a diachronic explanation. Thus, to summarize our claim, *the observed pattern of SDH is an outcome of the process of language change over the past centuries, the process being governed by functional forces such as syllable economy and acoustic distinctiveness.* Nevertheless, this claim cannot be established beyond doubt unless one shows how the process of language change can lead to the emergence of the optimal pattern of SDH, despite the fact that there is no volitional attempt on the part of the language users or the linguistic system towards the optimization.

In this chapter, we describe a MAS-based model for SDH, where the speakers are modeled as *linguistic agents* capable of speaking, listening and learning languages. The agents interact with each other through linguistic communication and learn from their mistakes or successes. In the process, their language changes over time and several interesting structural property of the language emerges in the system. The cognitive processes of the agents and the simulation experiments are appropriately designed to capture the real world dynamics of language change as closely as possible; at the same time, in order to keep the system computationally tractable and analyzable, suitable abstractions and simplifications are made, wherever possible.

We observe that under specific circumstances, the emergent pattern of the MAS model closely resembles that of SDH. This leads us to believe that the proposed MAS-based model entails a plausible emergent explanation for the pattern of SDH and its evolution. We further argue that the explanation so entailed supports the tenets of *phonetically-based phonology* (Ohala 1974; Ohala 1987; Ohala 1989; Ohala 1993; Hayes et al. 2004), even though the final pattern that emerges can also be explained in terms of *functional phonology*.

The chapter is organized as follows. Sec. 5.1 presents the MAS framework by introducing the concept of *imitation games* and detailing out the design of the *linguistic agent*. Sec. 5.2 describes the various properties and parameters associated with the model, initialization conditions, and the simulation environment. The experiments and observations are reported in Sec. 5.3. The analysis of the results and their interpretations in the context of SDH are presented in Sec. 5.4. Sec. 5.5 summarizes and concludes the chapter.

As we shall see below, the MAS model presented in this chapter has certain limitations, for which it is not possible to carry out simulation experiments with a realistic Hindi lexicon. To overcome these limitations, we enhance this model in several ways in the next chapter. For this reason, we shall refer to the MAS model presented in this chapter as “MAS Model I” and the model presented in the next chapter as “MAS Model II”.

5.1 The MAS Framework

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

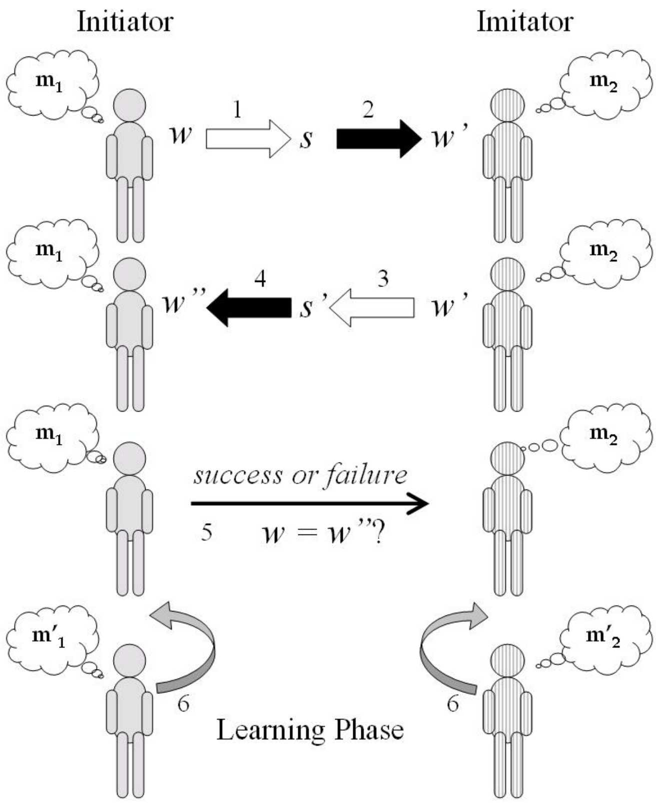


Figure 5.1: Schematic representation of an *imitation game*. The arrows represent events, which are numbered according to their occurrence and oriented according to the direction of information flow. The thick white and black arrows represent the process of articulation and perception respectively. The thin black arrow represents extra-linguistic communication and the gray arrows represent learning

emphasize however, that this is not a limitation and the framework can be extended to capture other phonological as well as syntactic phenomena.

5.1.1 Imitation Games

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 Fig. 5.1. 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* $\mathcal{A}$ (described in Sec. 5.1.4). The imitator perceives the signal and tries to map it to some valid linguistic unit w' in its language model using the *perceptor model* $\mathcal{P}$ (described in Sec. 5.1.5). 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 $\mathcal{P}$. 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. At the end of the game, both the agents *learn* on the basis of the result of the last language game as well as their past history.

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.

- Here we do not make any distinction between the *adult phase* when the rate of learning is very low, and the *learning phase* when the rate of learning is high of an agent.
- 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, here we assume that the noise in communication channel is modeled within the articulator model itself.
- The extra-linguistic communication of the success of a game (step 5 in Fig. 5.1) 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. In fact, it has been shown that successful communication can emerge in a system even without the assumption of extra-linguistic communication (Smith 2005).

5.1.2 Linguistic Agent Model

An agent (Russell and Norvig 2002) 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. Fig. 5.2 shows the block diagram of a linguistic agent. Formally, a *linguistic agent* **LA** is defined as a 4-tuple

$$\mathbf{LA} : \langle \mathbf{M}, \mathcal{A}, \mathcal{P}, \textsc{Learn} \rangle$$

where,

$$\begin{array}{rcl}\mathbf{M} & : & \text{Mental model of the agent (a set of mental states)} \\ \mathcal{A} & : & \text{The articulator model (or actuator)} \\ \mathcal{P} & : & \text{The perceptor model (or sensor)} \\ \textsc{Learn} & : & \mathbf{M} \rightarrow \mathbf{M} \text{ is the learning algorithm that maps a mental state} \\ & & \text{to another mental state}\end{array}$$

Ideally, in a MAS based model of language change, one would like to model $\mathcal{A}, \mathcal{P}, \mathbf{M}$ and Learn 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.

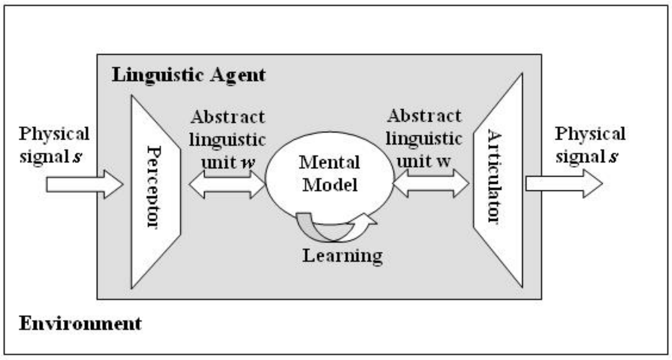


Figure 5.2: The architecture of a linguistic agent

5.1.3 Mental Model

The mental model $\mathbf{M}$ is defined as the set of all possible mental states of an agent. At any particular instant, one and only one of these mental states is associated with a agent. Let m_i and m_j be two distinct mental states belonging to $\mathbf{M}$. An agent can change its mental state from m_i to m_j through *learning*. Therefore, a learning mechanism LEARN can be conceived as a function from the set $\mathbf{M}$ to itself that maps a mental state m_i to another mental state m_j. LEARN 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. LEARN can be considered as a part of the mental model as well, but for the simplicity of presentation, we define LEARN separately. Here, we describe $\mathbf{M}$ for linguistic agents that share a common vocabulary, but possibly different surface forms (pronunciations) of the words.

Recall that the (finite) set of phonemes in a language is represented by Σ_P, a word $w \in \Sigma_P^*$, and the lexicon $\Lambda \subset \Sigma_P^*$ denotes the set of valid words in the language. We define the *realization of a word* w, represented by $r(w)$, as a string of 2-tuples $\langle p_i, t_i \rangle$, for $1 \leq i \leq n$, such that $w = p_1p_2 \ldots p_n$, (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 Λ, represented as $r(\Lambda)$, is obtained by replacing each element of Λ (say w), by a realization $r(w)$. A mental state m_i consists of a realization of Λ, which we shall represent as $r_i(\Lambda)$ or simply R_i. Note that by defining the realization as above, we have eliminated *intonation* and other *supra-segmental features* from the pronunciation of the words. Below, we illustrate this using a concrete example from Hindi.

Example 5.1 $\Sigma_P = \{a, A, i, k, m, r, h, \ldots\}$ is the set of all Hindi phonemes. Λ is the set of all Hindi words, but for the purpose of illustration let us define Λ as $\{hara, mara, amara, krama\}$, a set of four words. A specific realization $r(hara)$ of the word $hara$ looks like $\langle h, 1\rangle\langle a, 2\rangle\langle r, 1.5\rangle\langle a, 0\rangle$, 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 Λ, as given below (note that this realization is not representative of standard Hindi pronunciations).

$$\begin{array}{lll} r_i(hara) & \rightarrow & \langle h, 1\rangle\langle a, 2\rangle\langle r, 1.5\rangle\langle a, 0\rangle \\ r_i(mara) & \rightarrow & \langle m, 1\rangle\langle a, 2\rangle\langle r, 1.5\rangle\langle a, 2\rangle \\ r_i(amara) & \rightarrow & \langle a, 2\rangle\langle m, 1\rangle\langle a, 1.3\rangle\langle r, 1.5\rangle\langle a, 0.5\rangle \\ r_i(krama) & \rightarrow & \langle k, 0.5\rangle\langle r, 1.5\rangle\langle a, 1.6\rangle\langle m, 0.5\rangle\langle a, 0\rangle \end{array}$$

This particular state can also be represented as $r_i(\Lambda)$ or simply R_i. We can have another mental state m_j (also $r_j(\Lambda)$ or R_j) that looks like

$$\begin{array}{lll} r_j(hara) & \rightarrow & \langle h, 1\rangle\langle a, 2\rangle\langle r, 1.5\rangle\langle a, 0\rangle \\ r_j(mara) & \rightarrow & \langle m, 1\rangle\langle a, 2\rangle\langle r, 1.5\rangle\langle a, 0\rangle \\ r_j(amara) & \rightarrow & \langle a, 2\rangle\langle m, 1\rangle\langle a, 1.3\rangle\langle r, 1.5\rangle\langle a, 0.5\rangle \\ r_j(krama) & \rightarrow & \langle k, 0.5\rangle\langle r, 1.5\rangle\langle a, 1.6\rangle\langle m, 0.5\rangle\langle a, 0\rangle \end{array}$$

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 instance, for the Λ defined in Example 5.1, if the durations are restricted to take one of only three possible values, then there are 3^{19} possible mental states that comprise $\mathbf{M}$.

According to this definition of mental model, the agents share a common lexicon Λ. However, there exists variation in pronunciation among the agents, which is represented through the mental state an 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 disadvantage of non-generalization is that the agents have to learn the pronunciation of each of the words individually, which makes the simulation experiments time-consuming.

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 remember the number of games played previously and the number of games that 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 the outcome of each of the games individually.

5.1.4 Articulator Model

The articulator model $\mathcal{A}$ is a *procedure* that maps a linguistic unit $r_i(w)$ from the mental state R_i of an agent to a physical signal s. We define $\mathcal{A}$ 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 imperfections of the human articulators that result in errors during speech production. Moreover, as has been stated previously, this randomness can also be thought of to model the noise in the environment. For representational convenience, we will denote a signal s generated for a unit $r_i(w)$ as $\mathcal{A}(r_i(w))$ or simply $\mathcal{A}_i(w)$. Note that although $\mathcal{A}$ 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.

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 $w = p_1p_2 \ldots p_n$, $\mathcal{A}_i(w)$ is a string of tuples $\langle \rho_j, \tau_j \rangle$, where j varies from 1 to $2n-1$. Here, ρ_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. τ_j represents the duration of the corresponding phoneme or transition. We also make the assumption that the duration of a consonant is 0. Only vowels and phoneme-to-phoneme transitions have non-zero durations. The following example illustrates the representation of a signal.

Example 5.2 Let w be *amara* (refer to Example 5.1 above). If the agent is currently in the mental state m_i (i.e. R_i), then the corresponding realization of w (according to Example 5.1) is

$$r_i(amara) \rightarrow \langle a, 2 \rangle \langle m, 1 \rangle \langle a, 1.3 \rangle \langle r, 1.5 \rangle \langle a, 0.5 \rangle$$

The signal s corresponding to *amara* has 9 units (rho_is) represented as follows:

$$a,\ a-m,\ m,\ m-a,\ a,\ a-r,\ r,\ r-a,\ a$$

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.

$$\langle a, 1.8 \rangle \langle a-m, 0.9 \rangle \langle m, 0 \rangle \langle m-a, 0.65 \rangle \langle a, 1.3 \rangle \langle a-r, 0.65 \rangle \langle r, 0 \rangle \langle r-a, 0.15 \rangle \langle a, 0.3 \rangle$$

□□

The procedure $\mathcal{A}$ takes place as follows. Suppose the word to be articulated is $r_i(w) = \langle p_1, t_1 \rangle \langle p_2, t_2 \rangle \ldots \langle p_n, t_n \rangle$. Initially, the string

$$s = \langle \rho_1, \tau_1 \rangle \langle \rho_2, \tau_2 \rangle \ldots \langle \rho_{2n-1}, \tau_{2n-1} \rangle$$

is generated, where each ρ_j is defined as above and the τ_j are all 0. With respect to example 2, this corresponds to the string

$$\langle a,0\rangle\langle a-m,0\rangle\langle m,0\rangle\langle m-a,0\rangle\langle a,0\rangle\langle a-r,0\rangle\langle r,0\rangle\langle r-a,0\rangle\langle a,0\rangle$$

Next, for all odd js that represent phonemes, τ_j is assigned a value $t_{\lceil j/2\rceil} + \mathcal{A}_\epsilon(t_{\lceil j/2\rceil})$ if ρ_j is a vowel. The term $\mathcal{A}_\epsilon(t_{\lceil j/2\rceil})$ denotes a small random perturbation representative of 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 pr_d by a fixed amount δ from the duration stored in the current mental state of the *speaker*. The durations of other vowels are kept unchanged. Suppose δ is chosen to be 0.2 and pr_d is 0.4, then with probability 0.4 we reduce the duration of the schwas by 0.2. In the context of Example 5.2, 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.

$$\langle a,1.8\rangle\langle a-m,0\rangle\langle m,0\rangle\langle m-a,0\rangle\langle a,1.3\rangle\langle a-r,0\rangle\langle r,0\rangle\langle r-a,0\rangle\langle a,0.3\rangle$$

The duration of the transitions (i.e. τ_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:

$$\langle a,1.8\rangle\langle a-m,0.9\rangle\langle m,0\rangle\langle m-a,0.65\rangle\langle a,1.3\rangle\langle a-r,0.65\rangle\langle r,0\rangle\langle r-a,0.15\rangle\langle a,0.3\rangle$$

Assumptions and Justifications

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 (e.g., 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 quite a controversial concept (see Sec. 4.2). For a general overview on challenges and issues related to human perception, see (Jusczyk and Luce 2002) and references therein.

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 lead to deviation from the correct pronunciation (Ohala 1974; Ohala 1993). 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 for SDH.

Assumption 3: Duration of consonants is 0.
Justification: According to the sonority scale (Clements 1990), the vowels, which are placed at the top of the scale, can almost always form the syllable nucleus and have a longer duration. On the other hand, *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, in the context of the current work, due to the following reason it is a good abstraction to make. We want to model the deletion of the schwas. When a schwa gets deleted, the consonants that were part of the syllable have now to be placed within other syllables. For example, if both the schwas of the word $ha - ra$ 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 not pronounceable. 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.

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 δ (a model parameter) and a probability pr_d (also a model parameter). 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.

5.1.5 Perceptor Model

The perceptor model $\mathcal{P}$ maps a signal s to a word w in Λ. 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 R_i of an agent. Although these two actions might proceed hand in hand in human beings, separating them out simplifies $\mathcal{P}$. These two modules can be identified with the acoustic and the pronunciation models of automatic speech recognition systems (Jurafsky and Martin 2000, Chapter 7). Given a signal $s = \langle \rho_1, \tau_1 \rangle \langle \rho_2, \tau_2 \rangle \ldots \langle \rho_n, \tau_n \rangle$, the procedure $\mathcal{P}$ tries to perceive a phoneme p_i either from its realization $\langle \rho_{2i-1}, \tau_{2i-1} \rangle$ or from the transitions $\langle \rho_{2i-2}, \tau_{2i-2} \rangle$ or $\langle \rho_{2i-1}, \tau_{2i-1} \rangle$. 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.

$$Prob(p_j \text{ is correctly perceived from realization } \rho_j) = \tau_j/2 \tag{5.1}$$

Since the consonants are assigned a duration of 0, therefore, they can be perceived only from the transitions. We relax the probability for transition perception as follows, because transitions are assigned half the duration of the neighboring phonemes.

$$Prob(p_j \text{ is correctly perceived from transition } \rho_j) = \tau_j \tag{5.2}$$

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 the *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.

Example 5.3 Let the signal s corresponding to the word $w = amara$ be (taken from Example 5.2)

$$\langle a, 1.8 \rangle \langle a - m, 0.9 \rangle \langle m, 0 \rangle \langle m - a, 0.65 \rangle \langle a, 1.3 \rangle \langle a - r, 0.65 \rangle \langle r, 0 \rangle \langle r - a, 0.15 \rangle \langle a, 0.3 \rangle$$

Phoneme p	Left transition		Realization		Right transition		$Pr(p_i)$
	p_{lt}	$\neg p_{lt}$	p_{rz}	$\neg p_{rz}$	p_{rt}	$\neg p_{rt}$	
a	0	1	0.9	0.1	0.9	0.1	0.99
m	0.9	0.1	0	1	0.65	0.35	0.96
a	0.65	0.35	0.65	0.35	0.65	0.35	0.96
r	0.65	0.35	0	1	0.15	0.85	0.7
a	0.15	0.85	0.15	0.85	0	1	0.23

Table 5.1: Computation of perception probabilities. $Pr(p_i)$ denotes the probability of perceiving the ith phoneme and is computed using the expression $1 - (\neg p_{lt}) \times (\neg p_{rz}) \times (\neg p_{rt})$.

Let us estimate the probabilities of perception of the phonemes based on the above 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 5.1 illustrates the computations. We see that for the specific signal s presented in this example, the first three phonemes are almost always perceived (probability > 0.95), whereas the last one is hardly perceived (probability $= 0.23$). Also, the phoneme r is usually perceived (probability $= 0.7$). Therefore, the probability that v, the 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 this 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 Λ. The next task, therefore, is to map v to the nearest word w in Λ (the cognition step). This is accomplished by comparing v with realization of every word $r_i(w)$ in R_i. A score is calculated based on the *minimum edit distance* (Levenshtein 1966) of w and v, taking into account the duration of the vowels in $r_i(w)$ as well. The process is described below. If a vowel has a short duration in $r_i(w)$ and

$Cost(x, y)$	p	Schwa	ϕ
q	0 (if $p = q$), 2 (otherwise)	2	2
Schwa	2	0	2
ϕ	2	t	NA

Table 5.2: The cost matrix for aligning the phonemes x and y, where x (the column entries) belongs to the string w and y (the row entries) belong to v. ϕ is the null phoneme. p and q stand for any phonemes other than schwa. t represents the duration of the schwa according to the current mental state of the agent.

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 5.2. The case, where schwa in w is aligned with nothing in v (last row, second column), corresponds to schwa deletion, which is penalized by t the duration of the schwa according to the current mental state of listener. Note that the cost matrix is asymmetric and therefore, cost of aligning w to v is not the same as that of v to w.

The word $w^* \in \Lambda$ that has the lowest score (i.e. the word that is nearest to v) is then chosen as the output of the procedure $\mathcal{P}$. 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 5.4.

Example 5.4 Let us consider an agent, whose current mental state is R_i given in Example 5.1. Let the perceived string v be *amar* (refer to Example 5.3). The minimum edit distance of v from *amara* is calculated by finding out the best alignment, which is shown in Table 5.3. 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 5.4. The scores are also displayed when calculated according to the mental state R_j (Example 5.1.We observe that for both the mental states, the perceived string

w	a	m	a	r	$a(0.5)$
v	a	m	a	r	ϕ
Cost	0	0	0	0	0.5

Table 5.3: Minimum distance alignment

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

Table 5.4: Cost of alignment of *amar* with the different words, in mental states R_i and R_j

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 5.3, this has a probability of 0.005), the perceived word would have been *mara*. □□

Assumptions and Justifications

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 for SDH (Ohala 1983b).

5.1.6 Learning

The procedure LEARN defines transitions from one mental state to another on the basis of the outcome of the imitation games. As we have seen in Sec. 3.3.1, several approaches to learning have been explored for MAS models in the context of language evolution and change. For MAS Model I, we choose a very simple *trigger learning algorithm* (Gibson and Wexler 1994; Niyogi 2002). 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.

Suppose the initiator articulates a signal s corresponding to a word w, which is perceived by the imitator as w'. The imitator then articulates a signal s' corresponding to w', which is perceived by the initiator as w''. The procedure LEARN is as follows.

```
LEARN(k_sn, k_sm, k_f, pr_sn, pr_sm, pr_fn, pr_fm)
if w = w''
    then if the agent is the initiator
        then if success rate of the agent > k_sn
            then generate a random number r ∈ [0, 1]
            if r ≤ pr_sn then UPDATEDURATION(w, s)
        else if success rate of the agent > k_sm
            then generate a random number r ∈ [0, 1]
                if r ≤ pr_sm then UPDATEDURATION(w', s')
    else if success rate of the agent < k_f
        then if the agent is initiator
            then generate a random number r ∈ [0, 1]
                if r ≤ pr_fn then INCREASEDURATION(w)
            else generate a random number r ∈ [0, 1]
                if r ≤ pr_fm then INCREASEDURATION(w')
```

```
UPDATEDURATION(w, s)
w = p_1 p_2 ··· p_n,  s = {⟨ρ_j, τ_j⟩}_{2n−1}
for i = 1 to n
    do if p_i = /a/
        then t_i ← min(τ_{2i−1}, t_i)
```

Recall that t_i stands for the duration of p_i according to the realization of the word w in the mental model of the agent.

```
INCREASEDURATION(w)
w = p_1 p_2 ··· p_n
for i = 1 to n
    do if p_i = /a/
        then if t_i + δ > 2
            then t_i ← 2
            else t_i ← t_i + δ
```

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.

Note that the learners (i.e., the agents) here are *memoryless* in the sense that they do not remember the details of the previous games and the interaction history. Only the last imitation game and the overall communicative success are remembered by the agents.

5.2 The Simulation Set up

In this section, we describe the simulation setup for MAS Model I. We start with an outline of the simulation mechanism; this is followed by a discussion on issues related to lexicon selection and initialization of pronunciations. We also enlist a comprehensive list of the model parameters, which help us design the various simulation experiments. The details of the simulation environment and empirical results related to simulation time are also presented.

5.2.1 Simulation Mechanism

A population of N agents is initialized with identical mental states, say m_0 (the details of the initial mental states are discussed in the next subsection). The simulation is continued for several *rounds* with rn games per round. At the end of each round, the mental states of the agents are documented in a log file. Thus, the result file generated at the end of the simulation records the mental states of the agents only at the end of each round. Note that the mental states of every agent after every game can be recorded by setting $rn = 1$ and there is also provision for observing the words and signals generated/perceived by the agents during the games. Nevertheless, the huge number of games required for convergence makes it easier to analyze and visualize the results when the mental states are tapped only after a sufficiently large number of games. This is precisely the reason for introducing the concept of rounds, which is otherwise insignificant.

The steps in a language game are:

1. Two agents are selected randomly 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 Λ and generates a signal s corresponding to w using the articulator model $\mathcal{A}$,
3. The imitator receives the same signal s.
4. Imitator uses the perceptor model $\mathcal{P}$ to map the signal s to a valid word in Λ (if possible). Let the word perceived be w'.
5. Imitator generates a signal s' corresponding to w' using $\mathcal{A}$.
6. Initiator tries to perceive s' using $\mathcal{P}$. 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 initiator and 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.

5.2.2 Lexicon Selection and Initialization

Ideally, we would like to run the simulation experiments with the complete lexicon of Hindi. Since the lexicon of real languages are open systems, we could choose to work with *core lexicon* of the language defined as the set of most frequent M words. The value of M typically is between 10000 to 50000. However, as shown in Table 5.7, the time required to run simulation experiments with a lexicon of size 2000 is quite large and it is practically impossible to run meaningful simulations with lexicon of size 10000 for MAS Model I.

At the same time, the lexicon Λ has an important impact on the emerging pattern. This is due to the fact that perception is based on the closest word in Λ corresponding to the given string of phonemes. Table 5.5 shows two runs of a simple experiment with two agents and two words in the lexicon. With the exception of Λ, the values of the model parameters for the two runs are identical. We see that the emergent pronunciation of the word *amara* is different for the two runs.

Run	Words	Emergent pronunciation	Correct pronunciation
I	*hara*	*hra*	*har*
	amara	*mar*	*amar*
II	*mara*	*mar*	*mar*
	amara	*amar*	*amar*

Table 5.5: Two runs of the experiment with different lexica

The apparent discrepancies in the emergent pronunciations in Run I 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 Λ 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. In short, the structure of the lexicon defines the acoustic distinctiveness constraints and consequently the emergent SDH pattern.

In order to nullify the effect of the vocabulary the experiments we introduce the concept of normalized lexicon, where only two consonants and two vowels are used in many possible combinations to generate the words. One such lexicon is given below, which has been used as the Λ for most of the experiments with MAS Model I.

$$\Lambda_{normalized} = \{karaka, karakA, karAka, karAkA, kAraka, kArakA, kArAka\}$$

The motivation behind this definition comes from the observations made on the structure of the mental lexicon and the phonological neighborhood of words (Luce and Pisoni 1998; Vitevitch 2005; Kapatsinski 2006). The words of the real language lexica are found to be phonologically similar to each other making the structure of the network of phonological neighbors very dense and small. Note that in the normalized lexicon so defined, the average edit distance between a pair of words is 1.71 (assuming that cost of substitution is 1). Considering the fact that the minimum edit distance between two words cannot exceed 1, this value is quite small. Thus, the phonological neighborhood of a word in the normalized lexicon is very dense.

Apart from Λ, another important aspect of the model is the choice of initial pronunciations. Since we want to model the phenomenon of SDH, this choice is obvious, i.e., all the schwas have duration of 2 units. The initial states, thus, corresponds to the Sanskrit (or old-Hindi) pronunciations, where all the schwas are prominently pronounced.

5.2.3 Model Parameters

There are several parameters associated with MAS Model I that might have a significant effect on the emergent pattern. These parameters or *free variables* have already been defined and discussed in the previous section. Nevertheless, here we summarize them again in Table 5.6, which help us to understand the complexity and degrees of freedom of the model. Furthermore, this is important for the design of simulation experiments to systematically study the effect of these parameters on the emergent SDH pattern and their real life correlates.

Apart from the parameters listed in Table 5.6, an important issue governing a simulation experiment is the number of games for which the simulation is run, which is rn times the number of rounds. The values of all these parameters are specified in the beginning of a simulation experiment, and held fixed during that experiment.

5.2.4 Simulation Environment and Timing Analysis

All the simulation experiments reported in this chapter have been carried out on the Windows XP running on a Pentium-4 1.6GHz processor having 256MB RAM. The modules have been implemented in C.

Typically, it has been observed that the number of games required for convergence is linear in M – the size of the lexicon. This is because the agents learn the pronunciation of each of the words independently. Similarly, the number

Symbol	**Type**	**Range**	**Description**
N	integer	$[2, \infty]$	number of agents
δ	real	[0,2]	duration reduction step
pr_d	real	[0,1]	duration reduction probability
k_{sn}	real	$[0, 1]$	minimum initiator success rate for positive learning
k_{sm}	real	$[0, 1]$	minimum imitator success rate for positive learning
k_f	real	$[0, 1]$	maximum success rate for negative learning
pr_{sn}	real	$[0, 1]$	probability of positive learning by the initiator
pr_{sm}	real	$[0, 1]$	probability of positive learning by the imitator
pr_{fn}	real	$[0, 1]$	probability of negative learning by the initiator
pr_{fm}	real	$[0, 1]$	probability of negative learning by the imitator

Table 5.6: Model parameters for MAS Model I. Positive and negative learning implies learning from successes and failures respectively.

M	#Games	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	0.55M	680	12363	6.87

Table 5.7: 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

of games required for convergence is also proportionate to N – the number of agents, as every agent has to converge on a set of pronunciation independent of other agents. The time required for one game increases linearly with M, because, even though the articulation and learning phases take constant time, the perception step takes a time proportionate to M. This is due to the fact that the perceived string of phonemes v is compared with each of the words in Λ by computing the edit distance.

Therefore, the simulation time required for convergence is $O(M^2N)$. Table 5.7 shows the time required for a few different parameter settings. Assuming that convergence requires 10M games per word (see Sec. 5.3.1), 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)! Thus, even though the model (or a game) has an apparently manageable time complexity, it is not so in practice because of the large hidden constants. Stated differently, since the number of games required for convergence is very large even for small M and N (e.g., 8 and 2 respectively), we cannot afford to run experiments till convergence for sufficiently large M and/or N.

5.3 Experiments and Observations

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 pr_{sn} and pr_{sm} and 5) the deletion parameters pr_d and δ. 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 for different experiments.

5.3.1 Results for the Normalized Lexicon

The emergent pattern for the normalized lexicon for a specific run of the simulation has been presented in Table 5.8. Several runs of the simulation experiment under the same parameter settings show the emergence of very similar patterns, but the time required for convergence vary substantially between the runs.

Since the agents share the same lexicon $\Lambda_{normalized}$, 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.

There are two discrepancies in the emerging pattern, one is due to deletion of a schwa in a context where it should have been retained, and one is due to retention of a schwa, in a context where it is normally deleted. There are 12 schwas in the lexicon, and therefore, *the emergent 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

Words	Vowel duration	Emergent pronunciation	Pronunciation in std. 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

Table 5.8: Emergent pronunciations for the normalized lexicon. The parameters for this typical experiment were: $N = 4, k_{sn} = k_{sm} = 0.7, pr_{sn} = 0.6, pr_{sm} = 0.2, pr_{fn} = 0.6, pr_{fm} = 0.0, d = 0.01$. Games required for convergence: 70 million

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 ($karAkA \rightarrow^* krA - kA$). 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.

Fig. 5.3 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. We make the following observations and comments from Fig. 5.3.

- The transitions in the figure are very sharp (spanning over less than 10000 games) spaced by significantly longer periods of stable intermediate pronunciations.
- All the schwas that finally got deleted exhibit similar dynamics, which is represented by the so-called S-shaped curve.
- The fact that the MAS model I 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 Fig. 5.3. 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. Note that in this experiment the population size was $N = 4$.

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.

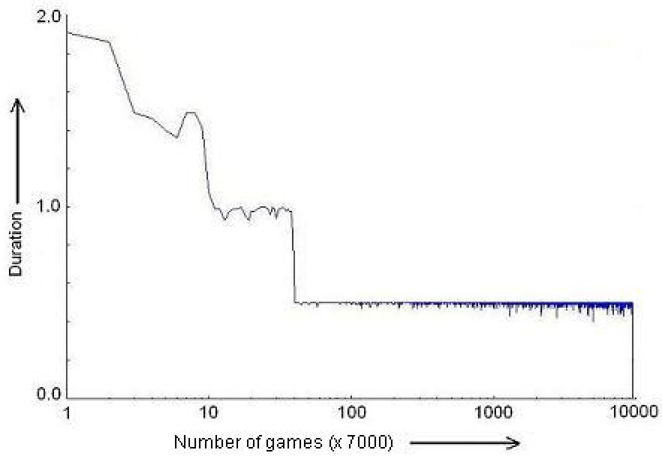


Figure 5.3: The average duration of a schwa vs. the number of games plotted for the final schwa of the word $karAka$.

5.3.2 Effects of other Parameters

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.

There are two parameters related to deletion – the duration reduction step δ and the duration reduction probability pr_d. The duration step parameter δ 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, δ and k_{sn} are the other two most influential parameters. Fig. 5.4 and 5.5 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 δ is negligible, but for smaller values of k_{sn} (less than or equal to 0.8), communicative success drops significantly for large δ. This can be explained as follows. When the agents greedily reduce the duration (large δ) 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 δ) 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 consistent global pattern.

The δ vs. average schwa duration curve (Fig. 5.5) 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 Fig. 5.4). However, when k is very close to 1, the system becomes too strict to allow schwa

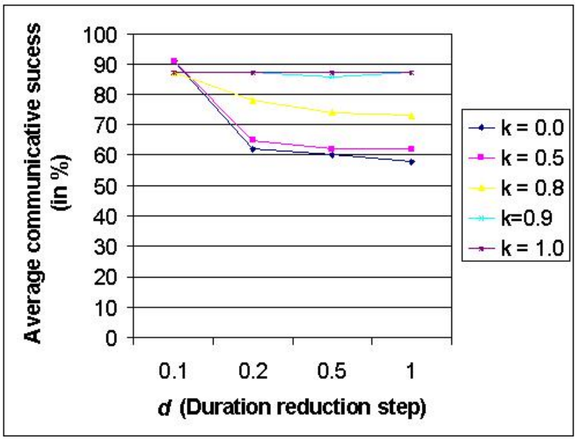


Figure 5.4: The dependence of average communication success rate on δ (duration reduction step) and learning threshold $k(=k_{sn}=k_{sm})$. Other simulation parameters: vocabulary size $=7$, $N=4$, $pr_{sn}=0.6$, $pr_{sm}=0.2$, $pr_{fn}=0.6$, $pr_{fm}=0.0$, number of games=300000.

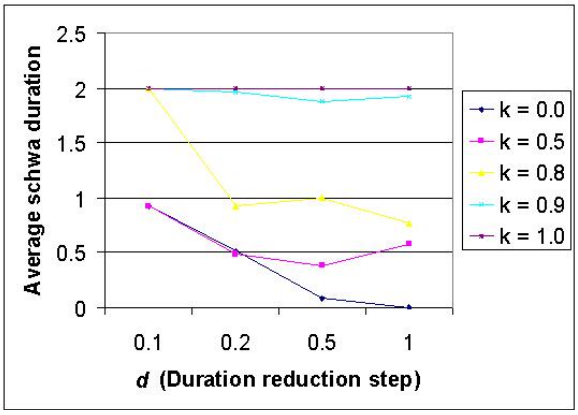


Figure 5.5: The dependence of average schwa duration on δ (duration reduction step) and learning threshold $k(=k_{sn}=k_{sm})$. Other simulation parameters: vocabulary size $=7$, $N=4$, $pr_{sn}=0.6$, $pr_{sm}=0.2$, $pr_{fn}=0.6$, $pr_{fm}=0.0$, number of games=300000. The expected duration according to Ohala's rule is 1.07

deletion and the original pronunciations are retained. Such a system has very high communicative success rate (as reflected in Fig. 5.4), 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.

5.3.3 Dialects and Synchronic Variation

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 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 Fig. 5.3 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 5.9 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 in Sec. 5.2.2, 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*.

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

Table 5.9: Different variants of a word that emerged during a simulation. The number of agents speaking that variant is given in the parentheses. Simulation parameters: vocabulary size = 7, $N = 10$, $d = 0.1$, $pr_{sn} = 0.6$, $pr_{sm} = 0.2$, $pr_{fn} = 0.6$, $pr_{fm} = 0.0$, $k_{sn} = k_{sm} = 0.9$, number of games=3M.

5.3.4 Robustness and Convergence Issues

What happens when under the same parameter settings we run two different simulations with different initial random seeds? Table 5.10 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 Fig. 5.3) 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.

We cannot, however, assume the schwas to be deleted in the same manner for every run. Observe that in Fig. 5.3 the last agent deleted the schwa much later than the other 3 agents (note that the x-axis is in logarithmic scale). In fact, theoretically it is impossible to predict the number of games after which

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

Table 5.10: The results for 10 different runs under the same parameter settings. Simulation parameters: vocabulary size = 7, $N = 4$, $d = 0.1$, $pr_{sn} = 0.6$, $pr_{sm} = 0.2$, $pr_{fn} = 0.6$, $pr_{fm} = 0.0$, $k_{sn} = k_{sm} = 0.9$, number of games=3M

a particular schwa will be deleted (see Bhat (2001) and references therein for a general discussion on predictability of sound change). 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 (Boersma 1997b; Boersma 1998; 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.

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.

5.4 Analysis and Interpretation

Let us analyze the reasons behind the structure of the emergent pattern of SDH in MAS Model I. We have already seen in the previous section that the specific pattern of SDH emerges only for certain range of the parameter values. Besides the model parameters, the agent model and the structure of the imitation game also play a major role in shaping the emergent behavior of MAS Model I.

5.4.1 Phonetic Factors Affecting the Emergent Pattern

The proponents of phonetically-based phonology (Ohala 1974; Ohala 1987; Ohala 1989; Ohala 1993; Hayes et al. 2004) argue that the course of sound change and other phonological changes are shaped by phonetic factors such as the common errors made while articulation and perception. Unlike functional phonology, phonetically based accounts of change do not advocate the optimality or communicative function of the linguistic structures as the driving force behind the change. We discuss below the properties of the articulatory and perceptual processes, which we think are primarily responsible for the emergent pattern of SDH as observed in MAS Model I. Consequently, we claim that MAS Model I entails a phonetically-based explanation for the SDH pattern of Hindi and its evolution over time.

Effect of the Articulatory Model

In MAS model I only one type of articulatory error, i.e. the reduction of schwa duration, has been modeled. This is because, other types of errors seemed 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. Nevertheless, the inherent tendency to reduce the schwa duration seems to be a necessary and sufficient 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.

Effect of the Perceptor Model

The perceptor model can explain the emergent pattern to a good extent. The three basic assumptions made while building 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.

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 5.8).

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 lexicon. The sensitivity of the emergent pronunciation to the lexicon is outcome of this assumption. Moreover, since the deletion of schwas can be less costly than that of other phonemes (due to the definition of the cost matrix in Table 5.2), the nearest word to a pattern, such as $kArak$ is the word $kAraka$ 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.

At the same time, we also mention that since in the beginning of the simulation, all the schwa durations are 2, the initial cost matrix is symmetric and therefore, unbiased. The matrix becomes asymmetric for an agent only after it has learnt to reduce the duration of a particular schwa. Therefore, in a sense, this bias is not preset in the system; rather it can also be viewed as an emergent behavior of the system. 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.

5.4.2 Analysis of other Model Parameters

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 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 hypothesize 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 (Fig. 5.5). This resembles the case of 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 (Sec. 5.3.1).

Other factors that are known to be significant for the schwa deletion pattern observed in a language include stress and morphology. The current work can be extended to model the stress pattern and the morphological features of a language and their effect on the emergent pattern can be studied.

We conclude our analysis of MAS Model I with one of the most interesting and counterintuitive observations: *the emergent pattern is sensitive to the lexicon.* This, as explained above, is clearly a consequence of the perception model. Nevertheless, the complete vocabulary of a language is expected to be optimally encoded with little redundancy such that the neighborhoods of words are dense (Luce and Pisoni 1998; Vitevitch 2005; Kapatsinski 2006). This in turn results in a situation where deletion of a consonant or a vowel makes correct recognition of a word less likely. At the same time, one must also consider the fact that in reality words are always uttered and recognized in context of other words and this provides extra clues for its perception. Moreover, a tendency towards phonological generalization, which is missing in the current model, is expected to even out any discrepancy arising due to the structure of the lexicon. Thus, we believe that, in general, the schwa deletion pattern observable in a language is independent of the lexicon.

5.5 Conclusion

In this chapter, we described a MAS based model for explaining the emergence of schwa deletion pattern in Hindi. Under suitable conditions, the pattern of SDH is found to evolve in the model with a significant S-shaped trajectory for duration reduction. These findings make MAS Model I a plausible explanation of the phonological change giving rise to the pattern of SDH. We also argue that, in effect, the model entails a phonetically based explanation of the phenomenon. Below we summarize some of the salient features of the model and important observations made from the experiments.

- A tendency towards reduction of schwa duration (i.e. fast speech) is pre-encoded in the model and seems to be a necessary as well as sufficient condition for emergence of schwa deletion.
- Under suitable conditions, the emergent pattern resembles the pattern of SDH with 83% similarity at the level of schwas and 71% similarity at the level of words.
- The pattern correct emerges only when neither the learning strategy nor the deletion rates are greedy. In other words, moderately stringent learning conditions and small deviations (articulatory errors) from the pronunciations are necessary preconditions for SDH.

- The articulatory and the perceptor models, and the lexicon have significant effects on the emergent pattern.
- The simulation experiments predict the coexistence of several variants of the words.
- 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.

MAS Model I supplements the constrained optimization model presented in the last chapter. While the optimization-based model provides an elegant and comprehensive view of the phenomenon that can be used to compute, in real time, the pattern of SDH, it cannot explain how the pattern might have emerged. MAS Model I, on the other hand, provides enough insight on how the pattern might have emerged, even though it is not possible to compute this pattern accurately or run the system in real time for the model.

The primary objective of this thesis is to develop computational models of “real world” phonological change. Although MAS Model I partially realizes this goal by putting forward a plausible model for SDH, the most serious limitation of the model towards realization of the objective of the thesis is in its inability to support meaningful experiments for a large and realistic lexicon of Hindi. In the next chapter, we extend and modify the current model in various ways and carry out experiments for real Hindi lexica.

Chapter 6

MAS Model II: Emergence of the Schwa Deletion Rule

In the previous chapter, we have seen that a tendency towards reduction of the duration of schwas coupled with appropriate learning mechanism gives rise to the schwa deletion pattern of Hindi. Nevertheless, the MAS Model I has the following limitations.

1. The articulatory as well as perceptual processes used in the model are costly in terms of the computations involved. We have seen that the simulation time required for convergence is very large (Table 5.7), making it impossible to carry out meaningful experiments with large and realistic lexica. The number of games required for convergence as well as the time required for each game is directly proportional to M, the size of the lexicon, which turns out to be rather costly as the constant of proportionality is very large.

2. The model assumes that the agents (read human beings) learn and store the pronunciation of each of the words independent of the pronunciation of any other words and without any generalization. This assumption is questionable as phonological rules are often generalized over the whole lexicon.

3. The duration of the vowels in the mental representation has been assumed to be a continuous variable between 0 and 1. However, past researchers have shown that phonological distinctions are often binary in nature. This fact is explicitly stated and argued for in the generative school (Chomsky and Halle 1968) and is implicit in several others including optimality theory (Prince and Smolensky 1993; Ellison 1994) .

In this chapter, we extend MAS Model I presented in chapter 5 in several ways to circumvent the aforementioned limitations. As we shall see, the modified model (henceforth referred to as MAS Model II) incorporates various abstractions, many of which are direct consequences of the experimental results of the previous model. These abstractions in turn help us to conduct experiments over a sufficiently large Hindi lexicon (20000 words), which grounds the model and the allied experiments more firmly in reality.

The chapter is organized as follows. Sec. 6.1 and 6.2 describes the modifications made in the mental and articulatory models, and the perceptory model respectively. The simulation setup including the initialization conditions, model parameters and analysis of simulation time, is presented in Sec. 6.3. The observations and findings of the simulation experiments are reported in Sec. 6.4, and the next section discusses the analysis and interpretations of the observations. Sec. 6.6 summarizes and concludes the chapter.

6.1 Representation and learning of Phonological Rules

In MAS Model I, the agents store and learn the pronunciation of the words individually, without any generalization. Furthermore, the vowel durations are stored as a continuous variable. However, we make the following important observations from the experiments with MAS Model I.

Observation: Upon convergence the duration of the schwas learnt by the agents are very close to either 0 or 2 (see section 5.3.1).

Implication: The agents essentially arrive at binary distinction, where a schwa is either fully pronounced (duration 2) or completely deleted (duration 0).

Thus, the first abstraction we introduce in MAS Model II is that in the pronunciations the schwas are either fully retained or deleted.

The second issue involving mental representation is generalization over individual pronunciations. Any generalization requires an appropriate bias and the choice of the bias is tricky, since it might render the emergent properties of the model trivial or impossible. Since other than the schwas, we treat all vowels and consonants alike, it is safe to assume that the pronunciations of the words that have the same consonant-vowel patterns are similar.

In this section, we describe the modifications made over the mental, articulatory and learning mechanisms based on the two aforementioned assumptions.

6.1.1 Mental Model

The words are clustered into groups in terms of their consonant vowel pattern or the *CV map* as defined below. Pronunciations are defined for each CV map in terms of probabilistic rules.

Definition 6.1 *For a word $w = p_1p_2 \dots p_n$, the CV map of w, represented by $CVMap(w)$ is a string of length n over the alphabet $\{C, V, a\}$, which is obtained by replacing every phoneme p_i in w by a corresponding C, V or a, depending on whether p_i is a consonant, a vowel (excluding schwa) or a schwa respectively. We shall use the symbols Ω and ω to denote CV maps.*

Note that the above definition of CV map differs slightly from the one presented in the context of syllable structure in Def. 4.8. Unlike the case of $CVMap(w)$, the schwa or /a/ is treated as a vowel, i.e. V, in $CVMap(\sigma)$. Nevertheless, conceptually the definitions are very similar and therefore, we refrain from using different notations for the two definitions. The concept of CV map is illustrated in Table 6.1. Note that several words can have the same CV Map.

Suppose a particular CV map Ω has k schwas. Based on whether a particular schwa in Ω is deleted or retained, there can be 2^k possible pronunciations associated with Ω. Let these pronunciations, which are also strings over $\{C, V, a\}$ be denoted by $\omega_0, \omega_1, \dots \omega_{2^k-1}$. Then a pronunciation rule for Ω is a set of 2^k

w	$CVMap(w)$	w	$CVMap(w)$
tuma	CVCa	kAma	CVCa
krama	CCaCa	karma	CaCCa
bAI	CVV	saradAra	CaCaCVCa

Table 6.1: CV maps for some common Hindi words.

2-tuples of the form $\langle \omega_i, pr_i \rangle$, where pr_i is the probability of pronouncing a word with CV map Ω as ω_i. Note that for every pronunciation rule, the probabilities pr_0 to pr_{2^k-1} must sum up to 1. The concept is illustrated through the following example.

Example 6.1 Let us consider the CV map $\Omega = CaCaCVCa$ of the word *saradAra*, which has $k = 3$ schwas. There are 8 possible pronunciations associated with Ω. A typical pronunciation rule associated with this CV map looks like

$$\begin{aligned} CaCaCVCa \rightarrow \quad & \langle CaCaCVCa, 0.10 \rangle \quad \langle CaCaCVC, 0.25 \rangle \\ & \langle CaCCVCa, 0.20 \rangle \quad \langle CCaCVCa, 0.05 \rangle \\ & \langle CaCCVC, 0.30 \rangle \quad \langle CCaCVC, 0.07 \rangle \\ & \langle CCCVCa, 0.02 \rangle \quad \langle CCCVC, 0.01 \rangle \end{aligned}$$

□□

We define the mental state of an agent as a tuple $\langle \Lambda, R \rangle$, where Λ is the lexicon comprising of M words and R is the set of pronunciation rules. In general, the number of pronunciation rules depends on the number of distinct CV maps for all the words in Λ. However, for a CV Map that does not contain any schwa (e.g. the CV map CVV of bAI in Table 6.1), there is just one trivial pronunciation with probability 1. Therefore, we can concieve of a default pronunciation rule r_d that maps a CV map to the same string with probability 1. Therefore, the set R consists of $\lambda + 1$ rules, where λ is the number of distinct CV maps that have at least one schwa and cover all the words in Λ.

According to this representation, all the agents share the lexicon Λ. Therefore, only the probabilities associated with the pronunciation rules vary across the mental states of the agents. The example below shows two possible mental states associated with a toy lexicon.

Example 6.2 Let us assume that $\Lambda = \{tuma, nAma, bAI, so, kala\}$. The distinct CV maps consisting of at least one schwa are $CVCa$ and $CaCa$. Therefore, two possible mental states m_i and m_j can be denoted by $\langle \Lambda, R_i \rangle$ and $\langle \Lambda, R_j \rangle$, where the pronunciation rules are defined as follows.

R_i	R_j
$CVCa \rightarrow \langle CVCa, 0.2 \rangle, \langle CVC, 0.8 \rangle$	$CVCa \rightarrow \langle CVCa, 0.5 \rangle, \langle CVC, 0.5 \rangle$
$CaCa \rightarrow \langle CaCa, 0.1 \rangle, \langle CaC, 0.7 \rangle,$	$CaCa \rightarrow \langle CaCa, 0.1 \rangle, \langle CaC, 0.2 \rangle,$
$\langle CCa, 0.2 \rangle, \langle CC, 0.0 \rangle,$	$\langle CCa, 0.5 \rangle, \langle CC, 0.2 \rangle,$
r_d	r_d

□□

Apart from the linguistic knowledge, the agents also possess a small short-term memory in which they can store the outcome of the recent imitation games played as well as the pronunciation patterns used during the last game played. The details of these are outlined in Sec. 6.1.3. Some of the salient features of the current mental representation are enumerated below.

1. Like **M** in MAS Model I, here too it is assumed that the agents share a common vocabulary.

2. Since the probabilities associated with the pronunciation rules can vary across the agents, there are infinite number of mental states, even though they have a fixed lexicon and a bounded number of pronunciation rules.

3. Learning in this representation implies change of probability distribution associated with the pronunciation rules. Thus, instead of learning M individual pronunciations as in the case Model I, here the agents need to learn $\lambda + 1$ probability distributions.

4. By definition, the scope of pronunciation variation is restricted to deletion of schwas only.

In the rest of this section, we discuss the modifications made on the articulator model and the learning mechanism that takes care of the new description of **M**.

6.1.2 Articulator Model

In MAS Model II, we choose to use the same representation for the acoustic signal as in MAS Model I (see Sec. 5.1.4). Given a word $w \in \Lambda$ and a mental state $m = \langle \Lambda, R \rangle$, the articulatory process $\mathcal{A}$ produces an acoustic signal s corresponding to w in the following manner.

1. Compute the CV-map of w. Let us denote the same by Ω.

2. If Ω does not contain any schwa, use the default rule $r_d \in R$; otherwise choose the appropriate pronunciation rule $r_\Omega \in R$, whose antecedent is Ω.

3. Out of the possible pronunciations $\omega_0, \omega_1, \ldots, \omega_{2^k-1}$, associated with Ω, stochastically one of the pronunciations, say ω^*, is chosen following the probability distribution $pr_0, pr_1, \ldots, pr_{2^k-1}$. If the rule is r_d, the pronunciation chosen is Ω.

4. The phonemes and transitions of the signal s are generated from w as described in Sec. 5.1.4. The durations of the vowels other than schwa are assigned as 2.

5. If a particular schwa in w (and consequently in Ω) surfaces in ω^*, then its duration is assigned as $2 - \delta$ with probability pr_d and 2 with probability $1 - pr_d$. We refer to pr_d as the *duration reduction probability* and δ as the *step of duration reduction.*

6. If a particular schwa in w (and consequently in Ω) does not surface in ω^*, then its duration is assigned as δ with probability pr_d and 0 with probability $1 - pr_d$.

7. The duration of transitions are computed from the duration of the phonemes as described in Sec. 5.1.4.

There are a couple of points about the articulatory model that require some elaboration. Firstly, the *step of duration reduction* δ must lie between 0 and 2. However, a more practical scenario would be one where $0 < \delta < 1$, as otherwise a schwa which is deleted according to the pronunciation rule chosen will surface as a long vowel and vice versa. Secondly, here we assume that the probability of duration reduction pr_d is same as that of duration increment. Although this is not necessarily true in reality, we make this simplification in order to reduce the number of free variables (parameters) associated with the model. The articulatory process is further explained below with the help of an example.

Example 6.3 Let us consider an agent with mental state m_i as described in Example 6.2 earlier. Suppose that the word chosen for articulation is $w = kala$. In order to find out the pronunciation rule applicable in this case, we construct the CV-map of w, which in this case is $\Omega = CaCa$. The rule associated with the antecedent Ω in R_i (refer to Example 6.2) is $\langle CaCa, 0.1\rangle$, $\langle CaC, 0.7\rangle$, $\langle CCa, 0.2\rangle$, $\langle CC, 0.0\rangle$. One of these pronunciations are chosen following the probability distribution. Let us assume that the pronunciation chosen is CaC (this has the highest probability 0.7).

The next step is identification of the schwas in w which are to be deleted according to the pronunciation pattern chosen. In this case, the absolute phoneme durations, therefore, are as follows.

$$\langle k, 0\rangle\langle a, 2\rangle\langle l, 0\rangle\langle a, 0\rangle$$

Further, we assume that $\delta = 0.5$ and $pr_d = 0.8$. Thus, with probability 0.8, the duration of the schwas are increased or decreased by 0.5. Suppose that after this process of random perturbation, the following durations are obtained for the phonemes.

$$\langle k, 0\rangle\langle a, 2\rangle\langle l, 0\rangle\langle a, 0.5\rangle$$

Finally, the duration of the transitions are computed from the phoneme durations to characterize the articulated signal s corresponding to w.

$$s = \langle k, 0\rangle\langle k \to a, 1\rangle\langle a, 2\rangle\langle a \to l, 1\rangle\langle l, 0\rangle\langle l \to a, 0.25\rangle\langle a, 0.5\rangle$$

□□

6.1.3 Learning

Learning in MAS Model II involves the change of the probabilities associated with the pronunciation patterns on the basis of the outcome of an imitation game. The learning strategy used here is similar to that of Model I, where every imitation game, whether a failure or a success, triggers the learning process. Recall that in MAS Model I (Sec. 5.1.6, the learning process is triggered with a particular probability (p_{sn} and p_{sm}) based on the success rate of the agent in the previous games (k_{sn} and k_{sm}). We observe that the parameters k_{sn} and k_{sm} have a significant effect on the emergent pattern, because in a sense they capture the tendency of the speakers to delete the schwas in spite of communication failure. Thus, for very low values of k_{sn} and k_{sm}, a coherent schwa deletion pattern fails to emerge and consequently the communication success rate of the system fails drastically. Clearly, this situation does not reflect the reality.

In Model II therefore, we introduce a single parameter, namely the *learning rate* η, which captures the tendency of the agents to learn a new pronunciation rule. We do not condition the learning on the communication success rate of an agent, because we assume that maintaining a high communication rate (as close to 1 as possible) is the goal of all the agents. Instead, the parameter η controls the communication success rate indirectly as follows. If η is very small, the agents learn very slowly maintaining a high communication rate in the system, but if η is large, it denotes that agents pay more importance to the outcome of recent past than the overall history. In other words, large value of η signifies that the agents are greedy. As we shall see shortly, unlike Model I, here η does not reflect any bias for schwa deletion; rather the tendency towards deletion is captured by the initial probability distributions associated with the pronunciation rules.

The learning proceeds as follows. Suppose in a particular game an agent (initiator or imitator) uses the pronunciation pattern ω_i associated with a CV-map Ω to articulate a word w for some i. Let the pronunciation rule for Ω according to the current mental state of the agent be as follows.

$$\Omega \rightarrow \langle \omega_0, pr_0 \rangle, \langle \omega_2, pr_2 \rangle, \ldots, \langle \omega_{2^k-1}, pr_{2^k-1} \rangle$$

If the game is successful, the probability pr_i of ω_i is increased according to the following rule (pr'_i is a temporary value).

$$pr'_i = pr_i + \eta pr_i \tag{6.1}$$

To maintain the sum of the probabilities at unity, the probabilities of the other rules are penalized equally as follows.

$$pr'_j = pr_j - \frac{\eta pr_i}{2^k - 1},\ i \neq j \tag{6.2}$$

However, if after the application of the above rule, we find that $pr'_j < 0$ for some j, then pr'_j is forcefully made 0. This might call for renormalization of the probability values. Thus, the final probability values after the learning process are obtained from normalization of the temporary values accordingly.

$$pr_i = \frac{pr'_i}{\sum_{t=0}^{t=2^k-1} pr'_t} \tag{6.3}$$

$$pr_j = \frac{pr'_j}{\sum_{t=0}^{t=2^k-1} pr'_t} \tag{6.4}$$

If the game is a failure, the probability of pr_i is decreased and the other probabilities are increased according to the following equations.

$$pr'_i = pr_i - \eta pr_i \tag{6.5}$$

$$pr'_j = pr_j + \frac{\eta pr_i}{2^k - 1},\ i \neq j \tag{6.6}$$

This is followed by the renormalization step, which is same as Eq. 6.3 and 6.4. The following example illustrates the learning process.

Example 6.4 Let us consider an agent whose current mental state is m_i as described in Example 6.2. Suppose that in the last game, the word articulated was *kala* and the rule used for articulation was

$$CaCa \rightarrow CaC, 0.7$$

Let us further assume that the last imitation game was successful and $\eta = 0.6$. Therefore, the probability of the pattern CaC is initially increased by 0.42 ($= 0.7 \times 0.6$) and the other values are decreased by 0.14 ($= 0.42/3$). This step (we shall call it *step 1*) is followed by *step 2*: assignment of negative probabilities to 0, and *step 3*: renormalization. The initial and final probabilities values for each of the above steps are shown below.

Pronunciation pattern	Initial value	After step 1	After step 2	After step 3
$CaCa$	0.10	-0.04	0.00	0.00
CaC	0.70	1.12	1.12	0.95
CCa	0.20	0.06	0.06	0.05
CC	0.00	-0.14	0.00	0.00

Thus, at the end of the learning process, the agent is in a mental state, say m'_i, where the pronunciation rule R'_i is identical to R_i except for the rule whose antecedent is $CaCa$. The revised probability values of this particular rule are those in the last column of the above table. □□

6.1.4 Implementation issues

In order to facilitate fast execution as well as lower memory usage, several optimizations and tricks have been used while implementing the aforementioned processes. The most important ones are described below.

- Since Λ is shared by all the agents, we store Λ as global variable, which can be read by all the agents. Each word $w \in \Lambda$ is assigned a unique id (number), which is nothing but its array index.

- The pronunciation rules are stored as a part of the agent, because they can vary across the agents. For a particular rule, it suffices to store the list of indices of schwas in the antecedent CV-map. Associated with the antecedent with k schwas, there is an array of 2^k elements, which store the probability values. The index itself is indicative of the pronunciation pattern, because we assume that in the binary representation of the index, a 0 means the schwa is deleted and 1 means it is not. Thus, index 0 means all schwas are deleted in the pronunciation, whereas index 2 (10 in binary), the last schwa is deleted but the second last is not.

- With every word $w \in \Lambda$ we also store the index of the rule that is applicable based on its CV-map. Thus, we do not need to compute the CV-map of w during the articulation process.

6.2 Perceptor Model

The task of the perceptor model $\mathcal{P}$ is to map a signal s to a word w in Λ. Recall that $\mathcal{P}$ involves two distinct phases - *perception* and *cognition* (Sec. 5.1.5). The perception step described in MAS Model I takes time linear in the length of the signal (number of phonemes and transitions) and produces a string of perceived phonemes v. Nevertheless, the cognition step, where we try to find out the word $w \in \Lambda$ that is nearest to v, takes a time linear in the number of words in Λ. This takes a significant amount of time for simulation when Λ is large, and consequently prevents us from experimenting with large lexicon in MAS Model I.

In MAS Model II, the perception phase of $\mathcal{P}$ is identical to that of MAS Model I. However, we introduce major modifications in the cognition phase. Firstly, since the structure of the mental model in Model II is significantly different from that of Model I, we introduce a new notion of *nearness* between the words based on the noisy channel approach (introduced in Shannon (1948), but also see Kernighan et al. (1990), Brown et al. (1993) and Brill and Moore (2000) for examples of applications in language modeling and processing). We design an efficient search algorithm for the nearest word using the concept of *Universal Levenshtein Automaton* (Mihov and Schulz 2004). The time complexity of the algorithm is linear in the size of v - the sequence of perceived phonemes, but *almost independent* of the size of the lexicon.

6.2.1 Defining the Nearest Word

We can formulate the cognition process in terms of a noisy channel as follows. Let w be a word in Λ, which has been articulated by some agent. The process of articulation, transmission and the perception steps can be abstracted out in

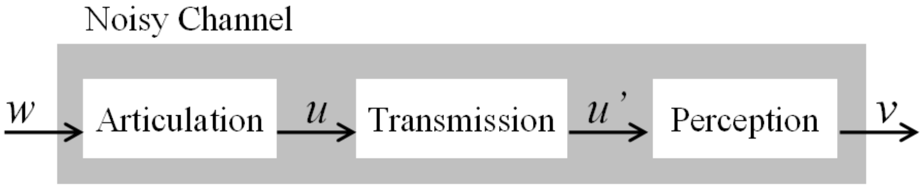


Figure 6.1: The noisy channel view of the articulation, transmission and perception processes. In the models described in this thesis, the transmission phase is assumed to be noise free. Therefore, u and u' are identical.

form of a noisy channel, so that after being transmitted through this channel, w is converted to as v. This process is illustrated in Fig. 6.1.

The noisy channel is characterized by the conditional probability $Pr(v|w)$. We can estimate the probability $Pr(w|v)$ from $Pr(v|w)$ using Bayes rule as shown below.

$$Pr(w|v) = \frac{Pr(v|w)Pr(w)}{Pr(v)} \tag{6.7}$$

The word nearest to v can be defined as $w^* \in \Lambda$ for which $Pr(w|v)$ is maximum. Thus,

$$w^* = \underset{w\in\Lambda}{argmax}\ Pr(v|w)Pr(w) \tag{6.8}$$

Note that we can neglect the term $Pr(v)$ in the denominator, as it is identical for all the words in Λ and as a result does not have any effect on Eq. 6.8. Throughout the experiments in MAS Model II, we shall assume that the probability of articulating any word in Λ is same, that is $1/|\Lambda|$. Under this assumption, the above equation can be rewritten as

$$w^* = \underset{w\in\Lambda}{argmax}\ Pr(v|w) \tag{6.9}$$

This assumption has been made in order to guarantee that the emergent pattern of SDH is independent of the distribution of the words in the language.

Nevertheless, an interesting direction of research can be to study the effect of the distribution of the words, if any, on phonological change, which we deem as a future work.

In order to compute the probability $Pr(v|w)$, we make the following assumptions:

Assumption 1: The transmission stage, as shown in Fig. 6.1, has no noise and consequently no visible effect on the value of $Pr(v|w)$. This fact is justified from the point of simulation as we do not add any noise during the transmission phase. Nevertheless, the perception phase is a stochastic process and can add to the distortion.

Assumption 2: The agents follow the *Saussurean convention* (Hurford 1989; Oliphant 1996) that states that the same model (i.e., grammar rules) are used for encoding and decoding the linguistic signals. In this case, it translates to the fact that the perceiver agent uses its own mental state to compute the probability $Pr(v|w)$ rather than assuming a different mental state or some statistical distribution over the mental states.

Therefore, we can express the probability $Pr(v|w)$ as a product of two independent probabilities $Pr_{\mathcal{P}}(v|u)$ and $Pr_{\mathcal{A}}(u|w)$, where the former represents the perception step and the latter the articulation step. u is the intermediate string of phonemes representative of the pronunciation pattern used for articulation. Thus we have

$$Pr(v|w) = \sum_{u \in \Sigma^*} Pr_{\mathcal{P}}(v|u) Pr_{\mathcal{A}}(u|w) \tag{6.10}$$

Suppose that according to the mental state of the perceiver agent, the word w is uttered using the rule

$$\Omega \to \langle \omega_0, pr_0 \rangle, \langle \omega_1, pr_1 \rangle, \ldots, \langle \omega_{2^k-1}, pr_{2^k-1} \rangle$$

Also, let u_i be the string of phonemes obtained when the pronunciation pattern ω_i is used to articulate w. Then, the probability that the articulated string was of the form u_i is given by pr_i. Note that the value of $Pr_{\mathcal{A}}(u|w)$ is greater than

0 only if u is a possible articulated pattern according to the rule Ω. Therefore, we can rewrite the above equation as

$$Pr(v|w) = \sum_{i=0}^{2^k-1} pr_i Pr_{\mathcal{P}}(v|u_i) \tag{6.11}$$

Estimation of $Pr_{\mathcal{P}}(v|u)$ is tricky, as there are several, possibly infinite, number of ways through which the string u could be perceived as v, and one needs to find the sum of the probabilities of all these possible paths to compute the value of $Pr_{\mathcal{P}}(v|u)$. However, since we are interested in finding out the word w^* for which $Pr(v|w)$ is maximum (Eq. 6.9), an exact estimate of $Pr_{\mathcal{P}}(v|u)$ is not necessary; rather it suffices to have a good approximation of the quantity, so that the values of $Pr(v|w)$ are comparable across the words in Λ. As it turns out, the quantity $\epsilon^{ed(u,v)}$, where ϵ is a positive constant less than 1 and $ed(u,v)$ is the edit distance between u and v, is a good estimate of $Pr_{\mathcal{P}}(v|u)$ (see, e.g., Bailey and Hahn (2001) for a similarity metric definition and Ellison and Kirby (2006) for an application of the same). Intuitively, this can be justified as follows. v can be obtained from u by making $ed(u,v)$ errors (insertion, substitution or deletion). If probability of making an error is ϵ, which is much less than 1, and errors are made independent of each other, then the quantity $\epsilon^{ed(u,v)}$ approximates the probability of making $ed(u,v)$ errors. Consequently, the quantity provides a rough estimate of $Pr_{\mathcal{P}}(v|u)$.

Thus, we can write

$$Pr(v|w) \approx \sum_{i=0}^{2^k-1} pr_i \epsilon^{ed(u_i,v)} \tag{6.12}$$

The value of ϵ depends on the noise introduced during the articulation and perception processes.

Note that equations 6.9 and 6.12 together imply that the cognition step requires the computation of the edit distance of the perceived string v from all words in Λ. This would make $\mathcal{P}$ linear in $|\Lambda|$, which is clearly undesirable. Moreover, the estimation of $Pr(v|w)$ as described by Eq. 6.12 is computationally intensive, and ideally we would like to carry out this computation for only a very small subset of Λ. Thus, if it is possible to define a set $\Lambda_v \subseteq \Lambda$, such that

$w^* \in \Lambda_v$, then we need to search for w^* only in Λ_v. Furthermore, if $|\Lambda_v| \ll |\Lambda|$, the cost of computation is significantly reduced, provided that there is a fast construction procedure for Λ_v.

The problem at hand is closely analogous to that of spell-checking (see Kukich (1992) for definition and review of different approaches to spell-checking), where v can be identified with the misspelling and w^* the correct or intended spelling. Spell-checking is often modeled as a noisy channel process (Kernighan et al. 1990; Brill and Moore 2000), and given a large lexicon Λ and the misspelling v, there are techniques to define and construct the set Λ_v. Here, we use a technique based on *similarity-key* (Odell and Russell 1918; Odell and Russell 1922; Kukich 1992; Zobel and Dart 1995) to define the set Λ_v, and construct the set efficiently using the concept of *Universal Levenshtein Automaton* (Mihov and Schulz 2004). The following subsections present the algorithm for the cognition step, outlining the problem-specific modifications made to the aforementioned concepts, but refrain from a general description of the techniques as such, for which the reader might refer to the relevant citations.

6.2.2 Defining Λ_v

The concept of *similarity-key* was introduced in the Soundex algorithm (Odell and Russell 1918; Odell and Russell 1922), where phonetically similar words get mapped to the same string, known as the *key*. Since the scope of the current work is limited to SDH, we do not consider a general mapping function that takes into account the confusability between every possible pairs of phonemes. Rather, we observe that a word w and a string v might be highly similar if one can be obtained from the other by only deletion and/or insertion of schwas. Stated differently, since only the schwas can be deleted during pronunciation, the perceived string v corresponding to an articulated word w is expected to share all phonemes with w, except for may be the schwas. This motivates us to define the *key* function as follows.

Definition 6.2 *Let w be a string of phonemes. We define a mapping key from Σ_P^* to $\{\Sigma_P - a\}^*$, such that in key(w) all the 'a' in w are mapped to null, keeping the other phonemes unchanged.*

w	$tuma$	$nAma$	bAI	so	$kala$
$key(w)$	tum	nAm	bAI	so	kl

Table 6.2: The *similarity-keys* for the words in the lexicon shown in Example 6.2

Table 6.2 shows the *keys* for all the words in the lexicon described in Example 6.2. Note that two different words, such as *karma* and *krama*, can have the same *key*, i.e. *krm*. This definition of the similarity-key enables us to write the following relations (u_i is as defined in eq. 6.11).

$$ed(v, w) \geq ed(key(v), u_i) \geq ed(key(v), key(w)) \tag{6.13}$$

A possible definition of Λ_v can be the set of all words $w \in \Lambda$, such that $key(w) = key(v)$ or equivalently $ed(key(w), key(v)) = 0$. However, a little inspection shows that we can construct practical examples, where $w^* \notin \Lambda_v$. For instance, suppose $v =$ "$karm$", so that $key(v) =$ "krm". Further suppose that Λ has words $w_1 =$ "$karma$" and $w_2 =$ "$karo$". Let us also assume that the pronunciation rules according to the current mental state of the agent are

$$CaCCa \rightarrow \langle CaCCa, 0.8\rangle, \langle CaCC, 0.0\rangle, \langle CCCa, 0.2\rangle, \langle CCC, 0.0\rangle$$

$$CaCV \rightarrow \langle CaCV, 1.0\rangle, \langle CCV, 0.0\rangle$$

Then, $Pr(v|w_1) = 0.2\epsilon(\epsilon + 4)$ and $Pr(v|w_2) = \epsilon$. Since, $\epsilon < 1$, we have $Pr(v|w_1) < Pr(v|w_2)$. However, by the definition of Λ_v, w_1 is in Λ_v, and w_2 is not. Thus, in order to ensure that $w^* \in \Lambda_v$, the definition of Λ_v must be less stringent. For Hindi, we have observed that if Λ is sufficiently large ($>$ 5000 words), then usually $ed(key(v), key(w^*)) \leq 1$. This motivates us to define Λ_v as follows.

Definition 6.3 *For a given string of phonemes v and a lexicon Λ, Λ_v is the set of all words $w \in \Lambda$, such that $ed(key(v), key(w)) \leq 1$.*

Although one can still construct examples with the above definition of Λ_v, where $w^* \notin \Lambda$, such examples seldom occur in practical cases. In other words, Def. 6.3 provides a heuristic description of the set Λ_v, which works quite well in practice.

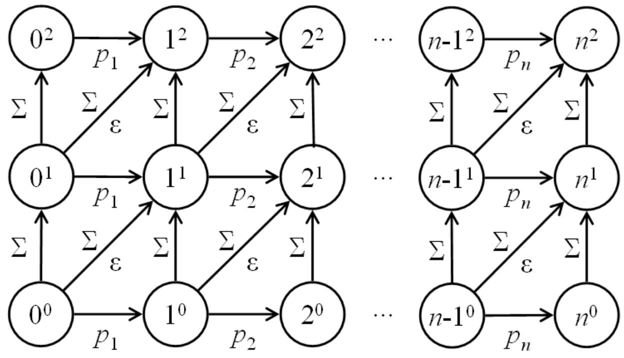


Figure 6.2: The structure of the Levenshtein automaton of order 2 for a string $s = p_1p_2 \ldots p_n$.

6.2.3 Levenshtein Automaton

In order to find out the set of all words $w \in \Lambda$, such that edit-distance between w and v is 0 or 1, we construct a *Levenshtein automaton* (Ukkonen 1985; Wu and Manber 1992; Baeza-Yates and Navarro 1999; Mihov and Schulz 2004) for v. By definition, the Levenshtein automaton of order k for a string v is a non-deterministic finite state automaton, which accepts a string s if and only if $ed(v, s) \leq k$. We describe the structure of the Levenshtein automaton with the help of Fig. 6.2.

Let us try to construct the Levenshtein automaton of order k for the string $s = p_1p_2 \ldots p_n$. The automaton can be visualized as a $(k+1)\times(n+1)$ rectangular grid of states. The row indices run from 0 to k and the column indices run from 0 to n. Let us denote the state in the i^{th} row and j^{th} column as j^i (refer to Fig. 6.2)[1]. The 0^0 is the *start-state* of the automaton. There are three types of transitions in the automaton.

- Horizontal or left to right transitions from a state j^i to $(j+1)^i$ are conditioned on the character p_{j+1}. These edges, when traversed, denote a match of character between the input string and s.

[1]Note that here superscript i is not an exponent. This notation has been adopted from (Mihov and Schulz 2004)

- Vertical or upward transitions from a state j^i to j^{i+1} are conditioned on any character (denoted by Σ in Fig. 6.2). These edges denote the insertion of a character.
- Diagonal transitions from a state j^i to $(j+1)^{i+1}$ are conditioned on any character as well as ϵ - the null transition marker. When the diagonal transition is traversed for a character, it denotes a substitution; whereas when the ϵ-transition is traversed, it denotes a deletion of character.

Thus, every move upward from a state in row i to row $i+1$ increases the edit distance by 1. The row index of the states, therefore, stores the value of the current edit-distance between the substring of the input seen from s. All the states in the $(n+1)^{th}$ column are the *final states*. Since the automaton is a non-deterministic one, after traversing an input string x, the automaton can be in more than one states; if any of these states is a final state then $ed(s,x) \leq k$. The exact edit-distance between s and x is given by the row index of the lowest (the one with smallest superscript) state.

Fig. 6.3 shows the Levenshtein automaton of order 1 for the string "*tuma*" and traces the state transitions of the automaton, when the input string is "*tum*". The states that are activated after traversing a part of the input are shown in gray. After scanning the input string "*tum*", the automaton can be in one of the four states - $3^0, 2^1, 3^1$ and 4^1, out of which 4^1 is a final state. Thus, the edit-distance between "*tum*" and "*tuma*" is 1.

Note that in Fig. 6.3, the set of activated states at any point of time, except for in the beginning, forms a triangular pattern, which shifts horizontally as the input is scanned. The set of activated states after scanning the j^{th} character of the input are $j^0, (j-1)^1, j^1$ and $(j+1)^1$. On the basis of similar observations, it is possible to define an automaton, where each state represents a subset of states in the Levenshtein automaton and the transitions are conditioned on string of phonemes of length $k+1$. Here k is the order of the Levenshtein automaton. A transition from state S to a state S' on phonemes $p_1p_2\ldots p_k$ in this automaton denotes that if the set of activated states of the Levenshtein automaton is represented by S and the last k phonemes observed including the current one are p_1 to p_k, then the next set of activated states in the Levenshtein

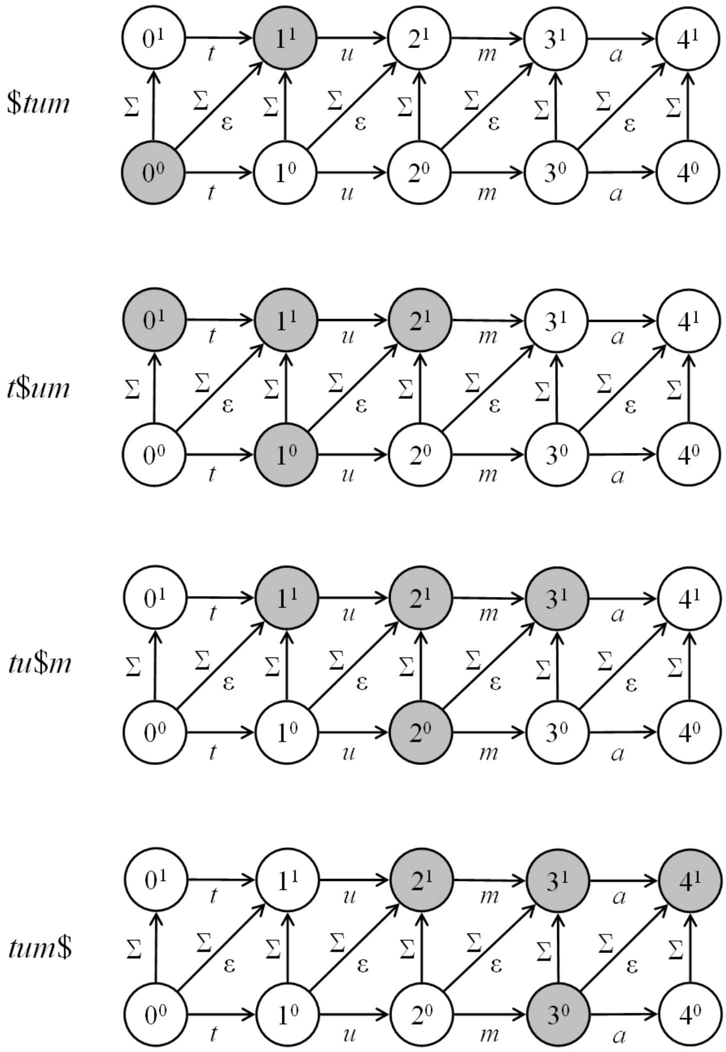


Figure 6.3: Tracing the state transitions for the string "*tum*" on the Levenshtein automaton of order 1 for "*tuma*". The currently activated states are shown in grey. $ separates the part of the input that has been scanned from the part to be scanned.

automaton is given by S'. This automaton that captures the traversal of a general Levenshtein automaton of order k is called an *Universal Levenshtein Automaton* (ULA) of order k (Mihov and Schulz 2004).

Thus, for a given string v, we can build a ULA of order 1, whose transitions are defined based on the phonemes in v. The lexicon Λ can also be conceived as an automaton. We can then construct the set Λ_v through a simultaneous traversal of the ULA and the lexicon. In the following subsection, we describe the data structures and algorithm that realize the aforementioned concept. Since we are interested in the ULA of order 1 only, here we refrain from any discussion on the general construction procedure of a ULA of order k; rather we describe a simpler algorithm for simultaneous traversal of the Levenshtein automaton of order 1. Nevertheless the basic idea is borrowed from the concept of ULA presented in (Mihov and Schulz 2004).

6.2.4 Algorithm for Searching w^*

The search for w^* consists of the following three distinct phases.

1. **Preprocessing:** The lexicon Λ is represented as an automaton. Since all the agents share the lexicon, a common automaton is constructed before the simulation begins. The automaton is accessible to all the agents during the simulation.

2. **Construction of Λ_v:** Λ_v is constructed in the runtime for the perceived string of phonemes - v through a traversal of the lexicon automaton.

3. **Exhaustive search for w^*:** For every word $w \in \Lambda_v$, the value of $Pr(w|v)$ is computed using Eq. 6.12 and the word for which $Pr(w|v)$ is highest is returned as w^*.

Each of these phases is detailed out in the following subsections.

Preprocessing

Recall that Λ_v is the set of all words in Λ for which the edit distance between the keys of v and the word are no greater than 1. The search for the words in

Λ_v, therefore, is performed in the domain of the key values of the words, rather than the words themselves. We construct a lexical automaton, denoted by $M_\Lambda : \langle \Sigma_P - a, Q, q_0, F, \delta_M \rangle$, for the lexicon Λ, where Q is the set of states, $q_0 \in Q$ is the start state, $F \subset Q$ is the set of final states and $\delta_M : Q \times \{\Sigma_P - a\} \rightarrow Q$ is the transition function. Given a string $s \in \{\Sigma_P - a\}^*$, the automaton M_Λ accepts s if and only if there exists a word $w \in \Lambda$, such that $key(w) = s$.

M_Λ has the following properties.

1. For every state $q \in Q$, there is a unique sequence of transitions from q_0 to q.
2. The sequence of transitions from q_0 to q defines a string of phonemes, say s_q.
3. $q \in F$ if and only if there exists a word $w \in \Lambda$, such that $key(w) = s_q$.
4. For every final state $q \in F$, we also store a list of all the words $w_q^i \in \Lambda$, such that $key(w_q^i) = s_q$ (i varying from 1 to the number of such words in Λ).

Note that M_Λ is essentially a *forward trie* (de la Briandais 1959; Fredkin 1960), that is used for efficient word search in Information Retrieval and other NLP applications. The node q_0 is the *root* of the trie. The construction of M_Λ is simple and we use the commonly used algorithm described in (Knuth 1997) for this purpose. The following example illustrates the structure of the lexicon automaton.

Example 6.5 Let us assume that the lexicon $\Lambda = \{sara, asara, so, saradAra, krama, karma\}$. The corresonding keys for the words are $sr, sr, so, srdAr, krm, tum$ and krm respectively. Fig. 6.4 shows the structure of M_Λ that accepts the keys corresonding to all the words in Λ. The final states of the automaton are shaded and the list of the words associated with a final state are also indicated in the diagram. □□

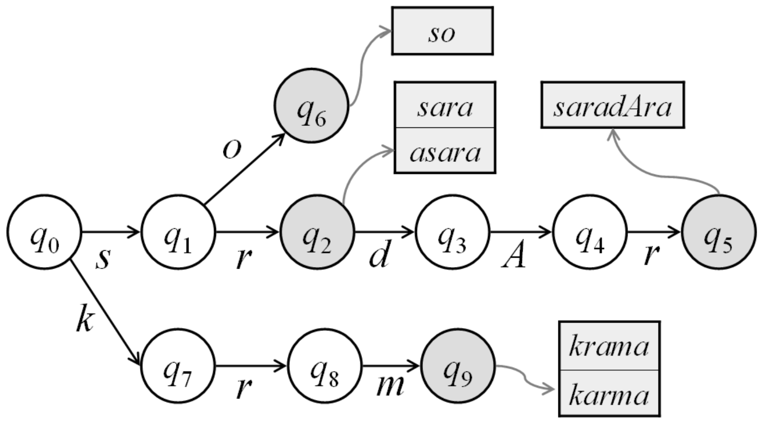


Figure 6.4: The lexicon automaton M_Λ for the lexicon shown in example 6.5

Construction of Λ_v

Given a string v and the lexicon automaton M_Λ, we construct the set Λ_v by simultaneous traversal of M_Λ and the Levenshtein automaton of order one for $key(v)$. Starting from the node q_0, we do a depth first traversal of M_Λ. Before presenting the algorithm, we introduce a few notations that will be used while describing the algorithm.

- We denote the transition function of the Levenshtein automaton of order 1 for $key(v)$ by δ_v and use the notation $\delta_v(j^i, p)$ to denote the set of states that can be reached from the state j^i, when the input is the phoneme p. Note that since the Levenshtein automaton is non-deterministic, $\delta_v(j^i, p)$ represents a set of states, which can also be null.
- We denote the transition function of M_Λ by δ_M and use the notation $\delta_M(q, p)$ to denote the state reached on input phoneme p, when the machine is in state q. Unlike the Levenshtein automaton, M_Λ is deterministic and therefore, $\delta_\Lambda(q, p)$ is unique, unless no transitions from q are defined on p, for which we assume that $\delta_\Lambda(q, p)$ is some predefined *trap state*, say q_{TS}.
- The list of words associated with a state q of M_Λ is denoted by $wlist(q)$. Note that if $q \notin F$, then $wlist(q)$ is empty.

The algorithm consists of an initialization step – Init, followed by a call to the recursive function – Traverse.

```
Init(M_Λ, v)
  v′ ← key(v)
  M_v ← Levenshtein automaton of order 1 for v′
  Λ_v ← NULL
  la_states ← {0^0, 1^1}
  Traverse(q_0, la_states)
```

The variable *la_states* contain the list of active states of the Levenshtein automaton M_v. Recall that in the beginning, that is before scanning any input, the states 0^0 and 1^1 are active (refer to Fig. 6.3). Also, we assume that the automata M_Λ and M_v, as well as the set Λ_v are declared *globally*. The function Traverse is as shown below.

```
Traverse(q, S)
  if q = q_TS or S = NULL
      return
  for each p ∈ {Σ_P − a}
      la_states ← NULL
      for each s ∈ S
          la_states ← la_states ∪ δ_v(s, p)
      Traverse(δ_Λ(q, p), la_states)
  for each s ∈ S
      if s is a final state of M_v
          Λ_v ← Λ_v ∪ wlist(q)
          return
  return
```

It can be shown that the number of active states in M_v for any input string never exceeds 4. In other words, the size of the variable *la_states* is bounded. Moreover, it is not necessary to explicitly construct the automaton M_v in order to formulate the transition function δ_v. The row and column indices of a state can be readily computed from the current row and column indices of the current state, say j^i, by comparing the j^{th} phoneme of the string $key(v)$ with the input phoneme p.

Proof of Correctness

In order to prove that the set of words Λ_v constructed by the above algorithm conforms to the definition 6.3 we first prove the following lemma.

Lemma. For every call of the function TRAVERSE(q, S), S contains the set of active states in M_v for the string s_q (the unique string of phonemes defined by the sequence of transitions from q_0 to q).

Proof. This can be proved by induction on q as follows. The basis is when $q = q_0$, for which this is trivially true (see step 4 of procedure INIT). For the induction step, notice that in the steps 4, 5 and 6 of procedure TRAVERSE, we construct the set of active states of M_v, when p is the observed input and the machine M_v is in one of the states in S. Thus, if on the input string s_q the set of active states of M_v is S (true by the induction hypothesis), then for the input string $s_q \cdot p$, the active state in M_Λ is $\delta_\Lambda(q, p)$ (by definition) and the set of active states in M_v is stored in the variable *la_states* after the completion of the **for**-loop in step 4 (by construction). This completes the proof.

Step 10 in the procedure TRAVERSE is the only place where Λ_v is updated. By the above lemma, for the input string s_q, in M_Λ we reach the state q, which implies that $\forall w \in wlist(q), s_q = key(w)$; and in M_v, we reach the set of states S, which contains at least one final state to ensure that Λ_v is updated (condition in step 9). If S contains a final state, then $ed(s_q, key(v)) \leq 1)$ (by definition of the Levenshtein automaton). Thus, we have

$$\forall w \in wlist(q), ed(key(w), key(v)) \leq 1$$

Hence, all the words in Λ_v satisfy the condition stated in Def. 6.3.

Similarly, if there is a word w satisfying the above condition, then by definition of M_Λ, there must be state $q \in Q$, such that $s_q = key(w)$. Since the procedure TRAVERSE explores all the transitions from a given state (step 3), the state q must be eventually traversed. Again by the above lemma, when TRAVERSE is called on q, the set of active states S at that point must contain at least one final state in M_v, because $ed(s_q, key(v)) \leq 1$. Therefore, all the words in Λ whose keys are within one edit-distance of $key(v)$ are included in

Λ_v. Thus, Λ_v constructed by the algorithm described above indeed conforms to the definition given in Def. 6.3.

Exhaustive Search for w^*

After the construction of the set Λ_v, we compute the value of $Pr(v|w)$ for each word $w \in \Lambda_v$ using Eq. 6.12. Note that even though we could compute the edit-distance between $key(w)$ and $key(v)$ while constructing Λ_v, it is not very clear how one can obtain the edit distance between v and w from the above. Therefore, we compute the exact edit-distance between the strings v and w during calculation of $Pr(v|w)$ and find out the word w^*, for which $Pr(v|w)$ is maximum. If there are more than one words with the maximum probability value, one of them is chosen at random as w^*.

6.3 The Simulation Setup

The structure of the agent for MAS Model II is presented in sections 6.1 and 6.2; the interaction between the agents takes place through *imitation games*, which has been discussed in section 5.1.1. In this section we describe the simulation setup for MAS Model II. The initialization of the mental model of the agents, which includes selection of an appropriate Λ and choice of initial pronunciation rules, is one of the most important issue in MAS Model II, and the same is discussed below. We also present a comprehensive list of the model parameters for MAS Model II, which help us design the various simulation experiments. To complete the description of the simulation setup, we also present the details of the simulation environment and some empirical results for the simulation time.

6.3.1 Lexicon Selection

The prime motivation behind the construction of MAS Model II lies in the fact that though we observed that the emergent pattern for SDH is dependent on the lexicon, we could not run the simulation experiments on a real lexicon for Model I. Consequently, one of the most important issues in Model II is the selection of an appropriate lexicon that is representative of the real Hindi lexicon. As is clear from the agent model described above and also substantiated in the following section through empirical analysis of the model, the time required for one imitation game as well as the number of games required for convergence in MAS Model II are *almost* independent of the lexicon size. Nevertheless, we must define a bound on the size of Λ as the vocabulary of the real languages are never *closed* and it makes no sense to work with the "entire" lexicon of Hindi.

In order to choose an appropriate value of M – the size of Λ, we proceed as follows. From a large corpus of Hindi, we enumerate the unigram (i.e. occurrence) frequencies of the words and sort the words in descending order of the frequency. We define Λ to consist of the top M words from this sorted list. Experimenting with different values of M (described in Sec. 6.4.4), we observe that other parameters being fixed, the emergent pattern of SDH changes as M increases from a very small value (say 1000) to 10000. However, as we further increase the lexicon size there is hardly any change in the emergent pattern. Therefore, for all of the experiments, excepting those meant for studying the dependency of the emergent pattern on M, Λ **is defined as the set of 10000 most frequent Hindi words.**

In this context, a few points are worth noting. Firstly, here the unigram frequencies have been estimated from the CIIL corpus[2], which is a 3 million word written corpus of standard Hindi. However, for the kind of experiments described here, ideally one should use a historical spoken corpus of Hindi (or some ancestor language of Hindi), which reflects the word usage patterns when there was no schwa deletion. Unavailability of such data forces us to resort to this alternative approach. Nevertheless, as discussed in Sec. 2.3, we assume that the current orthography of Hindi reflects the spoken forms of the past. Another important issue is that of inflectional and derivational morphology. Here, we consider the different inflected and derived forms of a root as distinct words.

Several features associated with Λ (such as the number of distinct CV-maps denoted by λ, words without any schwa (WWS) and states in M_Λ) grow with M.

[2]Developed by Central Institute of Indian Languages, Mysore, India. Currently the corpus is a part of a bigger set of language resources collectively known as EMILLE/CIIL corpus, developed and distributed under the EMILLE project (Xiao et al. 2004)

We describe below the behavior of these features as a function of M, because the scalability of MAS Model II heavily depends on them. As a matter of fact, for $M = 10000$, the values of λ, the number of WWS and the number of states in M_Λ are 352, 2682 and 16347 respectively.

Distinct CV-maps in Λ

Fig. 6.5 shows the variation of λ – the number of CV-maps with M. We observe that when plotted on a doubly logarithmic scale, the growth pattern of λ can be very well approximated by a straight line (indicated by the red line in Fig. 6.5). Through regression analysis, we obtain the following relationship between M and λ (on doubly logarithmic scales, the standard error for λ is 0.0235 and F-statistic is 293.7).

$$\lambda = 10.315M^{0.387} \tag{6.14}$$

Clearly, λ grows quite slowly with M. This sublinear growth supports the scalability of MAS Model II, because, since the agents have to learn λ probability distributions for the λ rules, the number of games required for convergence is directly proportional to λ and not M.

Words Without any Schwa

Fig. 6.6 shows the variation of the number of WWS with M. The growth pattern, in this case, cannot be approximated by any simple mathematical function. We note an exponential rise in the number of WWS initially, which grows at a much slower rate as M exceeds 10000. This is evident from Fig. 6.6, where the growth curve is approximately linear on the sub-logarithmic scale till $M = 10000$, after which it becomes nearly parallel to the x-axis. Although this parameter has no direct relevance to our experiments, there is one point, however, which calls for immediate attention. Recall that the emergent pattern in the simulation experiments are found to be heavily dependent on M, till M is 10000, but for $M > 10000$, we hardly observe any effect of M over the SDH pattern. It is quite plausible that the two observations have some causal connection, but further investigations are necessary before we can make any firm conclusions regarding the same.

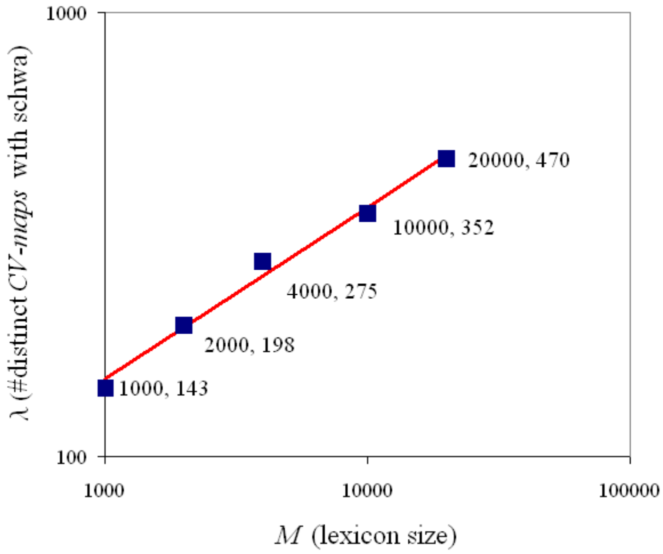


Figure 6.5: The growth pattern of the number of CV-maps versus lexicon size M for Hindi. The blue points corresond to emirical data, whereas the red line is the fit according to Eq. 6.14.

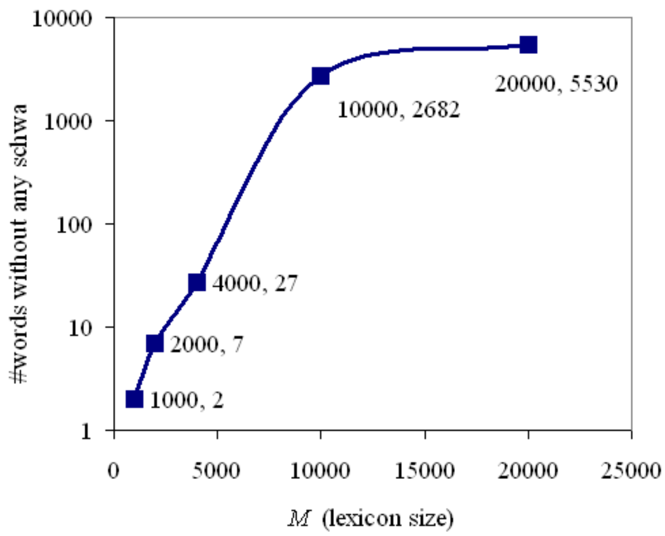


Figure 6.6: The growth pattern of the number of words without any schwa versus the lexicon size M for Hindi.

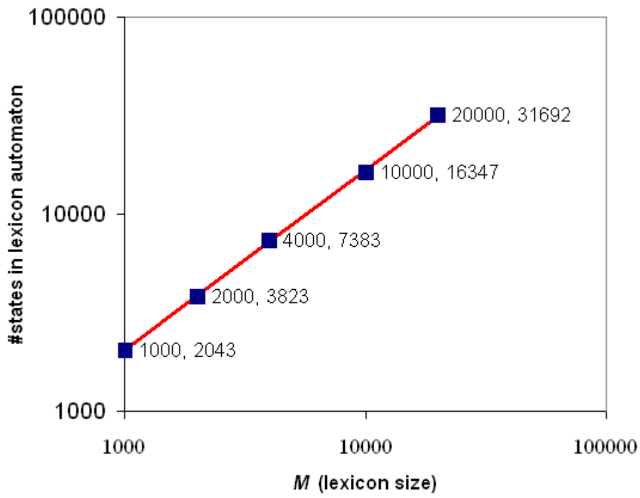


Figure 6.7: The growth pattern of the number of states in M_Λ versus the lexicon size M for Hindi. The blue points correspond to empirical data, whereas the red line is the fit according to eq. 6.15.

Number of States in M_Λ

The number of states in the lexicon automaton M_Λ grows almost linearly with M. Fig. 6.7 plots the growth pattern on a doubly-logarithmic scale. The line of regression, shown in red in the figure, represents the following function.

$$\text{The number of states in } M_\Lambda = 3.754M^{0.912} \tag{6.15}$$

The fit explains the data very well with a standard error less than 0.008 and F-statistic greater than 15200. The exponent 0.912 shows that the growth is nearly linear. The number of states in M_Λ is an indirect estimator of the memory and time required for the simulation. However, as we shall see shortly, although there is a strong correlation between the simulation time and the number of states, the former grows much more slowly than the latter.

6.3.2 Rule Initialization

The initialization of the mental models of the agent in MAS Model II involves appropriate instantiation of the pronunciation rules. Since the lexicon is shared by the agents, the number of pronunciation rules is also fixed and common among the agents. Rule instantiation in this case, therefore, boils down to assignment of the probability values for the different pronunciation patterns.

In MAS Model I, the pronunciation rules are deterministic in nature and the variation in pronunciation emerges due to a tendency towards reduction of the duration of schwas during articulatory process. As a result, the initial pronunciation rules in the model are shared by the agents, where all the schwas are pronounced. To the contrary, in MAS Model II the articulation process does not have any inherent bias towards deletion, even though it can add some noise to the generated signal. Therefore, the variation of pronunciation as well as the tendency towards vowel reduction must be captured through the probability values assigned to the different pronunciation patterns. This makes the choice of rule initialization far from trivial.

Let Ω be a particular CV-map with k schwas, whose associated pronunciation patterns are ω_0 to ω_{2^k-1} (see Sec. 6.1.1). There can be several strategies for initialization of the probability values pr_0 to pr_{2^k-1}. We list a few representative strategies below and also discuss their psycholinguistic implications.

Random probability assignment: This is a simple strategy where the values of pr_i are assigned randomly for every agent. Linguistically, this implies that the language-users do not share a common pronunciation grammar to start with. The tendency towards schwa deletion varies across the users as well as across rules for the same user. Although this model can provide some insightful results that can be used as a baseline for other models, its linguistic implications are rather counterintuitive. Moreover, the random assignment strategy does not allow us to systematically study the emergent behavior of the model with respect to the bias towards schwa deletion.

Uniform probability assignment: According to the uniform probability assignment strategy, all the probability values pr_i are assigned the same value that is 2^{-k}. Linguistically, this implies that the language-users agree on the pronunciation model[3] and there is no bias towards either retention or deletion

[3]Note that this does not mean that the speakers would agree on the pronunciation of a word in a particular game.

of the schwas. The strategy, though linguistically plausible, does not allow us to study the effect of the bias towards deletion, and thus can serve as a good baseline model.

Fixed schwa deletion probability based assignment: In this strategy, we define a parameter π, that is the probability of deletion of a schwa. We assume that every schwa is deleted with a probability of π (and consequently retained with a probability of $1 - \pi$) independent of the status of the other schwas. Let $|\omega_i|_a$ be the number of schwas in the pronunciation pattern ω_i. Then, the probability values according to this model are assigned as follows.

$$pr_i = \pi^{|\omega_i|_a}(1-\pi)^{k-|\omega_i|_a} \quad (6.16)$$

Note that when $\pi = 0.5$, the initial probability values assigned by this strategy is identical to those by the uniform probability assignment stratey. The psycholinguistic implications of the model are: 1) the language-users agree on the pronunciation grammar, and 2) initially, the schwas are deleted with a fixed probability independent of the pattern. The deletion pattern emerges due to the successes and failures of the users encountered during the interactions with other users.

Here, we choose the fixed schwa deletion probability based rule initialization strategy, not because of the fact that it seems linguistically more plausible than the other two, but for the single reason that this strategy allows us to control and study the effect of the bias towards schwa deletion using a single parameter π. Fig. 6.8 depicts the nature of the probability distributions obtained for pronunciation rules with different CV-maps for different values of π.

6.3.3 Model Parameters

Like MAS Model I, there are several parameters associated with Model II that might have a significant effect on the emergent pattern. These parameters or *free variables* have already been defined and discussed in the previous sections. Nevertheless, here we summarize them again in Table 6.3, which help us to understand the complexity and degrees of freedom of the model. Furthermore, this is important for the design of simulation experiments to systematically

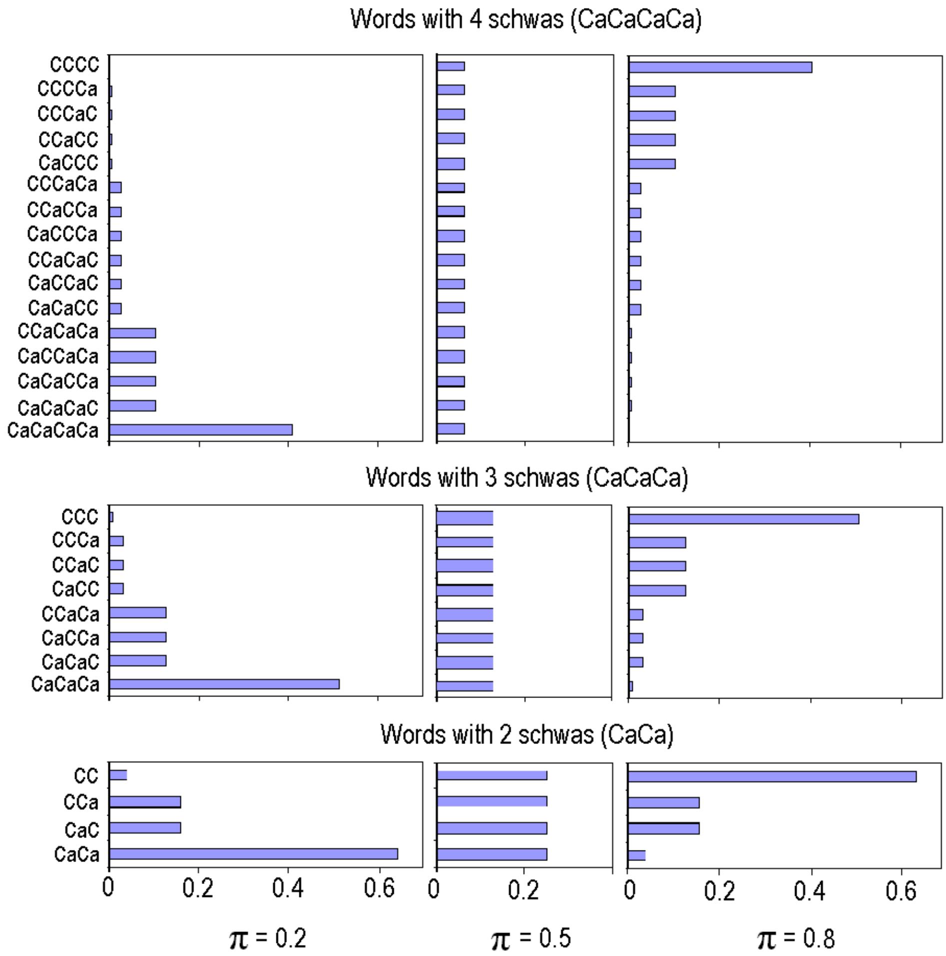


Figure 6.8: Initial probability distribution of the pronunciation rules for a few different CV maps for different values of π. The bar charts in the top, middle and bottom rows are for CV maps with 4, 3 and 2 schwas respectively. The bar charts in the left, middle and right columns are for $\pi = 0.2, 0.5$ and 0.8 respectively. The CV maps on the left shows the pronunciation patterns (ω_is) and the assigned probabilities (pr_is) are shown by the horizontal bars. Note that for $\pi = 0.5$, the assigned probabilities are same for all the patterns associated with a CV map.

Symbol	**Type**	**Range**	**Description**
N	integer	$[2, \infty]$	number of agents
M	integer	$[2, \infty]$	number of words in the lexicon
δ	real	[0,2]	duration reduction step
pr_d	real	[0,1]	duration reduction probability
η	real	$[0, \infty]$	learning rate
π	real	[0,1]	schwa deletion probability while initialization

Table 6.3: Parameters of MAS Model II

study the effect of these parameters on the emergent SDH pattern and their real life correlates.

Apart from the parameters listed in Table 6.3, two other important issues governing a simulation experiment are: 1) the number of games for which the simulation is run, and 2) initialization strategy. The effect of the parameters N, pr_d and δ has been studied in details for MAS Model I, and we observe that it is the number of games required for convergence, rather than the emergent pattern, that is dependent on N. Unlike Model I, where pr_d and δ encode the tendency and style of deletion for the agents, in Model II these parameters have been introduced to incorporate some noise in the articulatory model. Therefore, in our experiments with MAS Model II, we primarily focus on the effect of the parameters M, η and π on the emergent pattern.

6.3.4 Simulation Environment and Time Analysis

All the simulation experiments, reported in this chapter, have been carried out on the Windows XP running on a Pentium-4 1.6GHz processor having 256MB RAM. The modules have been implemented in C. In order to estimate the time required for simulation a set of two simulation experiments with 50000 and 100000 games has been carried out for five different values of M. For a particular value of M, the time required for 50000 games has been computed

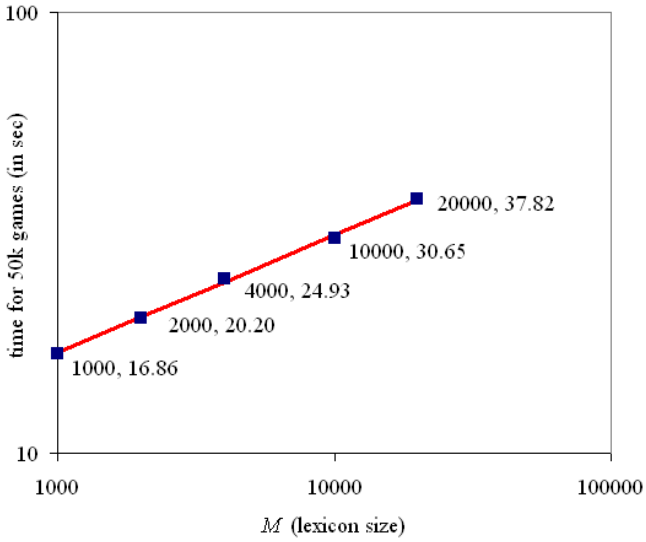


Figure 6.9: The plot of time required for 50000 games (in seconds) versus the lexicon size M for Hindi. The blue points correspond to empirical data, whereas the red line is the fit according to eq. 6.17.

by averaging over the observations of the two simulation runs. The results are plotted in Fig. 6.9. Note that by "time", here we refer to the *real time* that elapses between the beginning of the execution of the simulation code and the end of the same. Therefore, apart from the system configuration, the reported values of the run-time also depend on the system-load, that is the number of other processes and applications running simultaneously during the simulation experiments. It has been ensured that all the experiments are run under similar system-load conditions so that the time estimates are directly comparable.

From Fig. 6.9, we observe that on a doubly logarithmic scale, simulation time depends linearly on M. The empirical data can be very well approximated by the following equation (the standard error and the F-statistic for the fit are 0.007 and 1538 respectively).

$$\text{time per game} = 53.37M^{0.267}\,\mu sec \tag{6.17}$$

Thus, the time required for simulation grows much slowler than the size of the lexicon. We have seen that for MAS Model I, the time required for one million games for a vocabulary size of 1800 is 12363 seconds. According to Eq. 6.17, the time required for one million games when $M = 1800$ in MAS Model II is 397.3 seconds. This is a reduction of 96.8% over the time required in Model I. However, when $M = 8$, the time required for Model II is less than that of Model I by only 3.8%. Thus, Model II is far better than the Model I in terms of scalability with respect to the size of the lexicon.

This drastic reduction of the simulation time for Model II is an outcome of the modified perception algorithm; this is because the modifications made to the other modules (e.g. articulatory process and learning) are too insignificant to have some visible effect on simulation time. Therefore, it is worthwhile to analyze the time-complexity of the perception algorithm. The time required during the perception process is the sum of the time required for construction of Λ_v and searching for w^* in Λ_v. The former can be estimated by counting the number of states in M_Λ that are explored during the execution of TRAVERSE, while the latter is a linear function of the size of Λ_v.

Table 6.3.4 reports the average number of states explored during TRAVERSE for different lexicon sizes. The values have been obtained by counting the number of calls to TRAVERSE for each invocation of the perceptual procedure over 1000 games with randomly selected words from the lexicon. Since each game involves two invocations of the perceptual process, we have the data for 2000 perceptual experiments. We observe that unlike the number of states in M_Λ, which grows almost linearly with M, the number of states explored during the construction of Λ_v grows sublinearly with M. This is also reflected in the fraction of the total number of states explored, which significantly decreases with M. Fig. 6.10 shows the distribution of the number of states explored for the aforementioned experiments. It is evident from the histogram that the number of states explored is normally distributed and the mean and the standard deviation of the distribution increases with M.

In Table 6.3.4, we also report the average size of the set Λ_v as returned by TRAVERSE during the same set of experiments. The size of Λ_v is almost independent of M, and varies from 2 to 3. However, unlike the distribution of the number of traversed states, the distribution of the mean size of Λ_v is heavily skewed. For instance, when $M = 20000$, in 37% of the experiments, $|\Lambda_v| = 1$,

Size of the lexicon - M	Number of states in M_Λ	average number of states explored	Percentage of states explored	average size of Λ_v
4000	7383	829.2	11.23	1.95
10000	16347	1125.7	6.89	2.77
20000	31692	1450.7	4.56	3.04

Table 6.4: The number and the fraction (expressed in %) of states explored in M_Λ during TRAVERSE and the average size of the set Λ_v for different M.

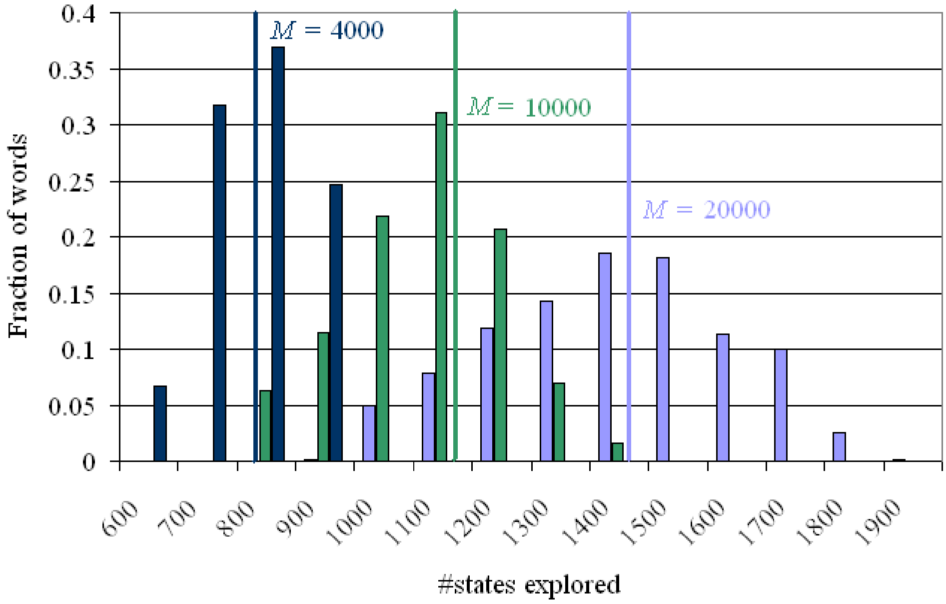


Figure 6.10: Distribution of the number of states explored for different values of M. The blue, green and cyan bars represent the distributions for $M = 4000$, 10000 and 20000 respectively. The thin vertical lines denote the averages for the distributions following the same color codes.

and in more than 88% of the cases, $|\Lambda_v| \leq 5$; nonetheless the maximum value observed for $|\Lambda_v|$ is 34.

These observations pertaining to the perceptual process can satisfactorily explain the timing behavior of MAS Model II as represented by Eq. 6.17.

6.4 Experiments and Observations

In this section, we systematically explore the effect of the different model parameters on the emergent schwa deletion pattern. The experiments with MAS Model I reveal that the parameter N has no effect as such on the emergent behavior, but it only affects the convergence time. The same is arguably true for MAS Model II, and therefore, here we do not conduct any further experiments to study the effect of N and for all other experiments, the value of N is fixed at 4.

The effect of the parameters δ and pr_d has also been studied in details for Model I. We have observed that realistic deletion patterns emerge for small values of δ and pr_d. However, the lower the values of these parameters, the higher is the time required for convergence. Since in Model II, these parameters are representative of the articulatory noise[4], we do not devote any elaborate experimentation and discussion regarding the effects of these parameters. In all the experiments, δ and pr_d are set to 0.1 and 0.6 respectively.

In the following subsections we focus on the effect of the parameters M, initialization strategy and π,and η on the emergent pattern. To begin with, we define certain evaluation metrics and present the results of some baseline experiments that will allow us to compare and comprehend the results of the simulation experiments that follow.

Note that unlike in MAS Model I, where all the words in $\Lambda_{normalized}$ has at least one schwa (by construction), in MAS Model II, a large number of words in Λ do not contain any schwa (Fig. 6.6. The outcome of an imitation game

[4]Even though according to the current implementation of the model, it is the articulatory process where the noise is being introduced, conceptually one may as well consider the noise to be a part of the transmission channel or the perception process.

with WWS is obvious, because there is only one rule r_d that is used with a probability 1, and such a game is always successful. Therefore, in order to strip of any unnecessary computation during the simulation experiments, *the initiator always chooses a word that has at least one schwa.* Nevertheless, as the game proceeds, the imitator might perceive and consequently, articulate a word, which does not have any schwa.

6.4.1 Evaluation Metrics

A simulation experiment can be thought to have the following three phases:

1. **Initialization:** The parameter values and the mental states of the agents are initialized.

2. **Execution:** The agents play a preset number of games, which are either successes or failures. At the end of a fixed number of games (one or more), which we call a *round* of the simulation, the state of the system is noted in a logfile.

3. **Conclusion:** At the end of the simulation, every agent is in a particular mental state. In other words, every agent learns a specific set of schwa deletion rules, which are reflected through the probabilities associated with the pronunciation pattern.

Table 6.4.1 shows a few pronunciation rules that emerge during a typical run of the MAS Model II. It is interesting to note that synchronic variation (i.e., variation across agents) is observed for almost all the pronunciation rules, despite a high communication success rate (94%). We also observe that for frequent CV maps, such as aCVCa, the pronunciation rules have probability of either 0 or 1. Thus, even though the agents may not converge to the same pronunciation rule, they do learn *strict* individual pronunciation rules. In other words, variation within an individual vanishes over time. The CV maps, such as VCCaCaCCV, for which we do not observe the aforementioned trend are infrequent, and very few games had been played with the words associated

Antecedent Ω	Pronunciation ω	Probabilities			
		Agent 1	Agent 2	Agent 3	Agent 4
CaCaCV	CCCV	0.0	0.0	0.0	0.003
	CaCCV	**1.0**	**1.0**	**1.0**	**0.990**
	CCaCV	0.0	0.0	0.0	0.003
	CaCaCV	0.0	0.0	0.0	0.003
CaCaCaCa	**CaCCaC**	**0.0**	**0.0**	**0.0**	**1.0**
	CaCCaCa	1.0	1.0	1.0	0.0
aCVCa	**aCVC**	**1.0**	**0.0**	**1.0**	**1.0**
	aCVCa	0.0	1.0	0.0	0.0
aCVCaCa	CVCC	0.014	0.006	0.014	0.0
	aCVCC	0.014	0.006	0.014	0.0
	CVCaC	0.014	0.96	0.014	0.0
	aCVCaC	**0.014**	**0.006**	**0.014**	**0.0**
	aCVCC	0.90	0.006	0.899	1.0
	aCVCCa	0.014	0.006	0.016	0.0
	CVCaCa	0.014	0.006	0.014	0.0
	aCVCaCa	0.014	0.006	0.014	0.0
VCCaCaCCV	VCCCCCV	0.54	0.54	0.0	1.0
	VCCaCCCV	0.42	0.13	0.0	0.0
	VCCCaCCV	**0.20**	**0.24**	**1.0**	**0.0**
	VCCaCaCCV	0.19	0.09	0.0	0.0

Table 6.5: Examples of emergent rules taken from a typical run of the MAS Model II. The probabilities associated with the pronunciation patterns of CV Map Ω are shown for all the four agents. The pronunciation according to standard Hindi is shown in bold font. The patterns for which all the agents have a probability of 0.0 are not shown in the table (e.g., CVC for the antecedent aCVCa). Values of the model parameters: $N = 4$, $M = 10000$, $\pi = 0.8$, $\delta = 0.1$, $pr_d = 0.6$, $\eta = 0.1$, number of games = 1 million, $\langle SuccessRate \rangle = 94\%$

with them. Consequently, the agents do not get sufficient time to converge to a particular pronunciation.

Since there is a large number of pronunciation rules, it is impossible to evaluate and characterize the emergent schwa deletion pattern of MAS Model II through manual inspection only, as we have done in the case of MAS Model I. Therefore, in order to understand the nature of the emergent pattern as well as the dynamics of the system, we define certain objective evaluation metrics that help us to – (a) talk about the stability of the emergent linguistic system, and (b) measure its similarity with its naturally occurring counterpart. Like MAS Model I, we measure the stability of the emergent linguistic system in terms of its *success rate*, which is defined as follows.

$$SuccessRate = \frac{\text{Number of successful games}}{\text{Total number of games played}} \tag{6.18}$$

After a sufficiently large number of games, the higher the value of *SuccessRate*, the higher is the stability of the emergent system. The *SuccessRate* of an ideal linguistic system is 1.

In order to measure the similarity of the emergent pattern with the pattern of SDH, we define four independent metrics based on the Ohala's rule. Recall that Ohala's rule succinctly captures the Hindi schwa deletion pattern through a regular expression for the context where a schwa can be deleted (see Sec. 2.3.2). However, deletion of a schwa can change the context of another schwa so that it cannot be deleted anymore, even if the deletion was supported by the context previously (refer to the case of *ajanabi* in Example 2.1). Such cases are resolved by the *right-to-left rule application* convention.

Thus, according to the context of deletion (as specified by Eq. 2.1), we can classify the schwas occurring in a CV-map Ω as *delible* and *indelible*. However, after applying the convention as well, the schwas get classified finally as either *deleted* or *retained*. Note that a schwa that is *indelible* is always *retained*. However, a schwa that is *delible* might be *deleted* or *retained*.

Let ω be a CV pattern obtained by deletion of zero or more schwas from Ω. In other words, ω is a pronunciation pattern associated with the CV-map Ω. Let us consider a particular schwa in Ω. The schwa may be *deleted* in Ω according to Ohala's rule as well as deleted in ω. Let us denote these types of schwas as *correct deletions*. Similarly, the schwa may be *retained* in both Ω and ω. We shall call such schwas as *correct retentions*. We define the following two quantities for a pair (Ω, ω), which reflect the similarity of ω to the deletion pattern of Ω as predicted by Ohala's rule.

$$CorrectDeletion(\Omega, \omega) = \frac{\text{No. of } \textit{correct deletions}}{\text{No. of } \textit{deleted} \text{ schwas in } \Omega} \tag{6.19}$$

$$CorrectRetention(\Omega, \omega) = \frac{\text{No. of } \textit{correct retentions}}{\text{No. of } \textit{retained} \text{ schwas in } \Omega} \tag{6.20}$$

The fractions of schwas in ω that are incorrectly retained or deleted are given by the expressions $1 - CorrectDeletion(\Omega, \omega)$ and $1 - CorrectRetention(\Omega, \omega)$. A schwa that is *undelible* in Ω, but deleted in ω, directly violates Ohala's rule. Such schwas will be referred to as *erroneous deletions*, for which we have the following equation.

$$ErroneousDeletion(\Omega, \omega) = \frac{\text{No. of } \textit{erroneous deletions}}{\text{No. of } \textit{undelible} \text{ schwas in } \Omega} \tag{6.21}$$

The above equations measure the similarity and error related to a single pronunciation pattern. In order to generalize the equations over a pronunciation rule of the type

$$\Omega \rightarrow \langle \omega_0, pr_0 \rangle, \langle \omega_2, pr_2 \rangle, \ldots, \langle \omega_{2^k-1}, pr_{2^k-1} \rangle$$

we simply take a weighted sum of the respective quantities, where the weights are the probabilities. Thus, if we denote the above rule as $\mathbf{r}$, we have

$$CorrectDeletion(\mathbf{r}) = \sum_{i=0}^{2^k-1} pr_i \times CorrectDeletion(\Omega, \omega_i) \tag{6.22}$$

$$CorrectRetention(\mathbf{r}) = \sum_{i=0}^{2^k-1} pr_i \times CorrectRetention(\Omega, \omega_i) \tag{6.23}$$

$$ErroneousDeletion(\mathbf{r}) = \sum_{i=0}^{2^k-1} pr_i \times ErroneousDeletion(\Omega, \omega_i) \tag{6.24}$$

For a particular simulation experiment, the similarity metrics can be further averaged over all the rules in R and over all the agents in the system. While averaging over all rules, we also weigh the quantity associated with each rule by the fraction of words in the lexicon for which the rule is used. We shall denote these final averages for a particular simulation as $\langle CorrectDeletion\rangle$, $\langle CorrectRetention\rangle$ and $\langle ErroneousDeletion\rangle$.

Thus, to summarize, we use four evaluation metrics:

1. *SuccessRate*: Higher the value, higher is the stability of the emergent system.

2. $\langle CorrectDeletion\rangle$: Higher the value, the closer is the emergent pattern to Ohala's rule and the pattern features proper deletion characteristics.

3. $\langle CorrectRetention\rangle$:Higher the value, the closer is the emergent pattern to Ohala's rule and the pattern features proper retention characteristics.

4. $\langle ErroneousDeletion\rangle$: Lower the value, the closer is the emergent pattern to Ohala's rule and the pattern does not violate the constraints on deletion.

6.4.2 Baseline Measurements

We define a baseline experiment as the one where the learning rate $\eta = 0$. Naturally, when $\eta = 0$, the pronunciation rules in the mental models of the agents do not change over time, and as a result all the evaluation metrics defined in the previous subsection remain fixed over time.

We conduct two sets of baseline experiments: the first one is for studying the effect of initialization strategies and the second one for studying the effect of lexicon size. Fig. 6.11 and 6.12 respectively shows the variation of the average *success rate* and the other three evaluation metrics with π - the probability of deletion of an individual schwa. All the results have been obtained by averaging the respective values over 10000 games. The variation of $\langle SuccessRate\rangle$ is reported for two different initialization strategies - fixed schwa deletion probability based assignment and random probability assignment. The significance of π for the former initialization strategy has been described earlier. In the case of latter, π denotes the average schwa deletion probability of all the rules for all the agents. In other words, in the fixed probability based initialization strategy the system has no synchronic variation in the sense that all the agents share

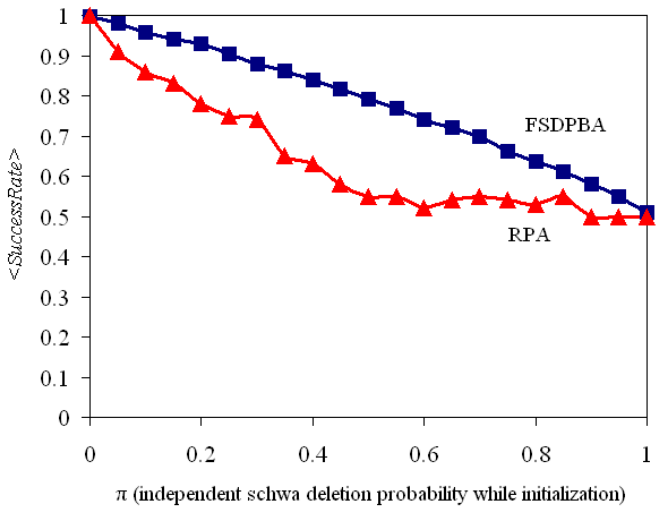


Figure 6.11: The average *SuccessRate* for the baseline experiments with $M = 10000, N = 4, \delta = 0.1, pr_d = 0.6$. The probability of deletion of an individual schwa - π is plotted in the x-axis. FSDPBA and RPA denote fixed schwa deletion probability based assignment and random probability assignment.

the same pronunciation rules, whereas with random assignment, the system features synchronic variation.

It can be shown from the definitions of the baseline experiments and fixed probability based initialization strategy that the values for $\langle CorrectDeletion\rangle$, $\langle CorrectRetention\rangle$ and $\langle ErroneousDeletion\rangle$ with this initialization strategy (when the schwa deletion probability is set to π,) are π, $1-\pi$ and π respectively.Therefore, in Fig. 6.12, we report the variation of these evaluation metrics only for the random assignment strategy.

The observations are summarized below.

- $\langle SuccessRate\rangle$ monotonically decreases with π. When $\pi = 1$, none of the schwas are deleted. This condition refers to the case of Sanskrit, and we observe that it is an ideal linguistic system with $\langle SuccessRate\rangle = 1$. $\langle SuccessRate\rangle$ gradually falls to 50% as π increases to 1.
- As is expected, the system without synchronic variation exhibits better

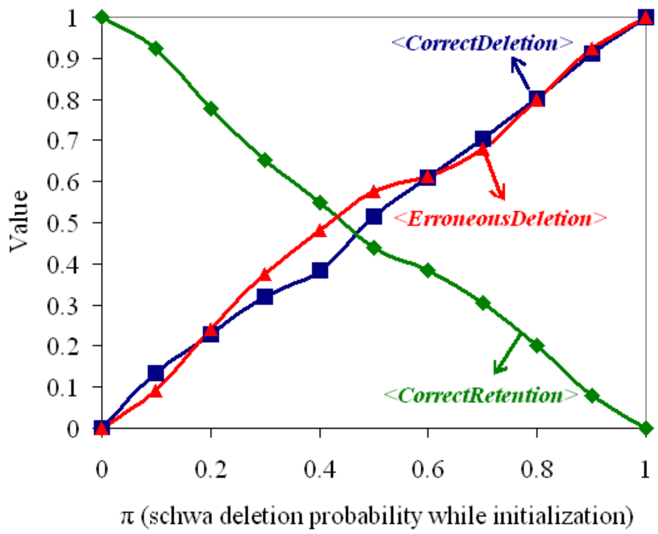


Figure 6.12: The average values of $CorrectDeletion$, $CorrectRetention$ and $ErroneousDeletion$ for the baseline experiments with a random probability assignment based initialization strategy. $M = 10000, N = 4, \delta = 0.1, pr_d = 0.6$.

communicative success than the system with synchronic variation. A maximum difference of 24% in the $\langle SuccessRate \rangle$ of the two cases is observed when $\pi = 0.5$. This can be explained as follows. For $\pi = 0$ or 1, we cannot have any synchronic variation in the system. As one moves away from these extremes, the scope of variation also increases. Therefore, the difference between the two systems (with and without variation) and consequently that in the value of $\langle SuccessRate \rangle$ is highest, when $\pi = 0.5$.

- The similarity of the rules to that of Ohala's, as reflected by the quantities $\langle CorrectDeletion \rangle$, $\langle CorrectRetention \rangle$ and $\langle ErroneousDeletion \rangle$, is a linear function of π and the results for the random assignment strategy matches quite well with that of the fixed schwa deletion based probability assignment strategy as discussed above.

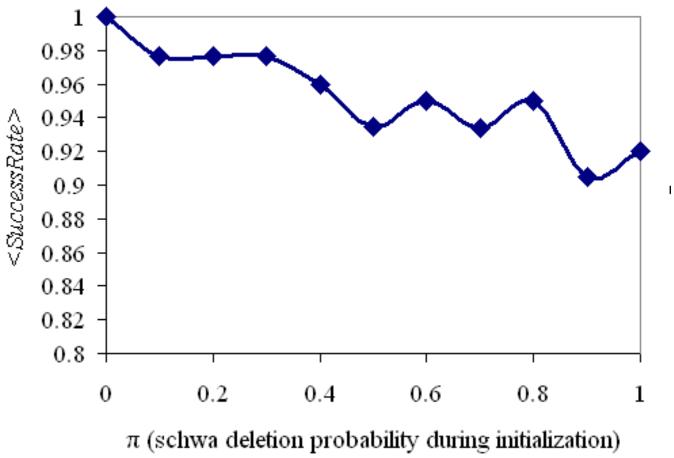


Figure 6.13: The plot of π versus $\langle SuccessRate \rangle$. The values of other model parameters: $N = 4, M = 10000$, $\eta = 0.1, \delta = 0.1, pr_d = 0.6$, number of games $= 10$ million

6.4.3 Effect of Rule Initialization

We choose to experiment with the fixed schwa deletion probability based initialization strategy only, because it is equivalent to the uniform assignment strategy for $\pi = 0.5$ and as is evident from the baseline experiments, the results of random assignment strategy are not significantly different from that of this strategy. We have only one parameter π associated with this strategy. Keeping other model parameters fixed ($N = 4, M = 10000$, $\eta = 0.1, \delta = 0.1, pr_d = 0.6$, number of games $= 10$ million) we conduct experiments for different values of π ranging from 0.0 to 1.0. The results are shown in Fig. 6.13 and 6.14.

In Fig. 6.13, we observe that after 10 million games $\langle SuccessRate \rangle$ is quite high and varies slightly with π. Like the baseline experiments, the value of $\langle SuccessRate \rangle$ is highest for small values of π, but unlike the case of baseline experiments, with learning communicative success also increases for large values of π. For instance, with $\pi = 0.8$, the value of $\langle SuccessRate \rangle$ is 0.61 for the baseline experiment, which increases to 0.95 when the agents are allowed to learn. In general, the high communicative success for all π indicates the inherent

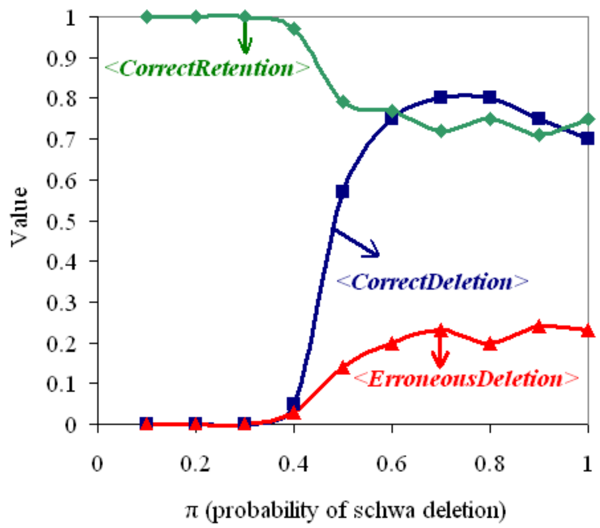


Figure 6.14: The plot of π versus $\langle CorrectDeletion \rangle$, $\langle CorrectRetention \rangle$ and $\langle ErroneousDeletion \rangle$. The values of other model parameters: $N = 4, M = 10000$, $\eta = 0.1, \delta = 0.1, pr_d = 0.6$, number of games = 10 million

stability of the linguistic system entailed by the MAS Model II.

Nevertheless, for all values π, the emergent pattern is not similar to the one that is predicted by Ohala's rule. It is evident from Fig. 6.14 that the system undergoes a *bifurcation* at $\pi = 0.5$. When $\pi < 0.5$, the near-zero values of $\langle CorrectDeletion \rangle$ and $\langle ErroneousDeletion \rangle$, and very high value of $\langle CorrectRetention \rangle$ indicate that hardly any schwas are deleted in the emergent pattern. The pronunciation of the system, therefore, matches with Sanskrit. On the contrary, when $\pi > 0.5$, we observe that the value of $\langle CorrectDeletion \rangle$ abruptly increases to around 0.8 and $\langle CorrectRetention \rangle$ drops from 1 to almost the same level. The cases of errorneous deletion also increase to 20%. Thus, the emergent pattern in this case shows around 80% match with Ohala's rule.

In order to understand the nature of the erroneous deletions, we classify the deletion errors into three categories: (1) deletion of a schwa word initially (eg. $aCaCV \rightarrow CaCV$), (2) deletion before a consonant cluster (eg. $CVCaCCV \rightarrow CVCCCV$, and (3) deletion in the first syllable (eg. $CaCV \rightarrow CCV$). Fig. 6.15 shows the absolute fraction of these three types

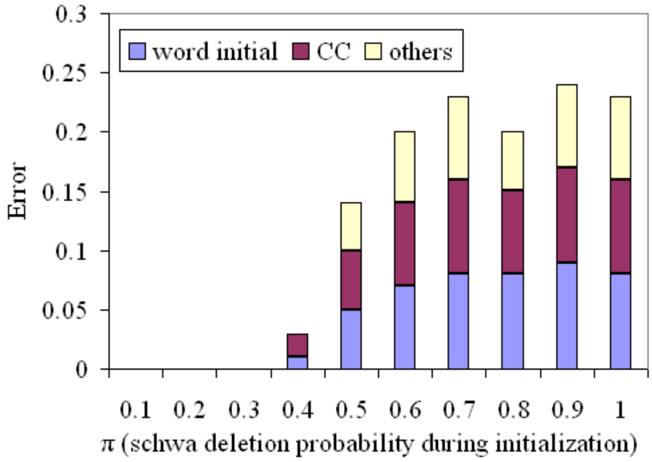


Figure 6.15: The histogram of π versus the absolute fraction of the different types of deletion errors. "Word Initial" and "CC" indicates the incorrect deletion of a schwa word initially and before a consonant cluster respectively.

of errors that sum up to $\langle ErroneousDeletion \rangle$. We observe that these three types of deletion errors almost equally account for the erroneous deletions.

6.4.4 Effect of Lexicon Size

Next, we study the effect of the lexicon on the emergent pattern. For this we conduct experiments, where all the parameters, but the lexicon size M, are kept fixed. More precisely, we choose the values of N, η, π, δ and pr_d to be 4, 0.1, 0.8, 0.1 and 0.6 respectively. The value of π is so chosen, because it has been observed from the experiments described in the last subsection that the emergent pattern resembles best to Ohala's rule for this particular value of π. Since the number of games required for convergence is dependent on M, and since it is difficult to define the convergence point, we choose to run the simulation till the system has an average $SuccessRate$ of 0.95. Table 6.6 summarizes the number of games required to achieve this success rate for different M.

Fig. 6.16 shows the variation of the different evaluation metrics with M. As M increases, the value of $\langle CorrectDeletion \rangle$ decreases, whereas that of

M	1000	2000	4000	10000	20000
Number of games	0.5M	1M	2M	2M	2M

Table 6.6: Number of games required for different values of M such that $\langle SuccessRate \langle = 0.95$. The values of other model parameters: $N = 4$, $\eta = 0.1, \pi = 0.8, \delta = 0.1$, and $pr_d = 0.6$.

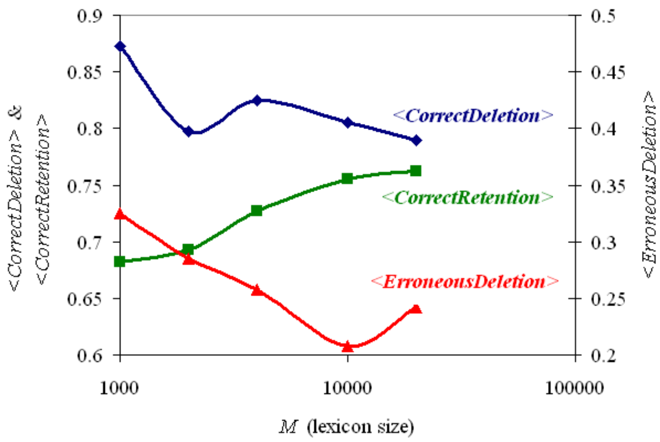


Figure 6.16: The variation of $\langle CorrectDeletion \rangle$, $\langle CorrectRetention \rangle$ and $\langle ErroneousDeletion \rangle$ with lexicon size - M. The values of other parameters: $N = 4$, $\eta = 0.1, \pi = 0.8, \delta = 0.1, pr_d = 0.6$, $\langle SuccessRate \rangle \approx 0.95$. The scale for $\langle CorrectDeletion \rangle$ and $\langle CorrectRetention \rangle$ are shown on the left, and the scale for $\langle ErroneousDeletion \rangle$ is shown on the right. The x-axis is in logscale.

$\langle CorrectRetention \rangle$ increases. The value of $\langle ErroneousDeletion \rangle$ initially decreases with M, but it achieves a minimum value for $M = 10000$, after which it increases again. These observations can be explained as follows. As more and more words are packed into the lexicon, the chance that an arbitrary deletion leads to the loss of lexical distinctions increases and consequently, the deletions are treated with a greater strictness (restrictions). Naturally, the cases of correct retention and erroneous deletions go up. As an unavoidable consequence of the same, the cases of correct deletions are also reduced.

Nevertheless, as M increases from 1000 to 10000, the value of $\langle CorrectDeletion \rangle$ falls by around 0.07 (i.e. 7.7%), whereas the value of $\langle ErroneousDeletion \rangle$ reduces by 0.12 (i.e. 36.2%). Thus, the net effect of increasing the lexicon size on the emergent pattern is positive in the sense that the loss incurred in correct deletions is by far compensated by the gain with respect to erroneous deletions. In other words, with an increase in M, the number of deletions gets reduced, but majority of the deletions that are prevented are those which were erroneous according to Ohala's rule. However, we cannot explain the increase in the value of $\langle ErroneousDeletion \rangle$ as M increases from 10000 to 20000. Due to the lowest error rates observed, M has been set to 10000 for the experiments with fixed lexicon.

Fig. 6.17 shows the detailed break-up in the different types of deletion errors. We observe that the reduction in the number of erroneous deletion with increasing M is primarily due to the reduction in the number of deletion errors in the word initial position. The fraction of deletions accounted for by the two other types of errors, namely deletion before consonant cluster and deletion in the first syllable, are more or less independent of M. This observation further corroborates our aforementioned claim that as the lexicon gets larger, deletion of a word initial schwa leads to the loss of lexical distinctions.

6.4.5 Effect of Learning Rate

The parameter η (learning rate) affects the dynamic behavior as well as the emergent pattern of the system to a large extent. η determines the relative weightage an agent pays to its previous experiences and the outcome of the most recent game. A high value of η signifies that the outcome of the last game is what is more important. Stated differently, the agents do not have a strong memory, and as a result pays least importance to the past experiences. This corresponds to a greedy strategy. To the contrary, when η is very small, the overall history is more important than just the outcome of the current game. In such a case, the aim of the agents is to optimize the success rate over a long period of time.

Fig. 6.18 shows the growth pattern of $\langle SuccessRate \rangle$ with time for three different values of η. The other model parameters are same for the three cases. We observe that the for a very small η (=0.01), the average communicative success of the system increases slowly with time (read number of games) and thus,

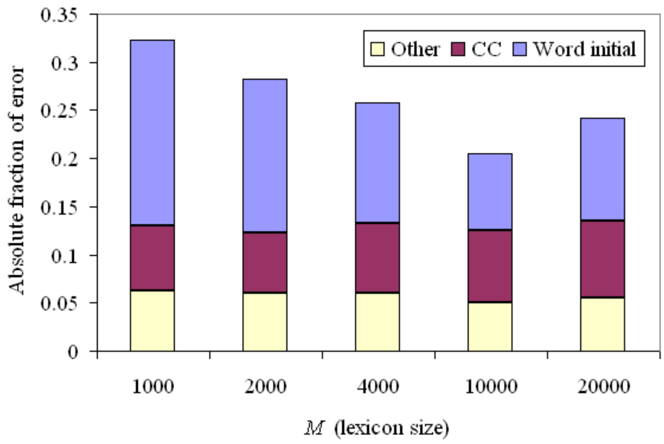


Figure 6.17: The histogram of M versus the absolute fraction of the different types of deletion errors. "Word Initial" and "CC" indicates the incorrect deletion of a schwa word initially and before a consonant cluster respectively. The values of other parameters: $N = 4$, $\eta = 0.1, \pi = 0.8, \delta = 0.1, pr_d = 0.6$, $\langle SuccessRate \rangle \approx 0.95$

a large number of games are required for convergence. The final communicative success of the system, however, is quite high. On the other hand, when η is large ($= 1.0$), $\langle SuccessRate \rangle$ grows and stabilizes quickly. Nevertheless, this fast stabilization is obtained at the expense of a lower communicative success eventually. For medium values of η ($= 0.1$), we have both fast stabilization and high value of $\langle SuccessRate \rangle$.

The fact that learning rate also has a significant effect on the emergent pattern, is substantiated in Fig. 6.19. Quite strangely, $\langle CorrectRetention \rangle$ increases steadily with η, whereas $\langle CorrectDeletion \rangle$ decreases sharply with η. This can be explained as follows. When η is large, the system cannot achieve high communicative success rate due to the greedy decisions made by the agents during learning. Consequently, the agents try to retain the schwas to maintain a high communicative success. Thus, $\langle CorrectDeletion \rangle$ decreases and $\langle CorrectRetention \rangle$ increases with an increase in η. However, the value of $\langle ErroneousDeletion \rangle$ does not decrease monotonically with η, because when the learning rate is high, the agents do not learn much from the errors made

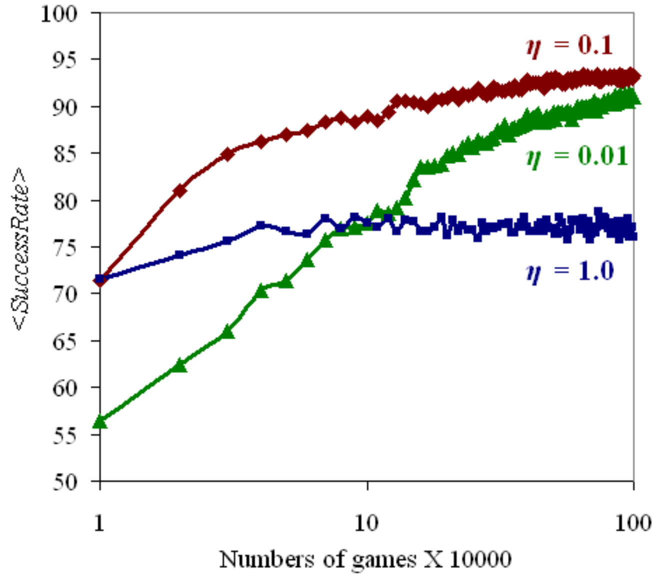


Figure 6.18: The plot of time (measured in number of games) versus $\langle SuccessRate \rangle$ for different learning rates - η. The values of other parameters: $N = 4$, $M = 10000, \pi = 0.8, \delta = 0.1, pr_d = 0.6$

earlier in the history. The schwas that are deleted or retained are merely an outcome of random chance!

The high deletion error rate for $\eta = 0.01$ as compared to that when $\eta = 0.1$, is due to the fact that with such a slow learning rate, the system requires more time to converge. This is also evident from the fact that the curve of average success rate of the system (Fig. 6.18) for $\eta = 0.01$ shows a positive slope after 1 million games, which indicates that the system (and consequently the pattern) is yet to converge.

6.5 Analysis and Interpretation

The observations made from the experiments with MAS Model II (as discussed in the previous section) agree with those made for MAS Model I. We see that in both the cases, under suitable conditions, the emergent pattern closely re-

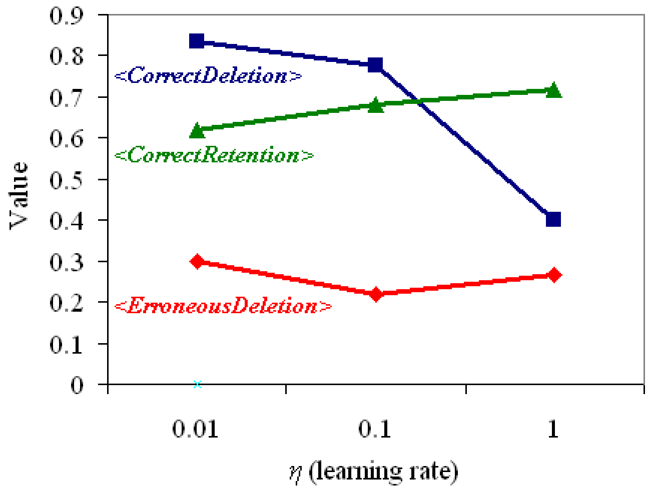


Figure 6.19: The plot of learning rate η versus the values of the evaluation metrics $\langle CorrectDeletion\rangle$, $\langle CorrectRetention\rangle$ and $\langle ErroneousDeletion\rangle$. The values of other parameters: $N = 4$, $M = 10000, \pi = 0.8$, $\delta = 0.1, pr_d = 0.6$, number of games = 1 million

sembles the context of SDH. Furthermore, the conditions necessary for the emergence of the pattern, such as non-greedy deletion and learning strategies, dependence of the emergent pattern on the lexicon and a tendency towards reduction of vowel deletion, are also comparable for the two models. Thus, the explanations offered by MAS Models I and II are similar in nature. As we have argued in Sec. 5.4 that MAS Model I entails a phonetically-based explanation for SDH, in the same lines it can be argued that the explanation offered by MAS Model II is also based on phonetic factors.

Apart from the fact that in MAS Model II we propose a more efficient computational method for realizing the perceptual process, the articulatory and perceptual processes of both the models are identical in essence. Since a phonetically-based account of sound change primarily banks on the nature of articulatory and perceptual errors to explain a phonological change, it is not surprising that the two models yield similar results.

Plausibility of MAS Model II

However, the main difference between the MAS Models I and II resides in the definition of the mental model **M** and the learning process LEARN of an agent. While in Model I, the pronunciations are stored and learnt explicitly for each of the words in Λ, in Model II, the agents learn generalized pronunciation rules, which are probabilistic in nature. In the context of this thesis, there are two important advantages of the latter representation:

- **Plausibility:** Since almost all of the popular schools of phonology (e.g., generative phonology (Chomsky and Halle 1968), optimality theory (Prince and Smolensky 1993; Prince and Smolensky 1997), functional phonology (Boersma 1998)) argue in favor of phonological generalizations, mental model for MAS Model II is linguistically much more plausible than that of Model I.

- **Computability:** Generalization of pronunciations over words solves the problem of pronunciation-rule explosion, and consequently, makes the model scalable. The obvious advantage of this is the possibility of experimentation with large and realistic lexica. This, in turn, places the model as well as its results on a firmer linguistic ground.

Lexicon and the Emergent Pattern

In Sec. 5.4, we have already discussed the various issues related to the perceptual and articulatory processes of the MAS model. Therefore, here we focus on the effect of the lexicon on the emergent pattern, which could not have been studied for MAS Model I.

Let us analyze the results presented in Sec. 6.4.4 and more specifically, those shown in Fig. 6.16. As we increase the size of the lexicon, the accuracy of the emergent pattern increases, primarily due to the decrease in the number of erroneous deletions and a corresponding rise in the number of correct retentions. This observation can be explained as follows. As we increase the size of the lexicon, the acoustic distinctiveness constraints become stringent, because, now there is a higher chance that the deletion of a schwa leads to the loss of lexical distinctions (i.e., after deletion, the word "maps to" or "sounds like" another word in Λ). Since this reduces the communicative success of an agent, deletions that lead to the loss of lexical distinctions are prohibited. As a result, with an increase in the size of Λ, the context of deletion becomes more stringent and in turn, leads to the emergence of the correct pattern.

Phase-transition and S-shaped Dynamics

We conclude our analysis of MAS Model II with some comments on its behavior from the perspective of an adaptive dynamical system. In Fig. 6.13, we see that the average *SuccessRate* of the system varies slowly with π – the initial bias towards deletion. Thus, whatever may be the initial bias, the system always stabilizes at a high communication success rate ($> 90\%$). However, as shown in Fig. 6.14, the nature of the emergent pattern shows a sharp transition at $\pi = 0.5$. For $\pi < 0.5$, there is no deletion, and thus, the emergent pronunciation resembles that of Sanskrit or Old-Hindi. On the other hand, for $\pi > 0.5$, the emergent pattern resembles that of modern Hindi (i.e., the context for SDH as predicted by Ohala's rule). A sharp transition takes place at $\pi = 0.5$; stated differently, the system undergoes a *bifurcation* or equivalently, a *phase-transition* at $\pi = 0.5$.

The S-shaped curve observable for all the three metrics, and more prominently for $\langle CorrectDeletion \rangle$ (Fig. 6.14), is a direct consequence (as well as evidence) of the bifurcation. Note that the S-shaped curve shown in Fig. 5.3 is for the duration of a single schwa, whereas the curve currently under consideration is for the whole pattern. Thus, we conclude that the S-shaped dynamics is an inherent feature of language change and is observed at every level.

From the perspective of sound change, this particular dynamics implies that prediction of the occurrence of a sound change is impossible, even if the prediction of its direction is possible. In the context of schwa deletion, one can say that if there is a phonological change leading to schwa deletion, then the deletion pattern is expected to be like that of SDH. However, it is not possible to predict whether the change will ever take place, because that depends on the critical value of π – the propensity towards deletion. Note that this is a very simplistic analysis of the problem, as we do not consider here several other factors, including socio-economic issues, that might trigger or inhibit a particular case of change. Nevertheless, the basic dynamics of the change remains the same, as observed in the MAS experiments.

6.6 Conclusion

In this chapter, we have presented a MAS model for explaining the SDH pattern of Hindi. The model captures the real world scenario to a good extent as it supports experimentation with real world lexicon. The emergent behavior of the model resembles that of the real world, leading us to believe that it indeed offers a plausible explanation for SDH.

It is worthwhile to discuss the relationship between MAS Models I and II. While we have already discussed the benefits of MAS Model II over I in the previous section, one cannot ignore the importance of Model I. Model I makes far less assumptions regarding the mental representations and cognitive processes of human beings. In other words, Model I is more general, and consequently, more powerful than Model II. For every pattern or dynamics that emerge in Model II, there is a possibility that they also emerge in Model I. However, the reverse is not true. For example, in Model I, an agent might learn to pronounce $katarA$ as $katrA$ and $kadali$ as $kadali$, which is impossible for Model II.

Nevertheless, this extra power for Model I comes with a cost – the model is computationally intensive. On the other hand, the extra assumptions made in Model II are not baseless. In fact, most of them are directly based on the findings from Model I (Sec. 6.1). This we refer to as the *hierarchical abstraction methodology* towards modeling of real world language change, and is an important contribution of this thesis. The methodology can be stated as follows:

> *The findings of a more general, but computationally intensive, model are used to abstract out the details in the subsequent phases of modeling. This in turn leads to models that are not as general as*

their precursors, but computationally more tractable and linguistically equally plausible.

In the last three chapters, including this one, we have presented three models for explaining the pattern of SDH. Schwa deletion, nevertheless, is a simple phonological phenomenon, where only a single vowel is deleted in specific contexts. Phonological changes, in general, take place simultaneously over the words of a lexicon such that the resulting lexicon (or lexica) is significantly different from its original counterpart. To complicate things further, other strata of the language, such as morphology, syntax and semantics, influence and get influenced by the phonological changes.

In the next chapter, we attempt to model the changes that affected the morpho-phonology of Bengali verb-inflections by utilizing the aforementioned hierarchical abstraction methodology. We presume that a simple phonological changes, such as deletion or insertion of a vowel, are explainable by more general models similar to the MAS models presented here, and take them as the basic units of change. These abstract units of change are used to develop a constrained optimization model for the problem of Bengali verb morphology.

Chapter 7

Phonological Change of Bengali Verb Inflections

In the last three chapters we have dealt with the problem of SDH. SDH, however, is a simple case of phonological change. In general, several phonological changes take effect in sequence, giving rise to one or more than one language or dialects of a language. The objective of this chapter is to develop computational models of more complex phonological changes, where at the end of the process of change we have a set of dialects rather than a single emergent language.

In the past, researchers have shown that dialect diversity emerges through the process of language change in MAS-based models (Fyfe and Livingstone 1997; Livingstone 2002; Livingstone 2003), especially when the agents are grounded in a social network (Nettle 1999; Gong et al. 2004). In fact, we have observed synchronic variations among the agents in both the MAS Models, even though there were no linguistic communities, precisely due to the uniform probability of interaction between any pair of agents. However, none of the aforementioned works pertaining to dialect diversity are based on real languages.

Modeling a case of emergence of dialects in the real world is difficult for several reasons. Language can evolve along infinitely many different paths, giving rise to infinitely many dialects. Nevertheless, in reality, only some of these paths are manifested and we can observe only a very small fraction of the possible dialects. Therefore, it is impossible to validate a model satisfactorily, even when one has a large amount of real world data.

In this chapter, we present an initial attempt towards modeling and validation of dialect diversities in the real world. For this purpose, we choose the case of phonological change affecting the Bengali Verb Inflections (BVI), which has been discussed in Sec. 2.4. The chapter is organized as follows. In Sec. 7.1, a computational formulation of the problem is presented, followed by a functional explanation for the same. Sec. 7.2 describes a multi-objective genetic algorithm (MOGA) based model for capturing the functional explanation. Sec. 7.3 details out the experiments and their results. The analysis and interpretations of the results are presented in Sec. 7.4. Sec. 7.5 discuss a potential application of the diachronic model. Sec. 7.6 summarizes the findings of the MOGA models and enumerates some of the future works.

7.1 Problem Formulation

We choose the classical Bengali forms, i.e. the dialect that was spoken around 1200 AD, as our starting point. It is observed that a sequence of simple phonological changes, which we shall call the *Atomic Phonological Operators* or APO for short, when applied over the classical Bengali lexicon gives rise to the modern dialects. Our objective here is to model the emergence of the BVIs in the modern dialects from their classical counterparts through appropriate sequence of APOs. Note that the derivations of the BVIs can be constructed manually, as has been done in (Chatterji 1926), or automatically using language reconstruction tools such as (Lowe and Mazaudon 1994; Nakhleh et al. 2005). However, our aim is not limited to the derivations of a few dialects; rather, we want to show that certain sequences of APOs, and consequently, the resultant dialects are preferred and observed in nature, among the myriads of possibilities.

We shall attempt to encode this "preference" as a functional selection, where there are several conflicting functional pressures over the evolving language.

7.1.1 Dialect Data

Several dialects of Bengali are spoken allover Bangladesh and parts of India. Since early 1900, there has been some effort to collect and classify the dialects. Azad (2001), a collection of several scholarly essays on Bengali published between 1743 and 1983, contains 11 articles related to dialectology and dialect data of Bengali. Among these, Grierson (1903) is one of the earliest and most comprehensive collection of the different Bengali dialectal forms; Chatterji (1927) emphasizes on the collection and preservation of rural dialectal forms and attempts to construct the etymological information of the rural dialects; Maniruzzaman (1974) analyzes the dialect spoken in Dhaka and identifies interesting phonological transformation rules between SCB and the Dhakan dialect.

Traditionally, the Bengali dialects are classified into four groups (Islam 1979):

- North Bengali (spoken in Dinajpur, Rajshahi, Bagura and Pabna)
- Rajbanshi (spoken in Rangpur)
- East Bengali – further classified into the dialects of the three regions
  a. Dhaka, Maimansing, Tangail, Kumilla, Barishal, Patuwakhali
  b. Faridpur, Jessore, Khulna
  c. Sylhet
- South Bengali (spoken in Chittagong and Nowakhali)

Although a few previous studies report some of the verbal inflections used in the different dialects (see, e.g., Chatterji (1926) for list of inflections in SCB, Classical Bengali and a few other dialects, and Islam (1979) for very brief discussion on some of the inflected forms), we do not know of any previous study, during which the different dialectal forms for BVI were collected and systematically listed. In the absence of any comprehensive dialect data for Bengali, we choose to work with primary data for three modern dialects, which have been collected during this work. The dialects are

Feature	Possible Attributes
Person	1^{st}, 2^{nd}, 3^{rd}, honorific*
Tense	past, present
Aspect	simple, continuous, perfect, habitual**

Table 7.1: Subset of attributes for BVI used in the model. * refers to the formal form in the 2^{nd} and 3^{rd} persons. ** Only in the past tense.

- *Standard Colloquial Bengali* (SCB) spoken in a region around Kolkata, the capital of West Bengal,
- *Agartala Colloquial Bengali* (ACB) spoken in and around Agartala, the capital of Tripura, and
- *Sylheti*, the dialect of the Sylhet region of Bangladesh.

The corresponding verb forms for these three dialects have been obtained by enquiring the näive informants and are listed in Appendix E.1. The geographical locations of the above places are shown in Appendix E.3.

Note that the alignment of the corresponding forms in the four dialects (the classical and 3 modern dialects) is straightforward: the inflections uniquely determine the morphological attributes of the form, such as tense, aspect and person (see Sec. 2.4.2 and Appendix D for details). Thus, we shall denote the meaning of a particular form by a subset of the morphological attributes as described in Table 7.1. We do not consider all the attributes, but only a subset of them, because in classical Bengali certain attributes are realized using separate words rather than inflections (e.g., the *second familiar* person or the negative polarity). Similarly, the future tense has been omitted as it features only one of the aspects (i.e., simple); moreover, the inflections for first person in the future tense in ACB and Sylheti – *um* and *mu* respectively, are not derivable from the classical Bengali inflection *ba* and the inflections have to be derived directly from Sanskri or Prakrit (*mi* or *mah*).

As described in Sec. 2.4.2, SCB has 19 morphological paradigms, which are also an outcome of the phonological change that affected the inflections. Neither classical Bengali, nor ACB or Sylheti feature those morphological paradigms. In other words, except for SCB, in all the other dialects, the verbs inflect uniformly irrespective of the structure of the root. Therefore, in our model we restrict our study to the various (28 to be precise) inflected forms of a single verb root, i.e., kar. This is not to say that the emergence of the morphological paradigms in SCB is uninteresting to study. We deem this to be a future extension of this work.

Thus, to summarize, every dialect is represented by a lexicon of 28 inflected forms of the root kar, where the meaning of each of the form is to be interpreted as the set of its morphological attributes.

7.1.2 Atomic Phonological Operators

We conceive a complex phonological change as a sequence of several basic phonological changes or APOs. We define four basic types of APOs:

- *Del* or *Deletion* of a phoneme
- Met or *Metathesis* of a string of two phonemes
- *Asm* or *Assimilation* of a phoneme in the context of another
- Mut or *Mutation* of a phoneme

The complete specification of an APO includes the specification of its type, the phoneme(s) that is(are) affected by the operation and the left and right context of application of the operator specified as regular expressions on phonemes. Thus, if *LC* and *RC* are two regular expressions denoting the left and right contexts respectively, then the semantics of the basic APOs in terms of rewrite rules are as shown in Table 7.2.

Here, p, p', p_i and p_j represent phonemes. In case of assimilation, p' is determined on the basis of assimilation of a particular phonological feature of p with respect to the neighboring phoneme (i.e. *LC* or *RC*). Therefore, the feature with respect to which the assimilation takes place must also be specified. However, since in Bengali, only vowel height assimilation is observed, we assume the feature to be the "height of the vowel". Also note that the operator Mut

APO	Semantics
$Del(p, LC, RC)$	$p \rightarrow \phi / LC—RC$
$Met(p_i p_j, LC, RC)$	$p_i p_j \rightarrow p_j p_i / LC—RC$
$Asm(p, LC, RC)$	$p \rightarrow p' / LC—RC$
$Mut(p, p', LC, RC)$	$p \rightarrow p' / LC—RC$

Table 7.2: Semantics of the basic APOs in terms of rewrite rules

can capture both *Del* and *Asm*, but we assume that *Mut* only refers to the change of a particular phonological feature of a phoneme such as devoicing or nasalization. We do not consider *epenthesis* or insertion as an APO, simply because epenthesis is not observed for the case of the change affecting BVI.

The motivation behind defining APOs rather than representing the change in terms of rewrite rules is as follows. Rewrite rules are quite expressive and therefore, it is possible to represent complex phonological changes using a single rewrite rule. Nevertheless, we have not explained the emergence of complex phonological changes. The models of SDH described so far, provide plausible explanations for a simple phonological change. SDH can be represented as an APO as shown below.

$$Del(a, VCC?, C\{V, \$\})$$

This leads us to believe that the emergence of APOs can be explained through appropriate computational models. Given this assumption, our objective is to explain more complex changes.

An important point of consideration is the complexity of the contexts LC and RC. Using our experiences from SDH, we define LC and RC to be short strings over the alphabet $\Sigma_P \cup \{C, V, -, \$\}$ ($-$ and \$ represent morpheme and word boundary respectively). The length of the contexts is restricted to two symbols. Table 7.3 shows the derivation of the SCB verb forms from classical Bengali in terms of APOs. The derivations are constructed based on the data provided in (Chatterji 1926).

Note that in Table 7.3, Rule 3 has no effect on the pronunciation, however the restrictions imposed by the APOs as well as the generality of the derivations over all verb roots necessitates the use of this extra rule. Since we shall be

Rule No.	APO	Example Derivations		
		$kar-iteChe$	$kar-iten$	$kar-iAChi$
1	$Del(e, \phi, Ch)$	$kar-itChe$	NA	NA
2	$Del(t, \phi, Ch)$	$kar-iChe$	NA	NA
3	$Met(-i, \phi, \phi)$	$kari-Che$	$kari-ten$	$kari-AChi$
4	$Met(Ci, \phi, -)$	$kair-Che$	$kair-ten$	$kair-AChi$
5	$Mut(A, e, \phi, Ch)$	NA	NA	kair-eChi
6	$Mut(A, e, \phi, \$)$	NA	NA	NA
7	$Asm(a, i, \phi, \phi)$	$koir-Che$	$koir-ten$	$koir-eChi$
8	$Mut(a, o, \phi, \$)$	NA	NA	NA
9	$Del(i, o, \phi)$	$kor-Che$	$kor-ten$	$kor-eChi$

Table 7.3: Derivations of the verb forms of SCB from classical Bengali using APOs. The rule number denotes the order of application of the operators. NA: The rule is not applicable for the form.

modeling only the 28 forms for one root verb, we can ignore Rule 6, which is applicable only for future tense. Similarly, Rule 3 and 4 can be combined as $Met(ri, \phi, \phi)$, where we ignore the morpheme boundary. Thus, it suffices to have only 7 APOs for derivation of the SCB forms. Derivations of ACB and Sylheti require 6 and 4 APOs respectively.

7.1.3 Functional Explanation for Change of BVI

Let Λ_0 be the lexicon of classical Bengali verb forms. Let $\Theta : \theta_1, \theta_2, \cdots \theta_r$ be a sequence of r APOs. Application of an APO on a lexicon implies the application of the operator on every word of the lexicon. Note that an APO θ can also be conceived of as a mapping from Σ_P^* to Σ_P^*. The sequence of operators Θ, thus, represent a dialect obtained through the process of change from Λ_0, which can be represented as follows.

$$\Theta(\Lambda_0) = \theta_r(\cdots \theta_2(\theta_1(\Lambda_0)) \cdots) = \Lambda'$$

The derivation of the dialect Λ' from Λ_0 can be constructed by following the APOs in the sequence of their application.

We propose the following functional explanation for the change of BVI.

> A sequence of APOs, Θ is preferred if $\Theta(\Lambda_0)$ has some functional benefit over Λ_0. Thus, all the modern Bengali dialects have some functional advantage over the classical dialect.

We would like to re-emphasize the term "some functional benefit", because it is quite possible that the modern dialects are inferior to Λ_0 with respect to "some other functional pressures".

Note that the above hypothesis is silent about the precise nature of the functional benefits, which we shall try to formulate in the next section. It only states that under certain realistic formulation of the functional forces as constraints and objectives, the present dialects of Bengali (and most probably classical Bengali as well) are expected to lie on the Pareto-optimal front of this constrained multi-objective optimization problem (see Fig. 4.1 and Sec. 4.1). We do not expect to see, however, a single optimum dialect that has advantages over all other dialects with respect to all the functional forces.

In order to validate the aforementioned hypothesis, we carry out a multi-objective and multi-constrained optimization of the possible dialectal forms of Bengali, thereby obtaining the Pareto-optimal front of the possible optimal dialects. We obtain the Pareto-optimal front using multi-objective genetic algorithm (MOGA) for various choices of functional objectives and constraints. The MOGA model is described in the subsequent sections.

7.2 The MOGA Model

Genetic algorithm (GA) is a special type of evolutionary algorithm that tries to mimic the biological evolution of species to compute optimal solutions for a problem. In a single objective GA, the objective is specified as a fitness function, which, without loss of generality, can be assumed to be minimizing in nature. A possible solution is represented as an *individual*, which has a *genotype* and a *phenotype*. A set of individuals constitute a population.

The phenotype of an individual is nothing but the solution that it represents, whereas the genotype is an underlying representation, over which the genetic operators act. The fitness function is defined for the phenotype, and there is a unique mapping from the genotype to the phenotype, though the converse may not be true.

There are three basic operators that act on the individuals: crossover, mutation and selection. Crossover refers to fragments of chromosomes (i.e., genotypes) being exchanged between two individuals. Mutation refers to a random change in a single gene (a part of the chromosome) and selection is the process of choosing a set of individuals that represent the next generation of the population. The selection process takes into account the fitness of an individual, which is computed using the fitness function. There are several variants of GAs differing mainly in the implementation of the selection process.

The output of a GA is a population, i.e., a set of solutions. Although it is not possible to provide a theoretical upper bound on the number of generations required for convergence to the optimum solution or even to a certain approximation of the optimum solution, GAs are known to perform very well (i.e., get quite close to the optimal solution) and converge quite fast for several real life problems.

In a multi-objective multi-constrained optimization problem, usually there is set of solutions that are non-dominated, rather than a single best solution. The aim of a multi-objective optimization problem is, therefore, to find out the non-dominated or Pareto-optimal solutions, which lie on the Pareto-optimal front. There are several techniques for multi-objective optimization. However, in the absence of any prior knowledge regarding the nature of the Pareto-optimal front, MOGAs are the best bets.

The basic challenge in formulating a problem in the MOGA framework is the appropriate choice of the genotype, so that the genetic operators such as crossover and mutation make sense. The rest of this section presents a MOGA formulation for the functional explanation of the change in BVI.

7.2.1 Phenotype and Genotype

We define the *phenotype* of a dialect d to be the lexicon of the dialect, Λ_d consisting of the 28 inflected forms of the root verb kar. This choice of phenotype is obvious because, at the end of the optimization process, we would like to obtain the Pareto-optimal dialects of Bengali and compare them with their real counterparts.

The *genotype* of a dialect d can also be defined as Λ_d, where the word forms are the genes. However, for such a choice of genotype, crossover and mutation lead to counter-intuitive results. For example, mutation would affect only a single word in the lexicon, which is against the *regularity* hypothesis of sound change. Similarly, exchanging a set of words between a pair of lexica, as crossover would lead to, seems insensible.

Therefore, considering the basic properties of sound change as well as the genetic operators used in MOGA, we define a chromosome (and thus the genotype) as a sequence of APOs. The salient features of a chromosome are described below.

- *Gene*: A gene is defined as an APO. Since in order to implement the MOGA, every gene (read APO) must be mapped to a number, we have chosen an 8-bit binary representation for a gene. This allows us to specify 256 distinct genes or APOs. However, for reasons described below, we use the first bit of a gene to denote whether the gene (i.e., the APO) is active (the bit is set to 1) or not. Thus, essentially we are left with 128 distinct choices for APOs. Since the number of words in the lexicon is only 28, the APOs for Del, Asm and Met are limited, even after accounting for the various contexts in which an APO is applicable. Nevertheless, there are numerous choices for Mut. To restrain the possible repertoire of APOs to 128, we avoided any APO related to the mutation of consonants. This allowed us to design a comprehensive set of APOs that are applicable on the classical Bengali lexicon or its modified versions.

- *Chromosome*: A chromosome is a sequence of g genes. Since the number of APOs invoked during the course of language change is variable, all the

g genes need not be fired. Therefore, the first bit of every gene is used for specifying the status of the gene. If the bit is set to one, the corresponding APO is active, otherwise the APO is assumed to be inactive. Note that the genes are ordered in a chromosome.

- *Genotype to phenotype mapping*: For a given chromosome, let the set of active APOs in sequence be $\theta_1, \theta_2, \cdots, \theta_r$. Then the phenotype corresponding to this chromosome is the lexicon $\Lambda' = \theta_r(\cdots \theta_2(\theta_1(\Lambda_0)) \cdots)$. In other words, the phenotype is the lexicon obtained by successive application of the active APOs on the chromosome on the lexicon of classical Bengali.

The concepts of gene, chromosome and the mapping from genotype to the phenotype are illustrated in Fig. 7.1. It is easy to see that the regularity hypothesis regarding the sound change holds good for the aforementioned choice of genotype. Moreover, according to this formulation, a chromosome not only models a dialect, but also the steps of its evolution from the classical forms.

In the next subsection, we discuss the significance of the genetic operators for the above choice of genotype in the context of language change.

7.2.2 Significance of Crossover and Mutation

Crossover and mutation are two basic genetic operators used in a GA. Figs. 7.2 and 7.3 graphically illustrate these processes. During crossover between two chromosomes (known as parents), certain fragment(s) of the chromosomes are exchanged leading to the formation of two new chromosomes (referred to as children). In the present context, a crossover can be interpreted as follows. Suppose there are two different sequences of APOs, representing two possible courses of language change. A crossover between these two courses of change gives rise to two different, possibly new, courses of change, where some part of the course of the change comes from one of the parents and rest of it is from another parent.

Thus, the resultant dialects (i.e., the phenotypes) of a crossover reflect a situation where there is a sudden change or bifurcation in the course of change

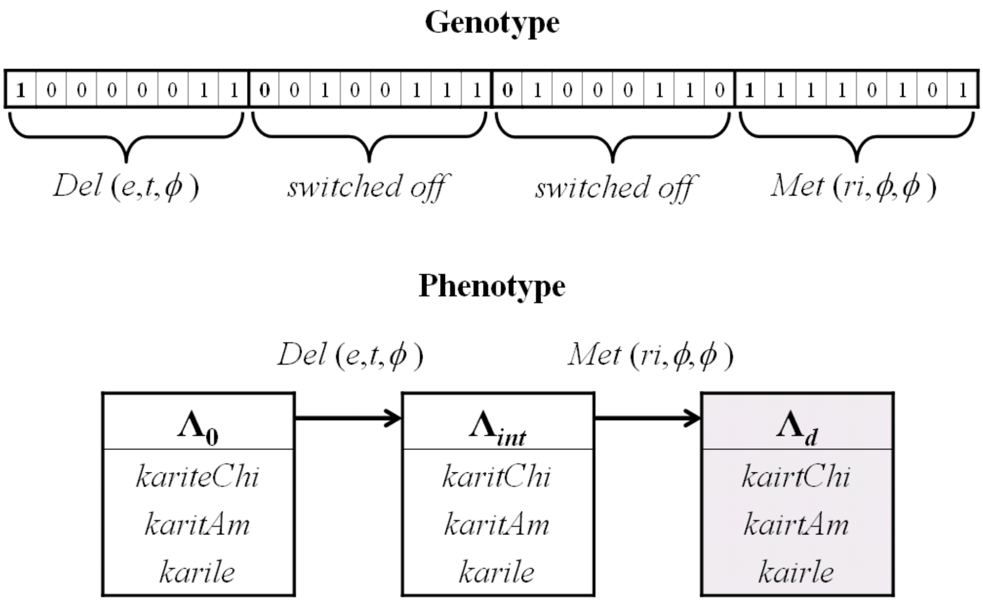


Figure 7.1: Schematic of genotype and phenotype. The chromosome shown here has 4 genes, two of which are switched off, because the first bit (in bold) is 0. The APOs corresponding to the genes are also shown. The mapping from genotype to phenotype is illustrated in the bottom, where the lexicon has only 3 forms. The final phenotype for the gene shown is the shaded lexicon Λ_d.

of the language. Note that the site(s) of a crossover can be in the middle of a gene, rather than its ends (as shown in Fig. 7.2). In such a case, the structure of the particular genes might change considerably giving rise to new genes that correspond to altogether different APOs. Nevertheless, the rest of the genes are unaffected as far as their internal structure is concerned.

The process of mutation refers to a random change in the chromosome. Since according to the current description of the genotype, a chromosome is a binary string, mutation maps a 0 to 1 and vice versa. In the current context, there are three possible contexts and consequent effects of mutations, as enlisted below.

- If the mutation affects the first bit of a gene (odds are 1/8), then it turns

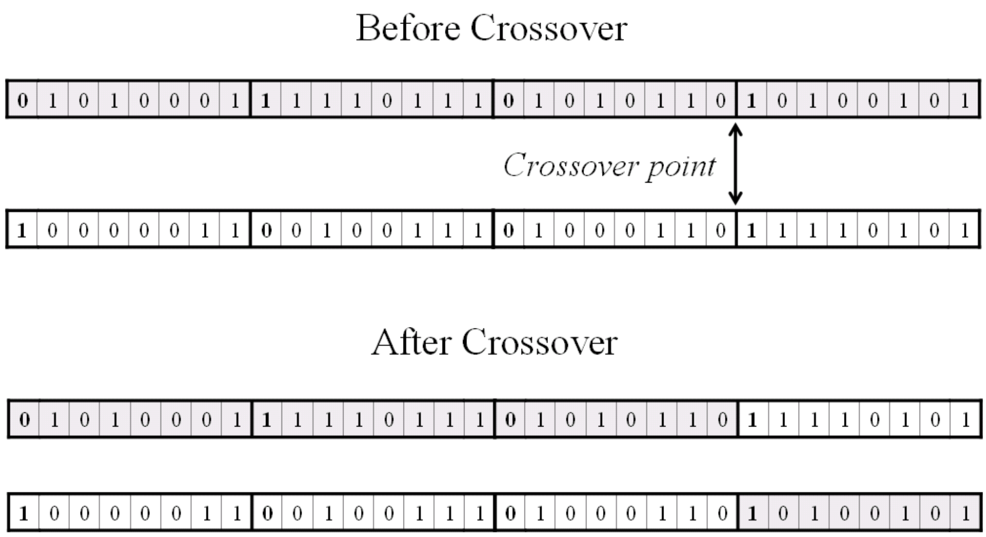


Figure 7.2: Crossover between two chromosomes

on (off) a gene, which was earlier switched off (on). Thus, such a mutation leads to the introduction (removal) of an APO from the course of change. This phenomenon is illustrated in Fig. 7.3 for the first and the second gene.

- If the mutation affects any other bit of a gene that is turned off (odds are 7/16), then there is no visible effect on the course of language change, and consequently the phenotype.

- However, when the mutation affects one of the non-first bits of a gene that is turned on (odds are 7/16), the APO corresponding to that location of the chromosome changes. This happens for the last gene in Fig. 7.3. It is counterintuitive that a case of deletion, say of t, suddenly changes to the metathesis of a sequence, say ri. However, it is quite possible in the real world, that an unconditioned metathesis of ri becomes conditioned over a t in the right (as shown in Fig. 7.3). In other words, generalization or specialization of the context of a phonological change is a common phenomena observed in the nature. Therefore, we assign the gene codes in such a way that APOs with same function, but conditioned over different

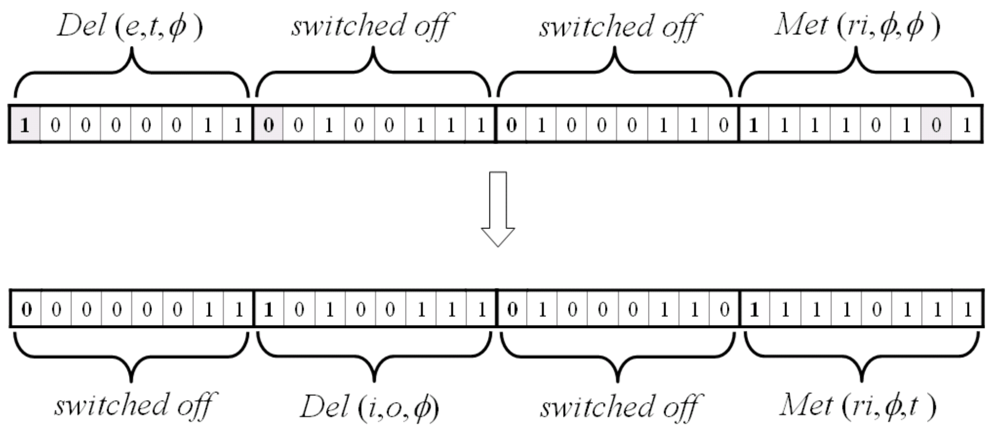


Figure 7.3: Mutation in a chromosome may switch on or off a gene or change the context of application of an APO. The gray cells indicate the sites for mutation.

contexts vary only over one bit-position. Nevertheless, it is impossible to come up with a coding scheme where mutation over a single bit only leads to the generalization or specialization of the context[1]. As a result, there is small probability (5/16 to be precise), when mutation would transform an APO to a completely unrelated one.

It is worth mentioning that since we are using MOGA as a multi-objective optimization tool, it is really not necessary to provide reasonable physical interpretations of the genetic operators. Nevertheless, such an interpretation helps us visualize the process of optimization to a good extent.

7.2.3 Objectives and Constraints

The formulation of the objective functions and constraints is crucial to the model because the linguistic plausibility, computational tractability and the results of the model are overtly dependent on them. We shall define here three

[1]We enumerate the 128 binary representations under a 7-bit gray code scheme. Successive bit strings are assigned similar APOs, so that there are two specific bit-positions in the representation of any APO, where mutation will result in similar APOs. However, for the rest of the five bit-positions, the aforementioned condition does not hold true.

basic objectives of ease of articulation, perceptual contrast and learnability, which can be expressed either as objective functions or constraints or both.

Several models have been proposed in the past for estimating the articulatory effort and perceptual distance between phonemes and/or syllables. See Chapters 2, 5 and 7 of (Boersma 1998) and references therein for an overview of the models estimating articulatory effort, and Chapters 3, 4 and 8 of the same for perception models. Since here we are interested in modeling the effort and contrast of the whole lexicon rather than a syllable, we choose to work with simpler formulations similar to those described in (Redford et al. 1998; Redford 1999; Redford and Diehl 1999; Redford et al. 2001). Nevertheless, the objective of the models presented in Redford et al. (1998; 2001), i.e., the emergence of universal syllable typologies, is different from ours because we are interested in emergence of "real dialects" rather than any arbitrary dialect with some desired property. Moreover, unlike Redford et al. (1998; 2001), which model the problem as a single objective optimization, we model the problem as a multi-objective optimization.

Estimation of learnability of a lexicon is difficult and debatable. Here we assume that a regular pattern is easier to learn and define learnability in terms of regularity of the lexicon. It must be emphasized, however, that it is easy and straightforward to extend the MOGA model for more complex and realistic estimates of these functions. Moreover, as we shall see shortly, despite the crude estimates of effort, contrast and learnability, the MOGA model yields encouraging results. Also note that we define the objective functions to be minimizing in nature as is required by the NSGA II (Deb et al. 2002) algorithm.

f_e: Articulatory Effort

Articulatory effort of a lexicon Λ is a positive real number that gives an estimate of the effort required to articulate the words in Λ in some unit. We define the sum of the effort required to pronounce the words in the lexicon as the effort associated with the lexicon. Thus, if f_e denotes the effort function, then

$$f_e(\Lambda) = \sum_{w \in \Lambda} f_e(w) \tag{7.1}$$

Note that the above definition presumes that all the words in the lexicon are used equally frequently. Furthermore, normalization with respect to the number of words is not necessary as all the lexica under consideration have the same number of words.

The term $f_e(w)$ depends on three parameters: 1) $f_{e1}(w)$ – the length of w in terms of phonemes, 2) $f_{e2}(w)$ – the structure of the syllables, and 3) $f_{e3}(w)$ – the features of adjacent phonemes, as it controls the effort spent in co-articulation or equivalently the phoneme to phoneme transitions. We define $f_e(w)$ to be a weighted sum of these three.

$$f_e(w) = \alpha_1 f_{e1}(w) + \alpha_2 f_{e2}(w) + \alpha_3 f_{e3}(w) \tag{7.2}$$

Here, α_1, α_2 and α_3 are the relative weights. Note that it is not necessary to constrain the sum of the three weights to 1, because we are defining f_e in an arbitrary unit. Initial experimentation with different combinations of the weights revealed that the following assignment leads to comparable estimates of the different effort parameters, and thereby, realistic dialectal patterns in the model.

$$\alpha_1 = 1, \quad \alpha_2 = 1, \quad \alpha_3 = 0.1$$

The value of f_{e1} is simply the length of the word, that is

$$f_{e1}(w) = |w| \tag{7.3}$$

Suppose $\psi = \sigma_1\sigma_2\cdots\sigma_k$ is the usual syllabification of w, where the usual or optimal syllabification for Bengali is defined similar to that of Hindi as described in Sec. 4.2.2. Then, we define f_{e2} as follows.

$$f_{e2}(w) = \sum_{i=1}^{k} hr(\sigma_i) \tag{7.4}$$

$hr(\sigma)$ measures the hardness of the syllable σ and is a function of the syllable structure of σ (Def. 4.8). The values of $hr(\sigma)$ for different syllable structures are enumerated in Table 7.4. The values have no absolute significance and are indicative of only the relative order of hardness. The values are arrived at by observing the frequency distribution of the different syllable types in Bengali.

$CVMap(\sigma)$	$hr(\sigma)$	$CVMap(\sigma)$	$hr(\sigma)$
CV	0	CVV	4
CVC	1	$CCVC$	4
V	2	$CCVV$	5
VC	3	VVC	6
CCV	3	$CVVC$	6

Table 7.4: The hardness of different syllable structures

Note that the definition of $hr(\sigma)$ differs from that of $H_\sigma(\sigma)$ as described in Def. 4.10. However, it can be shown that the strategy for optimal syllabification, which is presented in the proof of Theorem 4.1, does not change for the new definition of hardness provided in Table 7.4.

The effort spent in transitions, measured by f_{e3}, is dependent on several factors, a detailed modeling of which is beyond the scope of the present work. Since *vowel height assimilation* is the primary co-articulation phenomenon observed across the dialects of Bengali, we define f_{e3} so as to model only the effort required due to the difference in the heights of the adjacent vowels.

Let there be n vowels in w represented by V_i, where $1 \leq i \leq n$. Then f_{e3} is defined by the following equation.

$$f_{e3}(w) = \sum_{i=1}^{n-1} |ht(V_i) - ht(V_{i+1})| \tag{7.5}$$

The function $ht(V_i)$ is the tongue height associated with the vowel v_i. The value of the function $height(v_i)$ for different vowels is enumerated in Table 7.5. Again note that the values are indicative of the ordering of the vowels with respect to tongue height, and do not reflect the absolute height of the tongue in any sense.

Table7.6 illustrates the process of computation of the articulatory effort for a lexicon $\Lambda = \{kara, karite, koriAChi\}$.

Vowel V	$ht(V)$
A	0
a	1
e	2
o	2
i	3

Table 7.5: The tongue height associated with the different vowels of Bengali. The values are shown only for the vowels that are there in the lexicon of the verb forms and not for all the vowels of Bengali.

Partial estimates	Λ		
	ka-ra	*ka-ri-te*	*ko-riA-Chi*
$f_{e1}(w)$	4	6	7
$f_{e2}(w)$	$0+0=0$	$0+0+0=0$	$0+4+0=4$
$f_{e3}(w)$	$\|ht(a)-ht(a)\|$ $=0$	$\|ht(a)-ht(i)\|$ $+\|ht(i)-ht(e)\|$ $=2+1=3$	$\|ht(o)-ht(i)\|$ $+\|ht(i)-ht(A)\|$ $+\|ht(A)-ht(i)\|$ $=1+3+3=7$
$f_e(w)$	$4+0+0.1\times 0$ $=4$	$6+0+0.1\times 3$ $=6.3$	$7+0+0.1\times 7$ $=7.7$
$f_e(\Lambda)$	$4+6.3+7.7=18$		

Table 7.6: Computation of effort for an example lexicon

f_d: Acoustic Distinctiveness

We define the acoustic distinctiveness between two words w_i and w_j as the edit distance between them, which is denoted as $ed(w_i, w_j)$. While computing the edit distance, the cost of insertion and deletion of any phoneme is assumed to be 1; the cost of substitution of a vowel (consonant) for a vowel (consonant) is also 1, whereas that of a vowel (consonant) for a consonant (vowel) is 2, irrespective of the phonemes being compared. Since languages are expected to increase the acoustic distinctiveness between the words, we define a minimizing objective function f_d over a lexicon Λ as the sum of the inverse of the edit distance between all pair of words in Λ.

$$f_d(\Lambda) = \sum_{ij, i \neq j} ed(w_i, w_j)^{-1} \tag{7.6}$$

If for any pair of words w_i and w_j, $ed(w_i, w_j) = 0$, we redefine $ed(w_i, w_j)^{-1}$ as 20 (a large penalty).

C_d: Distinctiveness constraint

We say that a lexicon Λ violates the acoustic distinctiveness constraint C_d, if there are more than two pairs of identical words in Λ. The motivation behind the constraint is as follows. The functions f_e and f_d are conflicting in nature. Theoretically, one can obtain a lexicon Λ by deleting all the phonemes, such that $f_e(\Lambda) = 0$ and $f_d(\Lambda) = 10M(M-1)$, where M is the number of words in Λ. Such a lexicon is a part of the Pareto-optimal front, because it has the minimum possible value of f_e, but clearly, it is absurd and uninteresting. Thus, the constraint C_d has been introduced in order to restrict the search space of the MOGA model to realistic lexica, thereby reducing the computation time.

The allowance of two non-distinct pairs per lexicon is motivated by the fact that in the observed dialects, BVIs have one non-distinct pair. We allow one more to explore further possibilities.

C_p: Phonotactic constraints

A lexicon Λ is said to violate the constraint C_p if any of the words in Λ violates the phonotactic constraints of Bengali. As described in Sec. 4.2.1, the PCs are defined at the level of syllable onsets and codas and therefore, syllabification is a preprocessing step before evaluation of C_p. Since no dialect of Bengali is known to allow a complex coda (i.e., more than one consonant in the coda), the allowable codas are null (no consonant) or a single consonant. However, some of the consonant clusters comprising of upto two consonants (i.e., CC) are allowable in the onset. We manually enumerate the legality of the complex onsets that might emerge due to the process of change for the model under consideration.

f_r and C_r: Regularity

Although learnability is a complex notion, one can safely equate the learnability of a system to the regularity of the patterns within the system. In fact, in the context of morphology, it has been observed that the so called *learning bottleneck* has a regularizing effect on the morphological structures, thereby leaving out only the most frequently used roots to behave irregularly (Hare and Elman 1995; Kirby 2001; Kirby 2002).

In the present context, we define the regularity of the verb forms in a lexicon as the predictability of the inflectional suffix on the basis of the morphological attributes. For example, the inflections pertaining to the past tense have the phoneme l, and those pertaining to the continuous aspect have the phoneme Ch; similarly, n at the end is indicative of formal person. This fact can be restated as follows:

> If a lexicon is regular, two word forms are expected to be more similar if they share a larger number of morphological attributes.

Brighton et al. (2005) discuss the use of Pearson correlation between phonological edit distance and semantic/morphological hamming distance measures as a metric for learnability. On a similar note, we define the regularity function

f_r as follows: For two words $w_i, w_j \in \Lambda$, the (dis)similarity between them is given by $ed(w_i, w_j)$. Let $ma(w_i, w_j)$ be the number of morphological attributes shared by w_i and w_j. Since there are three distinct features – tense, aspect and person, the value of $ma(w_i, w_j)$ can be 0, 1 or 2. We define the regularity of Λ, $f_r(\Lambda)$, as the *Pearson correlation coefficient* between $ed(w_i, w_j)$ and $ma(w_i, w_j)$ for all pairs of words in Λ. Note that for a regular lexicon, $ed(w_i, w_j)$ decreases with an increase in $ma(w_i, w_j)$. Therefore, $f_r(\Lambda)$ is negative for a regular lexicon and 0 or positive for an irregular one. In other words, $f_r(\Lambda)$ is also a minimizing objective function.

Nevertheless, $f_r(\Lambda)$ is a very crude measure of regularity. Rather than formulating it as an objective function, we can formulate a regularity constraint C_r, such that a lexicon Λ violates C_r if $f_r(\Lambda) > minreg$, where $minreg$ is some constant greater than -1.

C_u: Constraint on Useless Genes

The functions and constraints described till now are computed for the lexicon, i.e., the phenotype. C_u is a constraint, which is defined on the genotype and has no relation to the functional forces as such. Rather, the constraint C_u helps to prune the search space by removing the chromosomes that encode useless information. Consider, for example, a chromosome that has $\theta_1\theta_2 \cdots \theta_r$ as the active APOs, such that after the application of the APOs θ_1 to θ_i on Λ_0, an intermediate lexicon Λ_{int} is obtained. Suppose that θ_{i+1} is inapplicable to any of the words in Λ_{int}. In other words, despite being switched on, the gene corresponding to θ_{i+1} is useless. We shall refer to them as *useless genes*.

There can be a lot of chromosomes that map to the same phenotype, but with several useless genes on it. If the phenotype they encode for is a non-dominated solution for the model, then the chromosomes will be selected as "good" individuals, generation after generation. On the contrary, these chromosomes do not provide us with any extra information and neither do they lead to better individuals, which would not have been obtainable from other chromosomes. Thus, in order to prune such useless chromosomes we introduce the constraint C_a. A chromosome is said to violate C_u, if it has a *useless gene*.

7.3 Experiments and Observations

In the previous section, we have presented several objective functions and constraints. In this section, we describe several experimental setups developed using different combinations of these functions. We begin with a brief sketch of NSGA II, the package that has been used for optimization in the MOGA framework. This is followed by the description of some general experiments designed to determine the number of generations required for convergence, population size and length of the chromosome g. Next, the experiments conducted for different combinations of objectives and constraints are discussed at length, and finally we attempt to design some objective evaluation metrics for the problem at hand.

7.3.1 Experimental Setup: NSGA - II

We use the Non-dominated Sorting GA-II or NSGA-II (Deb et al. 2002), which is a multi-objective, multi-constraint and elitist GA. The NSGA-II package is available online. The salient features of the package and their utilizations in the present context are enlisted below.

- NSGA-II is implemented in C, runs on the Linux platform, and uses GNUPlot for visualization.
- It can deal with both binary and real-valued genes. For the problem at hand, we have only one binary variable of length $8g$ bits.
- The population size, number of generations, number of objective functions and constraints are specified as inputs.
- The probability of crossover and mutation are also specified a priori by the user. In the present case, the probability of crossover has been set to 0.9 and that of mutation has been set to 0.1 for all the experiments. These values are arrived at through some amount of parameter space exploration.

- The package requires the specification of the objective functions and constraints in the form of modules in C. As stated earlier, the objectives are assumed to be minimizing and the constraints are violated if they evaluate to a negative value. Note that in the present context, computation of the objective functions and constraints require the genotype to phenotype mapping (i.e., application or the APOs or rewrite rules on Λ_0) and syllabification of the resulting lexicon.

- The output of NSGA-II is the set of best solutions (genotypes) obtained after running the algorithm till the number of generations specified. The algorithm also outputs the fitness of the chromosomes at the end of the process. The fitness is the values of the objective functions, which can be plotted using GNUPlot or otherwise to visualize the Pareto-optimal front.

- The package also provides option for visualizing the fitness plot of the individuals after every k generations (where k is an input parameter).

Thus, at the end of the optimization process, we obtain a set of Pareto-optimal dialects and a corresponding Pareto-optimal front. Fig. 7.4 shows the Pareto-optimal front obtained at the end of a typical run of NSGA-II. f_e and f_d are plotted on the x and y axes respectively. A dot on the $x - y$ plane corresponds to an individual and represents its fitness values. As one moves from the right to the left over the Pareto-optimal front, the dialects tend to loose distinctiveness between the verb forms, but in the process the articulatory effort is minimized.

Note that as the objectives are minimizing in nature, the area on the plot below and left of the Pareto-optimal front represents impossible languages, whereas the area to the right and top of the curve shows unstable or suboptimal languages. This is opposite of what has been shown in Fig. 4.1, because the objectives f_1 and f_2 in Fig. 4.1 are both maximizing in nature.

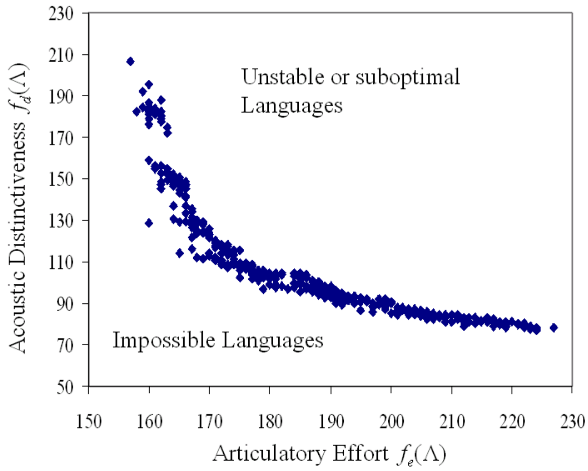


Figure 7.4: Pareto-optimal front obtained from a typical run of NSGA-II

7.3.2 Population, Generation and Chromosome length

The optimality of the solutions obtained through GA is dependent on the population size (larger population ensures variability) and the number of generations for which the algorithm is run (large number of generation ensures closer approximation of the Pareto-optimal front). However, the time complexity of NSGA II is quadratic in the population size and linear in the number of generations. Our aim is to strike the best trade-off between population size, generations and the quality of the Pareto-optimal front.

For this purpose, we have conducted a pilot experiment with different population sizes (100, 500, 1000, 2000, 5000) and number of generations (100, 500, 1000). The value of the other parameters for all the experiments were held fixed at $g = 7$, objective functions: f_e and f_d, no constraints, probability of crossover $= 0.9$, probability of mutation $= 0.1$, $|\Lambda| = 18$. The results of the experiments are shown in Fig. 7.5

We observe that the Pareto-optimal front obtained after 100 generations is not satisfactory, but the one obtained after 500 generations are reasonably good when the population size is more than or equal to 500. Therefore, for most of the experiments that we have conducted, both population and number of generations have been set to 500.

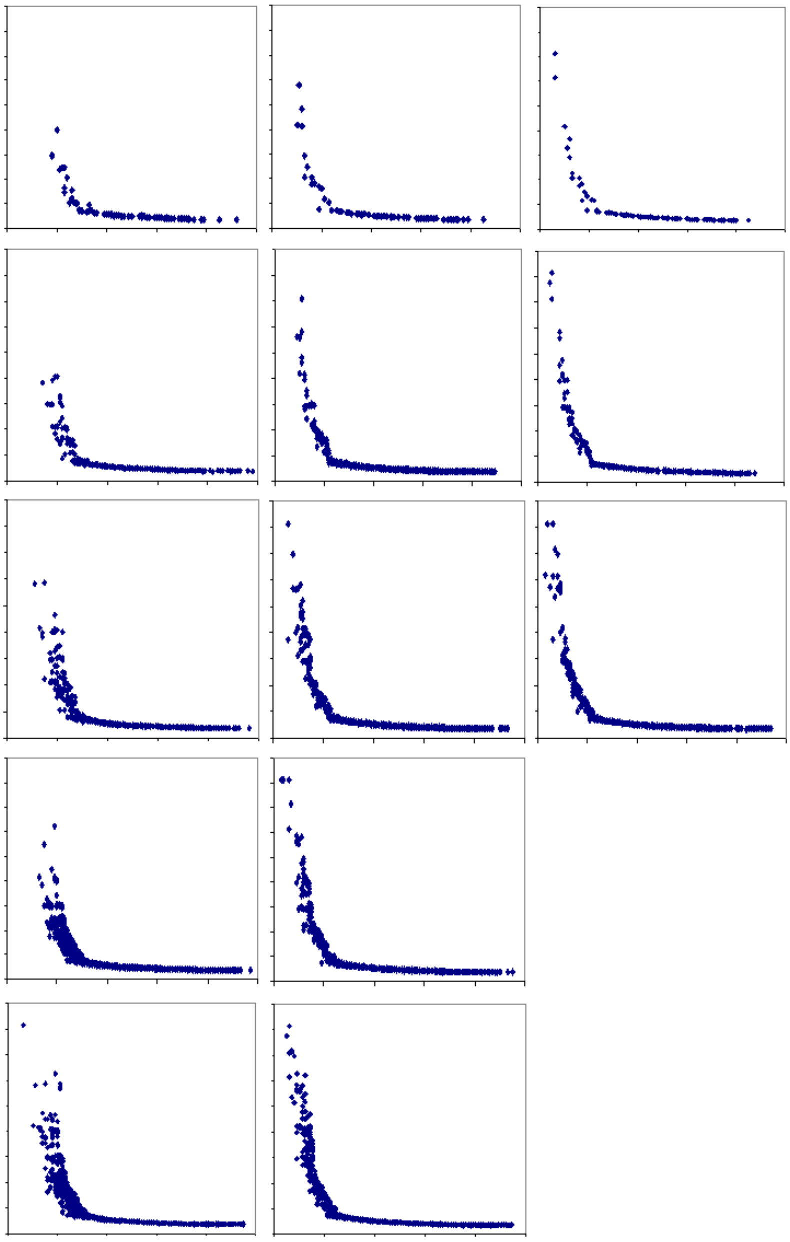


Figure 7.5: The Pareto-optimal front for the pilot experiments. The values of f_e and f_d are plotted on x and y axes respectively. The plots in the rows, from top to bottom, are for population sizes 100, 500, 1000, 2000 and 5000. The plots in the columns, from left to right, are for generations 100, 500 and 1000 respectively. Experiments could not be run for 1000 generations with population sizes of 2000 and 5000 as the time required is very large. For other details, see the text.

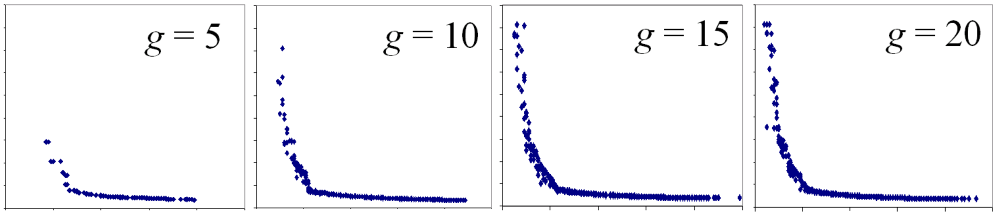


Figure 7.6: Pareto-optimal fronts for different lengths of the chromosome – g. The values of f_e and f_d are plotted on x and y axes respectively. Population = 500, generation = 500

We also carry out experiments to determine a suitable value for g, the number of genes on a chromosome. The results are presented in Fig. 7.6. The values of the other parameters are same as in the previous experiment. We observe that as g increases from 5 to 15, there is significant improvement in the Pareto-optimal front. However, we observe no further improvement of the front as g increases to 20. Thus, we infer that 15 is a suitable value for g, with which we conduct the other experiments.

Since the probability of any gene being switched off is 0.5, the expected number of active APOs on a chromosome with 15 genes is 7.5. It is interesting to note that this value is almost equal to the number of APOs required (7 to be precise) for derivation of the SCB verb forms.

7.3.3 Observations for Different Models

We formulate different MOGA models by various combinations of the objective functions and constraints described in Sec. 7.2.3 and closely inspect the emergent dialects. In fact, the motivation behind the incorporation of new objective or constraints comes from the observations made on the previous models. We start with the simplest possible model, which we shall refer to as MOGA Model 1, and refine it subsequently to construct five other models. Table 7.7 shows the set of objectives and constraints incorporated in each of these six models. Apart from the different combinations of the objectives and constraints, the models also differ in the size of the lexicon, length of the gene, and the number of APOs in the repertoire.

Model No.	$\|\Lambda_0\|$	g	APOs	f_e	f_d	f_r	C_p	C_d	C_r	C_u
1	18	10	64	$\surd$	$\surd$					
2	18	10	64	$\surd$	$\surd$		$\surd$	$\surd$		
3	18	10	64	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$		
4	18	10	64	$\surd$	$\surd$		$\surd$	$\surd$	$\surd$	
5	28	10	128	$\surd$	$\surd$		$\surd$	$\surd$	$\surd$	
6	28	15	128	$\surd$	$\surd$		$\surd$	$\surd$	$\surd$	$\surd$

Table 7.7: Description of the six MOGA Models. APOs refer to the total number of APOs in the repertoire. A $\surd$ indicates the presence of a particular objective or constraint in the model.

Each of the models and observations made from them are described below. In order to compare the Pareto-optimal fronts obtained for these models, we have appropriately normalized the values of f_e and f_d, as in Models 4 and 5, $|\Lambda| = 28$, whereas for the other four models, $|\Lambda| = 18$. The details of the emergent dialects for each of the models are reported in Appendix E. However, to have a feel of the realistic dialects that emerge in the system, in Table 7.8 we present some of the verb forms from MOGA Model 6 (described subsequently) vis-à-vis their real counterparts. Note that the phone $/Ch/$ in SCB or classical Bengali corresponds to $/s/$ in ACB. However, this change from $/Ch/$ to $/s/$ is not reflected in the emergent dialect because we have deliberately excluded any APO related to the mutation of consonants in the repertoire of genes.

MOGA Model 1

The simplest of the MOGA models, Model 1 has only two objective functions and no constraints. The Pareto-optimal front obtained for this model is shown in Fig. 7.7. Manual inspection of the dialects emerging in Model 1 reveals that although some of them resembles their real world counterparts loosely, quite a big fraction of the dialects are implausible because – 1) they violate the PC of Bengali, or 2) several forms are indistinguishable. See Appendix E.2.1 for examples of realistic as well as unrealistic dialects emerging in Model 1.

The fitness of the four real dialects are also plotted in Fig. 7.7, all of which lie

Classical Bengali	SCB		ACB	
	Real	Emergent	Real	Emergent
kari	*kori*	*kor*	*kori*	*kori*
kara	*karo*	*kora*	*kara*	*kora*
kare	*kare*	*kore*	*kare*	*kore*
karen	*karen*	*koren*	*karen*	*koren*
kariteChi	*korChi*	*karChi*	*kartAsi*	*karteChi*
kariteCha	*korCho*	*karCha*	*kartAsa*	*karteCha*
kariteChe	*korChe*	*karChe*	*kartAse*	*karteChe*
kariteChen	*korChen*	*karChen*	*kartAsen*	*karteChen*
kariAChi	*koreChi*	*korChi*	*korsi*	*koriChi*
kariACha	*koreCho*	*korCha*	*karsa*	*koriCha*
kariAChe	*koreChe*	*korChe*	*karse*	*koriChe*
kariAChen	*koreChen*	*korChen*	*karsen*	*koriChen*
karilAm	*korlAm*	*korlAm*	*karlAm*	*karlAm*
karile	*korle*	*korle*	*karlA*	*karle*
karila	*korlo*	*korla*	*karla*	*karla*
karilen	*korlen*	*korlen*	*karlen*	*karlen*
kariteChilAm	*korChilAm*	*karChilAm*	*kartAslAm*	*karteChilAm*
kariteChile	*korChile*	*karChile*	*kartAslA*	*karteChile*
kariteChila	*korChilo*	*karChila*	*kartAsla*	*karteChila*
kariteChilen	*korChilen*	*karChilen*	*kartAslen*	*karteChilen*

Table 7.8: Few example verb forms from the lexica of two real and emergent dialects. The emergent dialects are chosen from the Pareto-optimal set of MOGA Model 6 on the basis of their resemblance to the real dialects. The corresponding classical Bengali forms are also shown for reference.

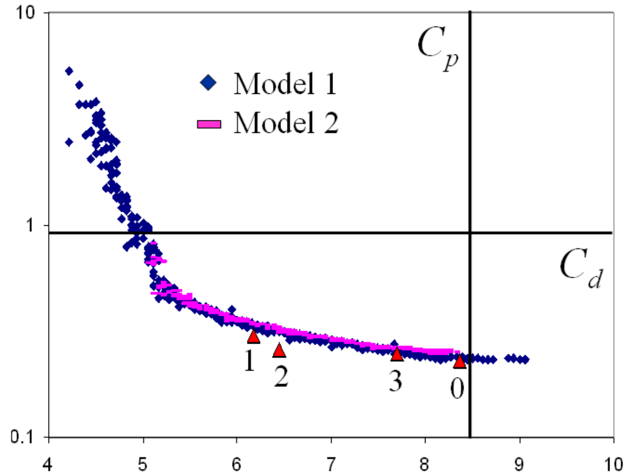


Figure 7.7: Pareto-optimal fronts obtained from MOGA Models 1 and 2. f_e and f_d are plotted on x and y-axes respectively. The triangles show the position of the real dialects in the plot: 0 – Classical Bengali, 1 – SCB, 2 – ACB, 3 – Sylheti. The vertical and horizontal lines marked C_p and C_d respectively, show the effects of these constraints in restraining the extent of the front for Model 2. Note that the y-axis is in logarithmic scale for better visualization.

below the Pareto-optimal front. This clearly reveals that the dialects obtained in Model 1 are suboptimal or inferior to the real dialects.

MOGA Model 2

In Model 2, we incorporate two constraints, C_p and C_d, to alleviate the problems of illegal clusters and indistinguishable forms in the emergent dialects. The Pareto-optimal front for Model 2 is plotted over that of Model 1 in Fig. 7.7. The effects of the constraints are clearly visible from the Pareto-optimal sets of the two models. While C_p restrains the Pareto-optimal front from having dialects with very high effort (shown by the vertical line marked C_p in Fig. 7.7), C_d restrains the front from having dialects with very low acoustic distinctiveness, i.e., high f_d (shown by the vertical line marked C_d in Fig. 7.7). Thus, the two constraints restrain the Pareto-optimal front to the left of the C_p line and below the C_d line.

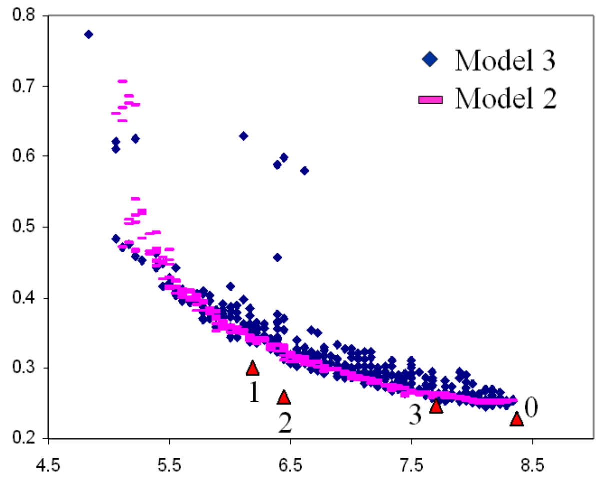


Figure 7.8: Pareto-optimal fronts obtained from MOGA Models 2 and 3. f_e and f_d are plotted on x and y-axes respectively. The triangles show the position of the real dialects in the plot: 0 – Classical Bengali, 1 – SCB, 2 – ACB, 3 – Sylheti.

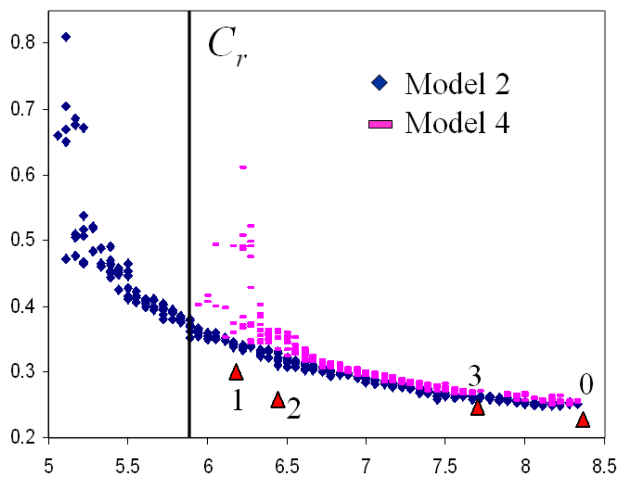


Figure 7.9: Pareto-optimal fronts obtained from MOGA Models 2 and 4. f_e and f_d are plotted on x and y-axes respectively. The triangles show the position of the real dialects in the plot: 0 – Classical Bengali, 1 – SCB, 2 – ACB, 3 – Sylheti.

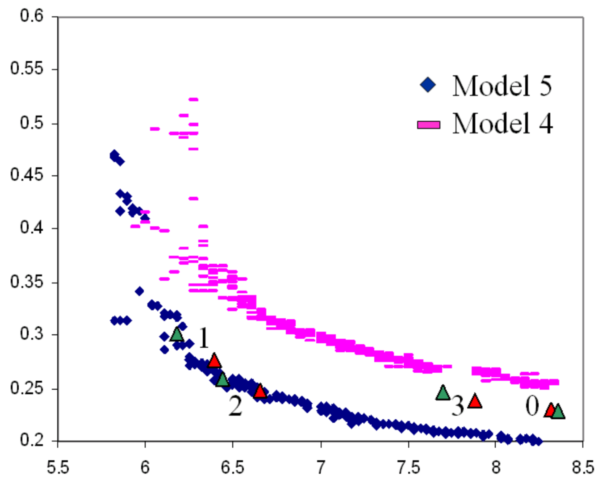


Figure 7.10: Pareto-optimal fronts obtained from MOGA Models 4 and 5. f_e and f_d are plotted on x and y-axes respectively. The triangles show the position of the real dialects in the plot: 0 – Classical Bengali, 1 – SCB, 2 – ACB, 3 – Sylheti (green and red for lexicon of size 18 and 28 respectively).

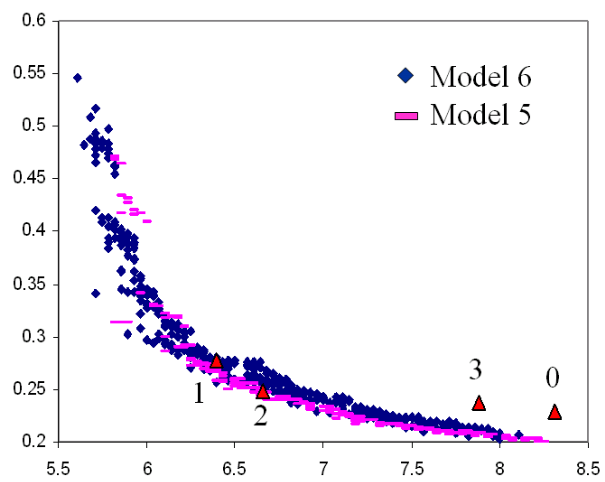


Figure 7.11: Pareto-optimal fronts obtained from MOGA Models 4 and 5. f_e and f_d are plotted on x and y-axes respectively. The triangles show the position of the real dialects in the plot: 0 – Classical Bengali, 1 – SCB, 2 – ACB, 3 – Sylheti

Although the constraints help to eliminate some of the implausible dialects, manual inspection reveals that several dialects emerging in Model 2 are irregular. See Appendix E.2.2 for examples of such dialects.

MOGA Model 3

In order to resolve the problem of irregular dialects, we introduce a new objective function of regularity, f_r, in Model 3. Note that Model 3 has three objectives and therefore, the Pareto-optimal front is a surface in 3-dimensional space, rather than a curve in a 2-dimensional plane. Nevertheless, for the sake of comparison, in Fig. 7.8 we plot the Pareto-optimal set of Model 3 with respect to only two of the objective functions. Although this results in existence of a few dots (i.e., solutions) in the middle of the $f_e - f_d$ plane, away from the Pareto-optimal curve of Model 2, still majority of the solutions flock towards the Pareto-optimal front obtained from Model 2.

This indicates that the new objective f_r plays an insignificant role in the optimization process. In fact, a closer look at the solutions emerging from Model 3 reveals that 90% of the individuals in the Pareto-optimal set has f_r values ranging from -0.39 to -0.84. On the other hand, we also observe that f_r of a reasonably regular dialect should be less than -0.8. Appendix E.2.3 enlists some of the dialects from Model 3.

MOGA Model 4

In Model 4, the objective function f_r has been replaced by the constraint C_r, where a lexicon Λ violates C_r, if $f_r(\Lambda) > -0.8$. Fig. 7.9 compares the Pareto-optimal fronts of Models 4 and 2. We observe that the constraint C_r, in effect, pushes the vertical limb of the front towards right. Indeed, the solutions with very low effort have been found to violate the regularity of the pattern. See Appendix E.2.4 for examples of dialects emerging in Model 4.

MOGA Model 5

In the previous four models, we have successfully restrained the Pareto-optimal front by incorporation of appropriate constraints and objectives, such that the emergent dialects are as plausible as possible. Nonetheless, in all the models, the Pareto-optimal set seems to be inferior to the real dialects. Therefore, we extend the previous models by increasing the size of the lexicon as well as the repertoire of APOs, so that the search space is extended. In Model 5, $\Lambda = 28$ and the number of APOs has been doubled to 128.

Fig. 7.10 compares the Pareto-optimal fronts of Model 4 and 5. The extension of the search space, due to increase in the repertoire of APOs, has pushed the front in Model 5 towards a more optimal configuration. As a result, two of the real dialects, namely SCB and ACB now lie on the front, whereas classical Bengali and Sylheti are suboptimal with respect to the front of Model 5. Some of the best solutions of Model 5 are listed in Appendix E.2.5

MOGA Model 6

In Model 6, we increase the length of the gene from 10 to 15, which in effect, further extends the search space. Initial experiments with this model show that several genes in the emergent solutions are useless, but they clutter the search space and in turn, negatively affect the convergence and variation of the population. Therefore, we also introduce the constraint C_u to eliminate the chromosomes with useless genes.

Fig. 7.11 compares the Pareto-optimal fronts obtained for Model 5 and 6. We observe that the fronts are almost similar, except for the fact that the larger search space now helps the algorithm to find out solutions with very low effort, but bad distinctiveness. A few emergent solutions from Model 6 are listed in Appendix E.2.6.

7.3.4 Objective Evaluation Metrics

Unlike the case of SDH, where Ohala's rule provides a basis for evaluating the emergent patterns of the MAS Models, quantifying the validity of the MOGA

Models is quite a difficult task. In the last subsection, we have compared the dialects obtained for different combinations of constraints and objectives with data from real dialects. Nevertheless, emergence of dialects in the experiments that lie on the Pareto-optimal front and closely resemble one of the modern dialects is hardly a measure of plausibility of the model. This is because we have data from only four real dialects, whereas the number of dialects that belong to the Pareto-optimal set is of the order of a several hundreds.

Thus, in the absence of any concrete and objective evaluation criterion, we have resorted to manual inspection of the dialects in the Pareto-optimal set, during which the plausibility of the dialects have been adjudged on the basis of linguistic features. The seemingly absurd dialects have helped us to further improve the model by incorporating appropriate constraints and/or objectives.

In order to automatically identify the solutions in the Pareto-optimal set that are closest to the real dialects, we introduce two metrics of similarity: one based on the phenotype and another based on the genotype. The first metric, $\overline{sim}_p(\Lambda, \Lambda')$, computes the average dissimilarity between the corresponding word forms in the lexica Λ and Λ' and is defined as follows.

$$\overline{sim}_p(\Lambda, \Lambda') = \frac{1}{|\Lambda|} \sum_{i=1}^{|\Lambda|} ed(w_i, w_i') \tag{7.7}$$

Here w_i and w_i' are the corresponding word forms in Λ and Λ' respectively. The lower the value of $\overline{sim}_p(\Lambda, \Lambda')$, the higher is the similarity between Λ and Λ'.

The second metric, $sim_g(\Theta, \Theta')$, compares two sequences of APOs and is defined as follows:

$$sim_g(\Theta, \Theta') = \sum_{i=1}^{r} \sum_{j=1}^{r'} \delta_{\theta_i, \theta_j'} \tag{7.8}$$

where, $\delta_{\theta_i, \theta_j'}$, is the *Kronecker's delta*, which is 1 if and only if $\theta_i = \theta_j'$. We assume that in Θ or Θ', any APO θ occurs atmost once. Thus, in words, $sim_g(\Theta, \Theta')$ counts the number of APOs shared by the sequences Θ and Θ'. The higher the value of $sim_g(\Theta, \Theta')$, the higher is the similarity between the dialects.

It may be noted that although the two measures are correlated, they are

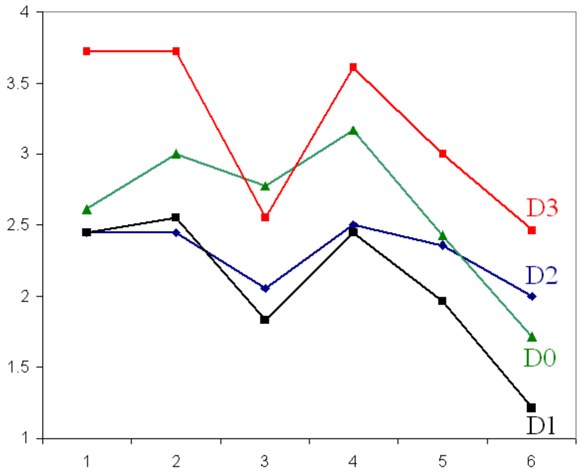


Figure 7.12: Resemblance to real dialects according to the metric $\overline{sim_p}$. The model numbers are shown in the horizontal axis and the value of $\overline{sim_p}$ is plotted along the vertical axis. D0 – Classical Bengali, D1 – SCB, D2 – ACB, D3 – Sylheti

not equivalent because, even a small difference between Θ and Θ' might lead to a large value of $\overline{sim_p}(\Theta(\Lambda_0), \Theta'(\Lambda_0))$.

Fig. 7.12 and 7.13 show the minimum value of $\overline{sim_p}$ and maximum value of sim_g respectively for the emergent dialects in the six MOGA Models, where the comparisons are made with respect to each of the four real dialects. We observe that for both the metrics, dialects closest to the real ones emerge in Model 6. Note that despite the limitations of Model 3 with respect to Model 4, the former is better than the latter as far as the similarity to the four real dialects are concerned.

7.4 Analysis and Interpretations

Several inferences can be drawn from the aforementioned experiments and observations. We discuss at length three such issues, which we deem to be the most significant findings of the MOGA Model.

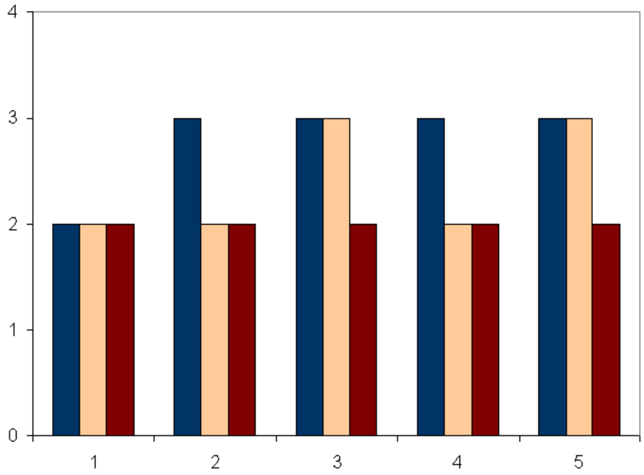


Figure 7.13: Resemblance to real dialects according to the metric sim_g. The model numbers are shown in the horizontal axis and the value of sim_g is shown along the vertical axis. The first second and the third column for every model correspond to SCB, ACB and Sylheti respectively.

The first one is regarding the distribution of real dialects over the Pareto-optimal front. We have observed that the Pareto-optimal front for all the MOGA Models look like rectangular hyperbolas that can be expressed as follows.

$$f_d(\Lambda) = af_e(\Lambda)^{-b} \tag{7.9}$$

We shall not attempt to solve for the values of a and b for the different models, because the values are dependent on the formulation of the objective functions and their precise linguistic significance is unclear. However, from the plots, it seems that both a and b are large. In other words, for all the models, the Pareto-optimal front has a horizontal and a vertical limb, with a sharp angular transition between the two. The limbs are almost parallel to the axis. This means that it is possible to minimize the effort for a lexicon, without compromising with its distinctiveness upto a certain critical value of $f_e(\cdot)$. Once this critical value is crossed, even a very small reduction in the effort seriously hampers the distinctiveness. Similar is the case for $f_d(\cdot)$.

Interestingly, all the real dialects lie on the horizontal limb of the Pareto-optimal front, Classical Bengali being placed at the extreme right. We also note the negative correlation between the value of $f_e(\cdot)$ for the real dialects, and the number of APOs invoked during derivation of these dialects from Classical Bengali. These facts together imply that the natural direction of language change in the case of BVIs has been along the horizontal limb of the Pareto-optimal front, leading to the formation of dialects with higher and higher articulatory ease. Among the four dialects, SCB has the minimum value for $f_e(\cdot)$ and it is positioned on the horizontal limb of the front just before the beginning of the vertical limb (see e.g., Fig. 7.11).

Therefore, it is natural to ask whether there are any real dialects of modern Bengali that lie on the vertical limb of the Pareto-optimal front. And if not, what may be the possible reasons behind their inexistence? In the absence of any comprehensive collection of Bengali dialects, we do not have a clear answer to the above questions. Nevertheless, it may be worthwhile to analyze the emergent dialects of the MOGA Models that lie on the vertical limb. As discussed earlier, in Model 1 the emergent dialects at the extreme regions of the vertical limb severely violate the distinctiveness criterion by featuring repetition of several word forms in the lexicon. The incorporation of the distinctiveness constraint C_d in Model 2 helped us eliminate such implausible dialects (refer to Fig. 7.7). The introduction of the regularity constraint C_r in Model 4 further shifted the vertical limb towards right (see Fig. 7.9). Thus, the vertical limb of the Pareto-optimal front has always featured implausible dialects that have been eradicated by incorporation of suitable constraints.

The emergent dialects of Model 6 with low $f_e(\cdot)$ and high $f_d(\cdot)$ are no exceptions. As shown in Appendix E.2.6, they feature repetition of forms as well as deletion of vital tense/aspect marking affixes (e.g. $/Ch/$). Thus, it is likely that the vertical limb of the Pareto-optimal front does not contain any real dialect of modern Bengali. The reason, perhaps, is the strong linguistic constraints imposed on the lexicon and morphology by the pressure of regularity and distinctiveness, which restrict any event of phonological change that pushes the lexicon upwards towards the vertical limb.

Note that the presensce of a strong distinctiveness constraint also implies that the system of BVIs have redundant phonemes that could be deleted to obtain dialects with smaller $f_e(\cdot)$. However, there seems to be a strong bias towards maintainance of this *redundancy* across all the dialects. It is a well known fact that redundancy is observed in all linguistic systems. It is often refered to as "Linguistic junk" (Lass 1997) that allows the speakers of different dialects to communicate more successfully because of the availability of more number of supplementary clues pertaining to what is being said. Furthermore, through agent-based simulations Livingstone (2003) shows that redundancy helps the agents to easily learn conflicting lexicons. Hence, redundancy seems to be beneficial to the speakers in the sense that it helps them improve their capability to learn a language from conflicting and contradictory evidences. This is perhaps one of the primary reasons for the redundancy in the BVIs of the modern Bengali dialects, and consequently, the absence of any real dialect on the vertical limb of the Pareto-optimal front.

The second inference we would like to draw is regarding the relative merits and demerits of the constraints and objective functions. The nature of the functional forces, as in whether they are binary or real valued functions, has been the subject matter of a long debate in synchronic as well as diachronic linguistics. Note that a binary valued function is equivalent to a constraint in the MOGA Model, whereas a real valued function is similar to the objective functions. While the proponents of Optimality theory claim that linguistic functions are always binary in nature (Prince and Smolensky 1993; Boersma 1997a), several other researchers (e.g., proponents of phonetically based phonology Ohala (1983a; 1990b), functionalists such as Liljencrants and Lindblom (1972) and Schwartz et al. (1997)) have postulated real valued linguistic functions.

In the context of MOGA models, we observe that multi-objective optimization with a set of real-valued objective functions give rise to a large variety of possible dialects, most of which are implausible (Model 1 and 3). Incorporation of suitable constraints however, restricts the set of possibilities to more plausible dialects. Indeed, the aforementioned discussion regarding the concentration of the real dialects on the horizontal limb further points to the fact that distinctiveness also acts as a constraint rather than a real valued function. Of course, it does not follow from the aforementioned facts that linguistic functions are always manifested as constraints. Nonetheless, the fact that the relaxation of the constraints does not beget richer models support the hypothesis, at least in the context of BVIs, that linguistic functions are binary valued in nature.

The third interesting aspect emerging out of the experiments is about the richness of the search space. The sudden shift of the Pareto-optimal front towards the origin, i.e., towards more optimal solutions, in MOGA Model 5 as compared to Models 1 to 4 (refer to Fig. 7.10), is arguably an effect of doubling the set of APOs. Note that, on the other hand, increase in the number of genes on a chromosome does not have any significant effect on the Pareto-optimal set (see Fig. 7.11). In fact, as is apparent from Fig. 7.6, for a given repertoire of APOs, there is an optimum value for g, after which increase in the length of the chromosome do not beget better solutions. Furthermore, increase in g does not shift the Pareto-optimal front towards origin, rather it has a visible effect of extending the vertical limb of the front further up. Nevertheless, the best solutions emerging in Model 6 resemble the real dialects much more closely than those from Model 5 (see Fig. 7.12).

We have not conducted any systematic experiment to study the effect of increase in the size of the repertoire of APOs on the front. However, it is likely that as we increase the size of the repertoire, the front shifts towards more optimal solutions up to a certain point. Clearly, the choice of the repertoire, in effect, defines the shape and extent of the search space or the set of possible courses of language change. Therefore, it follows from the above discussion that in the MOGA model, the appropriate choice of the APOs is as important as formulation of the objectives and constraints. Ideally, the repertoire should contain all possible APOs, but the set of all possible APOs can be infinite. Thus, it becomes necessary to restrict the set of APOs to some selected candidates.

The MOGA model can be interpreted in various ways. For instance, since a sequence of APOs can be equivalently represented using an FST (Kaplan and Kay 1994), the evolving chromosomes can be viewed as evolving FSTs. Although in the specific context of BVIs, FSTs would represent historical derivations, it is not difficult to imagine a case where the FSTs stand for synchronic grammar rules. Stating differently, the present MOGA framework can as well be thought of as a competition between synchronic grammars. Another interpretation could be that each individual in one instantiation of the MOGA model represents a *meme* or collection of memes (Dawkins 1976; Blackmore 1999), and the GA mimics the evolution of the memes. According to this interpretation, the objectives and constraints capture the goodness of the memes in terms of their imitability. Nevertheless, we would like to emphasize that the evolutionary algorithm or GA in the current MOGA framework has been used as a *multi-objective optimization tool*, and it should not be taken for simulation of the process of evolution, be it linguistic or memetic.

7.5 An Application

As an aside we describe a potential application of the model presented in this chapter. While studying the dialect data for Bengali, we observed that the verb forms of the modern dialects can be obtained through the application of a sequence of APOs, Θ, on the classical Bengali verb forms (see Sec. 7.1.2). This provides us with a noble method for natural language generation in a multi-dialectal scenario, which is described below.

Suppose we want to build a natural language generator (NLG) as a part of an automatic Question Answering system. It is expected that the answers are framed in the dialect of the end-user. We can choose a common ancestor of the dialects, say classical Bengali, and develop an NLG for that language. Any sentence generated by this NLG can then be converted to the required dialect through a set of transformations. As a first step towards this goal, we can conceive of a morphological generator as follows: given a verb root, a set of morphological attributes and a dialect d, the task is to generate the corresponding verb form in d. This task can be accomplished in two steps:

1. Generate the required verb form in classical Bengali using an appropriate morphological generator.

2. Apply a set of phonological transformation rules Θ_d to obtain the corresponding form in dialect d.

Such a system will be advantageous in a multi-dialectal scenario because development of morphological generators for n dialects is expected to be costlier (in terms of development effort) than developing a single generator for classical Bengali along with n transducers implementing the sequence of transformations. Furthermore, since verb morphology of classical Bengali is more regular compared to that of SCB and other modern dialects that feature several morphological paradigms, both the time and structural complexity of the generator for classical Bengali is lower than that of other dialects. Thus, the method provides an elegant and scalable framework for multi-dialectal morphology synthesis.

It should be mentioned, however, that to implement a complete NLG system it is not enough to transform the verb forms, and the framework needs to be extended to capture the morpho-syntactic as well as semantic variations observed across the dialects. Thus, implementation of the proposed framework and extension of the same to syntactic and semantic changes are interesting and challenging research problems that can be explored in the future.

7.6 Conclusions

In this chapter we have described a MOGA framework for modeling the morpho-phonological change of BVIs. The salient features and contributions of the work are

- The concept of atomic phonological operators has been defined. We assume that like the case of SDH, the emergence of an APO can be explained through the MAS Models.
- The problem has been looked upon from a functionalist perspective, whereby we have attempted to provide appropriate quantification of the functional forces, namely articulatory effort and acoustic distinctiveness, and several other constraints.
- The definition of a chromosome as a sequence of APOs is a novel and significant contribution of this chapter. The formulation helps us to simultaneously search for optimal dialects (i.e., phenotypes) and the courses of change leading to those dialects (i.e., genotype).

- We incrementally construct six MOGA models by introducing the different objectives and constraints, and thereby study the effect of the constraints on the emergent dialects.
- The comparison of the Pareto-optimal sets with a few real dialects reveal that not only the real dialects lie close to the front, but we also observe emergence of dialects in the system that are close to the real dialects.

The current framework can be extended in many ways. For instance, one can explore the emergence of the morphological paradigms of Bengali verbs in SCB in the MOGA model. For this, it is necessary to start with a lexicon Λ_0 that has the inflected forms of several verb roots. One can also explore other phonological changes that have in general affected the Bengali words. Here, we have evaluated the fitness functions for a sequence of APOs $\Theta = \theta_1\theta_2 \cdots \theta_r$ on $\Theta(\Lambda_0)$. This ignores the cases where the intermediate lexica, such as $\theta_1(\Lambda_0)$ or $\theta_2(\theta_1(\Lambda_0))$, are suboptimal. Therefore, a possible extension of the model can be to evaluate the fitness at every step of the derivation, the final fitness being a composition of all the intermediate ones. Such a definition of fitness is also capable of measuring the goodness of the course of change than just the dialects that finally emerge.

It is important to point out that the results of the experiments with the MOGA Models must be interpreted with caution. This is because the results are very much dependent on the formulation of the fitness functions and the choice of the constraints. The selection of the APOs in the repertoire also plays a major role in shaping the Pareto-optimal front of the models. This, in fact, is a general drawback of any computational model, which we discuss at length in the next and the last chapter of the thesis.

Chapter 8

Conclusion

Explaining and describing the dynamics of human language is a hard problem that has bedazzled the human minds over the centuries. The advent of sophisticated mathematical tools and powerful computers enabled present day researchers to explore newer computational and formal models of language evolution and change. Nevertheless, as described in Chapter 3, modeling of real world language change turns out to be a notoriously difficult problem. In this thesis, we set out to crack this hard nut by building formal as well as computational models for cases of real world phonological change.

The previous four chapters are essentially dedicated to the aforementioned theme, where we have attempted to build formal models for two cases of real language change – schwa deletion in Hindi and morpho-phonological change of the verb inflections in Bengali, through which we have been able to demonstrate "the thesis" that

> *with suitable abstractions and computational tricks, it is indeed possible to develop formal and/or computational models of real world language change.*

We believe that our thesis makes sense in the context of the fact that almost all the present day computational models of language evolution and change are built for artificial toy languages – a limitation that has been identified and criticized by several researchers (Hauser et al. 2002; Poibeau 2006) despite the otherwise well accepted advantages of such models.

A little reflection would reveal that the models presented in this thesis are also based on numerous simplifying assumptions. We have tried to justify our assumptions in the respective chapters, wherever possible. Nevertheless, it is worthwhile to ask whether the models developed during this work (and computational models of language change in general) qualify as valid explanations of language change. Similarly, one can ask which properties of a model and an experiment make them more plausible than some other model. Furthermore, since it is impossible to develop a model of language change without making any assumptions, it is also worth pondering what kind of assumptions are fairly reasonable or may be even necessary to build a model of real world language change.

In this concluding chapter, we discuss some of these meta-issues regarding the subject matter of the thesis analyzed in the light of the models described here. Since the contributions of this thesis have been summarized in Chapter 1, and the contributions and future works related to each of the models have been discussed at length in the respective chapters, we do not devote any further space on these.

8.1 Need for Abstractions

A model is an "abstraction of a real world scenario" and therefore, any model of language change, and any scientific model for that matter, is based on a series of abstractions. Abstractions, in turn, are based on simplifying assumptions. Hence, the predictions of an abstract model hold good as long as the underlying assumptions are *valid*. Stated differently, a model is as good as the assumptions are, on which it is based.

In Chapter 3 (Sec. 3.2.4) we have seen that the basic difference between the macroscopic and microscopic models of language is that unlike the latter, the macroscopic models assume that the behavior of a linguistic system can be derived from the average behavior of the speakers. Moreover, in order to analytically solve the macroscopic models so obtained, one usually makes assumptions of infinite population (of speakers) and non-overlapping generations. Although apparently these assumptions sound reasonable, some initial investigations have shown that the predictions made by some of the macroscopic models are qualitatively different from their microscopic counterparts (Briscoe 2000c; Stauffer and Schulze 2005).

In Chapter 4, we attempted to develop an optimization model for SDH, which is based on a functional explanation. The amount of reverse-engineering that has undergone behind the formulation of the model is quite obvious. Nevertheless, we have been able to arrive at a *falsifiable prediction* that according to the proposed model, Ohala's rule is equivalent to the optimal pattern if and only if the syllabification of the language does not follow the onset-maximization principle. In other words, our model predicts that there is no language in which the syllabification follows onset maximization and schwa deletion pattern matches that of Hindi's.

Similarly, the observations related to the MOGA model presented in Chapter 7 are overtly dependent on the formulation of the objective functions and constraints. However, we make three important remarks regarding the formulation of the MOGA models.

- The functional forms of the objectives and constraints have been kept as simple as possible.
- The introduction of new constraints or functions were always motivated by the failures associated with the previous models. Thus, in essence, we have attempted to build the "simplest possible" MOGA model that can explain the emergence of the real Bengali dialects – a strategem often referred to as *Occam's razor*.
- We also observe that the model parameters (such as the weights $\alpha_1, \alpha_2, \alpha_3$) have minimal effect on the emergent pattern.

Thus, the MOGA model provides a simple and robust explanation of the morpho-phonological change of BVI. In the two MAS models described in Chapter 5 and 6, a series of abstractions have been made to facilitate computational tractability. The utmost importance of the abstractions can be appreciated from the fact that experiments with a large Hindi lexicon have been possible only with MAS Model II, which is based on much stronger assumptions about the human cognitive faculty, compared to MAS Model I.

Thus, assumptions are necessary for building any model of language change. Moreover, for any model the stronger the assumptions are, the higher the level of abstraction, and consequently, the simpler the explanation. If the assumptions are stated clearly, the model can be used to make falsifiable predictions, which in turn makes the model as well as the underlying assumptions scientifically acceptable.

8.2 Principle of Hierarchical Abstraction

Can there be a formal technique of incorporating assumptions in a model that ensures the plausibility of the model and its predictions? If so then such a technique would be extremely helpful in computational modeling of language change and evolution, and more so in the context of real languages, because these models demand several strong assumptions to be made.

In this thesis, we have put forward the principle of hierarchical abstraction (Sec. 6.6), which can be stated as follows. Starting from an extremely detailed model that is based on fewer assumptions, more general and simpler models are incrementally built by introducing stronger assumptions that are based on the observations of the previous models. For instance, the optimization model presented in Chapter 4 provides us with the insight that SDH is possibly an outcome of an urge towards faster communication through reduction of the durations of schwas, which we incorporate as the primary bias in MAS Model I. Experiments with MAS Model I reveal that upon convergence, the schwas are either completely deleted or fully pronounced. This motivated us to incorporate the CV-pattern based mental model in MAS Model II. Finally, on the basis of the fact that SDH can be explained through computational models, we assumed that similar models can be constructed for other simple problems of phonological change and took them for granted. Based on this assumption, we built the MOGA model for BVI.

Thus, hierarchical abstraction seems to be a useful approach towards modeling of real language change. Nevertheless, while experimenting with the MOGA models, we have chosen just the reverse strategy. We started with the most general model and tried to move towards more specific models by incorporation of new assumptions. This difference between the two approaches crop up from the opposite nature of the two problems tackled. The aim of the models of SDH has been to explain Ohala's rule – a unique pattern. To the contrary, in the problem of BVI there are several solutions (dialects) that are observed in real life and therefore, our aim has been to eliminate the seemingly implausible dialects. Thus, in the former case, we move from the most specific model towards a general model of SDH, whereas in the latter case the reverse is true.

8.3 Plausibility of a Model

Throughout this thesis, we have been arguing in favor of the computational models of language change that are constructed for and validated against real linguistic data. Therefore, it is natural to ask whether and why the models of real language change are more plausible compared to those built for artificial languages.

As a final remark to this thesis, we would like to make a few comments regarding the above question. Consider the problem and models of SDH presented in this thesis. Our aim has been to explain the schwa deletion pattern of Hindi in the light of language change. For this purpose, we have proposed three models that attempt to explain the fact that starting from a state of no schwa deletion (i.e., Sanskrit or Vedic Sanskrit pronunciations) one can indeed arrive at the pronunciation patterns of modern Hindi. Nevertheless, only the initial (Sanskrit pronunciations) and final (Hindi pronunciations) states of the system have been compared with real language data. There are infinitely many paths through which Sanskrit could have given way to Hindi pronunciations. The models proposed here describe only one or some of those paths. In the absence of any further information regarding the nature of this path, is it possible to assert or comment upon the plausibility of the models?

In the case of BVI, the intermediate stages for the real dialect data are known. This information, nevertheless, does not alleviate the aforementioned problem. Classical Bengali has evolved through several historical paths giving rise to the modern dialects of Bengali. We have compared our model with a very few (only 3) of these dialects. Furthermore, classical Bengali could have evolved in many other ways giving rise to other dialectal patterns. In nature, we observe only a handful of those possibilities. Given these facts, what is the plausibility of the MOGA models?

In either case, we have to resort to an "inference to the best explanation". It is impossible to empirically prove the correctness of a computational model, and any linguistic explanation for that matter. However, when it comes to empirical validation of the models, diachronic linguistics is hardly different from the other branches of natural and social sciences. Niyogi (2006) summarizes this fact as follows:

> "... in a historical discipline like evolution, controlled experiments are difficult to design and so the kind of success one wishes is not the sort that one sees in certain areas of physics or engineering. Rather, ... one hopes that one will be able to separate plausible from implausible theories, sort out inconsistencies in reasoning, and generally obtain a deeper qualitative understanding of the phenomena."

In spite of the aforementioned methodological issues and philosophical problems related to their empirical validation, models that attempt to explain real data are more reliable and perhaps, also more plausible than those which explain only qualitative trends. Computational modeling of real world language change, therefore, seems to be a fruitful way of doing research in diachronic linguistics; this thesis shows that the process of modeling is hard, but not impossible.

References

Abrams, D. M. and Strogatz, S. H. 2003. Modelling the dynamics of language death. *Nature 424*, 900.

Albert, R. and Barabási, A. L. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics 74*, 47–97.

Andersen, H. 1973a. Abductive and deductive change. *Language 49*, 765–793.

Andersen, J. M. 1973b. *Structural aspects of language change*. Longmans.

Antiquiera, L., Nunes, M. G. V., Oliveira, O. N., and da F. Costa, L. 2007. Strong correlations between text quality and complex networks features. *Physica A 373*, 811–820.

Avesani, P. and Agostini, A. 2003. A peer-to-peer advertising game. In *Proceedings of the First International Conference on Service Oriented Computing*. Springer-Verlag LNCS 2910, 28–42.

Azad, H. 2001. *Bangla Bhasha: Bangla Bhasha Bishayak Prabandha Sankalan [1743 – 1983]*. Vol. 2. Agami prakashani.

Baeza-Yates, R. A. and Navarro, G. 1999. Faster approximate string matching. *Algorithmica 23,* 2, 127 – 158.

Bailey, C.-J. 1973. *Variation and Linguistics Theory*. Center for Applied Linguistics, Washington D.C.

Bailey, T. M. and Hahn, U. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language 44*, 568 – 591.

Bali, K., Ramakrishnan, A., Talukdar, P. P., and Krishna, N. S. 2004. Tools for the development of a hindi speech synthesis system. In *Fifth ISCA Speech synthesis workshop*. 109–114.

Batali, J. 1998. Computational simulations of the emergence of grammar. In *Approaches to the Evolution of Language: Social and Cognitive Bases*, J. R. Hurford, M. Studdert-Kennedy, and C. Knight, Eds. Cambridge University Press, Cambridge.

Baugh, A. C. and Cable, T. 1993. *A History of the English Language*. Prentice-Hall.

Baxter, G. J., Blythe, R. A., Croft, W., and McKane, A. J. 2006. Utterance selection model of language change. *Physical Review E 73,* 046118.

Belpaeme, T. 2002. Factors influencing the origins of colour categories. Ph.D. thesis, Vrije Universiteit Brussel, Artificial Intelligence Lab.

Belz, A. 1998. An approach to the automatic acquisition of phonotactic constraints. In *Proceedings of SIGPHON'98: The computation of phonological constraints*, T. M. Ellison, Ed. 33–44.

Bertsekas, D. P. 1996. *Constrained optimization and Lagrange multiplier methods*. Academic Press.

Bhat, D. 2001. *Sound Change*. Motilal Banarsidass, new Delhi.

Biemann, C. 2006. Unsupervised part-of-speech tagging employing efficient graph clustering. In *Proceedings of the COLING/ACL 2006 Student Research Workshop*. 7–12.

Blackmore, S. 1999. *The Meme Machine*. Oxford University Press.

Blevins, J. 2004. *Evolutionary Phonology*. Cambridge University Press, Cambridge, MA.

Bodik, P. 2003a. Emergence of language in spatial language games. diploma thesis.

Bodik, P. 2003b. Language change in multi-generational community. In *Proceedings of CALCI'03*.

Bodik, P. and Takac, M. 2003. Formation of a common spatial lexicon and its change in a community of moving agents. In *Frontiers in AI: Proceedings of SCAI'03*. IOS Press.

Boersma, P. 1997a. Functional optimality theory. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 21*, 37–42.

Boersma, P. 1997b. Sound change in functional phonology. Tech. rep.

Boersma, P. 1998. *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. Uitgave van Holland Academic Graphics, Hague, Netherlands.

Brighton, H., Kirby, S., and Smith, K. 2005. Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In *Language Origins: Perspectives on Evolution*, M. Tallerman, Ed. Oxford: Oxford University Press, Chapter 13.

Brighton, H., Smith, K., and Kirby, S. 2005. Language as an evolutionary system. *Physics of Life Reviews 2,* 3 (September), 177–226.

Brill, E. and Moore, R. C. 2000. An improved error model for noisy channel spelling correction. In *Proceedings of 38th Annual Meeting of the Association for Computational Linguistics*. 286–293.

Briscoe, E. J. 2000a. An evolutionary approach to (logistic-like) language change. Draft of DIGS6 talk, Maryland, 23rd May 2000 and Univ of Sussex, 8th Dec 2000.

Briscoe, E. J. 2000b. Evolutionary perspectives on diachronic syntax. In *Diachronic Syntax: Models and Mechanisms*, S. Pintzuk, G. Tsoulas, and A. Warner, Eds.

Briscoe, E. J. 2000c. Macro and micro models of linguistic evolution. unpublished notes on talk presented at The 3rd Int. Conf. on Lg. and Evolution, Paris, April, 2000.

Briscoe, T. 2002. *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press.

Brown, P. F., Pietra, S. A. D., Pietra, V. J. D., and Mercer, R. L. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics 19,* 2, 263–312.

Bybee, J. L. 1985. *Morphology: A Study of the Relation between meaning and form.* John Benjamins Publishing Company.

Bybee, J. L. 1988. The diachronic dimension in explanation. In *Explaining Language Universals.*

Bybee, J. L. 2001. *Phonology and Language Use.* Cambridge University Press.

Cangelosi, A. 2003. Neural network models of category learning and language. *Brain and Cognition 53,* 2, 106–107.

Cangelosi, A. 2005. The emergence of language: neural and adaptive agent models. *Connection Science 17,* 3-4 (December), 185–190.

Cangelosi, A. and Parisi, D. 1998. The emergence of a language in an evolving population of neural networks. *Connection Science 10,* 2, 83–97.

Cangelosi, A. and Parisi, D. 2001. How nouns and verbs differentially affect the behavior of artificial organisms. In *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, J. D. Moore and K. Stenning, Eds. Lawrence Erlbaum Associates, London, 170–175.

Cangelosi, A. and Parisi, D. 2002. Computer simulation: A new scientific approach to the study of language evolution. In *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds. Springer Verlag, London, Chapter 1, 3–28.

Carson-Berndsen, J. 1993. An event-based phonotactics for german. Tech. Rep. ASL-TR-29-92/UBI.

Cavalli-Sforza, L. L. 2000. *Genes, Peoples, and Languages.* North Point Press.

Chambers, J. K. 1995. *Sociolinguistic Theory.* Blackwell.

Chatterji, S. K. 1926. *The Origin and Development of the Bengali Language.* Rupa and Co., New Delhi, India.

Chatterji, S. K. 1927. Gramyashabda sankalan. *Sahitya-Parishad-Patrika 33,* 2, 105 – 108.

Chen, M. and Wang, W. S.-Y. 1975. Sound change: actuation and implementation. *Language 51,* 2, 255–281.

Chomsky, N. 1981. *Lectures on Government and Binding.* Foris.

Chomsky, N. 1995. *The Minimalist Program.* MIT Press, Cambridge, MA.

Chomsky, N. and Halle, M. 1968. *The sound pattern of English.* Harper Row, New York.

Chong, E. K. P. and Żak, S. H. 2001. *An introduction to optimization.* Wiley-Interscience.

Chopde, A. 2001. Itrans version 5.30: A package for printing text in indian languages using english-encoded input. *http://www.aczoom.com/itrans/*.

Christiansen, M. H. and Dale, R. 2003. Language evolution and change. In *Handbook of brain theory and neural networks (2nd ed.)*, M. Arbib, Ed. MIT Press, Cambridge, MA, 604–606.

Christiansen, M. H. and Devlin, J. T. 1997. Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Mahwah, NJ, 113–118.

Christiansen, M. H. and Kirby, S. 2003a. Language evolution: Consensus and controversies. *Trends in Cognitive Sciences 7,* 7 (7), 300–307.

Christiansen, M. H. and Kirby, S. 2003b. Language evolution: The hardest problem in science? In *Language Evolution: The States of the Art*, M. Christiansen and S. Kirby, Eds. Oxford University 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 sonority cycle in core syllabification. In *Papers in laboratory phonology I: Between the grammar and physics of speech*, J. Kingston and M. Beckman, Eds. Cambridge University Press, New York, 283 – 333.

Comrie, B. 1981. *Language Universals and Linguistic Typology.* Basil Blackwell.

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach.* Longman Linguistic Library.

Crothers, J. 1978. Typology and universals of vowel systems. In *Universals of Human Language (Phonology)*, J. H. Greenberg, Ed. Vol. 2. Stanford University Press, California, 93 – 152.

Dall'Asta, L., Baronchelli, A., Barrat, A., and Loreto, V. 2006a. Agreement dynamics on small-world networks. *Europhysics Letters 73,* 6, 969–975.

Dall'Asta, L., Baronchelli, A., Barrat, A., and Loreto, V. 2006b. Non-equilibrium dynamics of language games on complex networks. *Physical Review E 74,* 036105.

Dawkins, R. 1976. *The Selfish Gene*. Oxford University Press.

de Boer, B. 1997. Self organisation in vowel systems through imitation. In *Computational Phonology, Third Meeting of the ACL SIGPHON*, J. Coleman, Ed. 19–25.

de Boer, B. 1998. Emergence of sound systems through self-organisation. In *Proceedings of the Tenth Netherlands/Belgium Conference on Artificial Intelligence NAIC'98*, H. L. Poutre and J. van den Herik, Eds. CWI, Amsterdam, The Netherlands, 37–46.

de Boer, B. 1999a. Evolution and self-organisation in vowel systems. *Evolution of Communication 3,* 1, 79–103.

de Boer, B. 1999b. Self-organisation in vowel systems. Ph.D. thesis, Vrije Universiteit Brussel AI-lab.

de Boer, B. 2000a. Emergence of vowel systems through self-organisation. *AI Communications 13,* 1, 27–39.

de Boer, B. 2000b. Self-organization in vowel systems. *Journal of Phonetics 28,* 4 (October), 441–465.

de Boer, B. 2001. *The Origins of Vowel Systems*. Oxford University Press.

de Boer, B. 2005. Infant-directed speech and evolution of language. In *Language Origins: Perspectives on Evolution*, M. Tallerman, Ed. Oxford University Press, Chapter 5.

de Boer, B. 2006. Computer modeling as a tool for understanding language evolution. In *Studies in Linguistics companion series*, N. Gontier, Ed.

de la Briandais, R. 1959. File searching using variable length keys. In *Proceedings of the Western Joint Computer Conference*. 295 – 298.

de Oliveira, V. M., Campos, P. R., Gomes, M., and Tsang, I. 2006. Bounded fitness landscapes and the evolution of the linguistic diversity. *Physica A: Statistical Mechanics and its Applications 368,* 1 (August), 257–261.

de Oliveira, V. M., Gomes, M., and Tsang, I. 2006. Theoretical model for the evolution of the linguistic diversity. *Physica A: Statistical Mechanics and its Applications 361,* 1 (February), 361–370.

de Saussure, F. 1916. *Cours de linguistique générale*. Payot, Lousanne and Paris.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation 6*, 182–197.

Dircks, C. and Stoness, S. C. 1999. Effective lexicon change in the absence of population flux. In *ECAL99*, D. Floreano, J. Nicoud, and F. Mondada, Eds. Springer-Verlag, Berlin, 720–724.

Dorogovtsev, S. N. and Mendes, J. F. F. 2001. Language as an evolving word web. *Proceedings of The Royal Society of London. Series B, Biological Sciences 268,* 1485 (December), 2603–2606.

Dras, M., Harrison, D., and Kapicioglu, B. 2003. Emergent behavior in phonological pattern change. In *Artificial Life VIII*. MIT Press.

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*. Almqvist & Wiksell.

Ellison, T. M. 1994. Phonological derivations in optimality theory. In *Proceedings of 15th International Conference on Computational Linguistics (COLING)*. Vol. 1. 1007–1013.

Ellison, T. M. and Kirby, S. 2006. Measuring language divergence by intra-lexical comparison. In *Proceedings of COLING/ACL*. 273–280.

Ewens, W. J. 2004. *Mathematical Population Genetics*. Springer-Verlag Inc.

Ferrer-i-Cancho, R. 2003. Language: universals, principles and origins. Ph.D. thesis.

Ferrer-i-Cancho, R. 2005a. Decoding least effort and scaling in signal frequency distributions. *Physica A: Statistical Mechanics and its Applications 345,* 1-2 (January), 275–284.

Ferrer-i-Cancho, R. 2005b. The structure of syntactic dependency networks: insights from recent advances in network theory. In *Problems of quantitative linguistics*, L. V. and A. G., Eds. 60–75.

Ferrer-i-Cancho, R. and Sole, R. V. 2001. The small world of human language. *Proceedings of The Royal Society of London. Series B, Biological Sciences 268,* 1482 (November), 2261–2265.

Ferrer-i-Cancho, R. and Sole, R. V. 2004. Patterns in syntactic dependency networks. *Physical Review E 69,* 051915.

Fontaine, C. 1985. Application de méthodes quantitatives en diachronie: l'inversion du sujet en français. Master Thesis, Université du Québec à Montréal.

Fontana, J. M. 1985. Application phrase structure and the syntax of clitics in the history of spanish. Ph.D. thesis, University of Pennsylvania.

Fredkin, E. 1960. Trie memory. *Communications of the ACM 3,* 9, 490 – 499.

Fyfe, C. and Livingstone, D. 1997. Developing a community language. In *ECAL97*. Brighton, UK.

Gell-Mann, M. 1992. Complexity and complex adaptive systems. In *The Evolution of Human Languages*, J. A. Hawkins and M. Gell-Mann, Eds. Addison-Wesley, Reading, MA.

Gell-Mann, M. 1995. What is complexity? *Complexity 1,* 1, 16–19.

Gell-Mann, M. 2005. Language and complexity. In *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*, J. W. Minett and W. S.-Y. Wang, Eds. City University of Hong Kong Press.

Gibson, E. and Wexler, K. 1994. Triggers. *Linguistic Inquiry 25,* 4, 407–454.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

Goldsmith, J. A. 1990. *Autosegmental and metrical phonology*. Blackwell, Cambridge, MAS.

Gong, T., Ke, J., Minett, J. W., and Wang, W. S.-Y. 2004. A computational framework to simulate the co-evolution of language and social structure. In *Artificial Life IX*. Boston, MA, U.S.A.

Gong, T. and Wang, W. S.-Y. 2005. Computational modeling on language emergence: A coevolution model of lexicon, syntax and social structure. *Language and Linguistics 6,* 1, 1–41.

Goslin, J. and Frauenfelder, U. H. 2000. A comparison of theoretical and human syllabification. *Language and Speech 44,* 4, 409–436.

Greenberg, J. H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, J. H. Greenberg, Ed. MIT Press, Cambridge, MA, 73–113.

Grierson, G. A. 1903. Upabhashar namuna. In *Linguistic Survey of India, Vol V: part 1: Indo-Aryan family, Eastern Group, Bengali and Assamese languages.*

HARE, M. AND ELMAN, J. L. 1995. Learning and morphological change. *Cognition 56,* 1 (July), 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)*. New York, 217–236.

HAUSER, M. D. 1997. *The Evolution of Communication*. MIT Press/BradfordBooks.

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.

HAWKINS, J. A. 1983. *Word Order Universals*. Academic Press.

HAYES, B., KIRCHNER, R., AND STERIADE, D. 2004. *Phonetically Based Phonology*. Cambridge University Press.

HOCKETT, C. F. AND ALTMANN, S. A. 1968. A note on design features. In *Animal communications: techniques of study and results of research*, T. A. Sebeok, Ed. Indiana University Press, Bloomington, 61–72.

HOLANDA, A. D., PISA, I. T., KINOUCHI, O., MARTINEZ, A. S., AND RUIZ, E. E. S. 2004. Thesaurus as a complex network. *Physica A 334*, 530–536.

HOLLAND, J. H. 1975. *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor.

HOOPER, J. 1978. Constraints on schwa-deletion in american english. In *Recent Developments in Historical Linguistics*, J. Fisiak, Ed. Mouton, Hague, 183–207.

HOOPER, J. B. 1972. The syllable in phonological theory. *Language 48*, 525–540.

HUDSON, R. 2006. *Language Network: The New Word Grammar*. Oxford University Press.

HURFORD, J. 1989. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua 77*, 187–222.

HURFORD, J. 1991a. The evolution of the critical period for language acquisition. *Cognition 40,* 3, 159–201.

HURFORD, J. R. 1991b. Nativist and functional explanations in language acquisition. In *Logical Issues in language Acquisition*, I. Roca, Ed. Holland Foris publications, Dordretch, 85–136.

ISLAM, R. 1979. Upabhashatatta o bangladesher upabhasha bishleshan. *Sahitya Patrika 22,* 1, 83–96.

ITOH, Y. AND UEDA, S. 2004. The ising model for changes in word ordering rules in natural languages. *Physica D: Nonlinear Phenomena 198,* 3-4 (11), 333–339.

JIM, K.-C. AND GILES, C. L. 2000. Talking helps: Evolving communicating agents for the predator-prey pursuit problem. *Artificial Life 6,* 3, 237–254.

JOANISSE, M. F. 1999. Exploring syllable structure in connectionist networks. In *Proceedings of the 14th International Congress of the Phonetic Sciences (ICPhS-99)*. 731–734.

JOANISSE, M. F. AND SEIDENBERG, M. 1997. [i e a u] and sometimes [o]: perceptual and computational constraints on vowel inventories. In *Proceedings of the 15th Conference of the Cognitive Science Society*. Stanford, CA.

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

JUSCZYK, P. W. AND LUCE, P. A. 2002. Speech perception and spoken word recognition: Past and present. *Ear & Hearing 23*, 2–40.

KAHN, D. 1976. *Syllable-based generalizations in English phonology.* Indiana Univ. Linguistics Club, Bloomington, IN.

KAIRA, S. 1976. *Schwa-deletion in Hindi.* Language forum (back volumes). Bhari publications, New Delhi, India.

KAPATSINSKI, V. 2006. Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Speech research Lab Progress Report 27*, 133 – 152.

KAPLAN, R. M. AND KAY, M. 1994. Regular models for phonological rule systems. *Computational Linguistics 20,* 3, 331–378.

KE, J. 2004. Self-organization and language evolution: System, population and individual. Ph.D. thesis, Department of Electronic Engineering, City University of Hong Kong.

KE, J. AND COUPÉ, C. 2002. The fall and rise of homophones: a window to language evolution. In *Fourth International Conference on Language Evolution.* Harvard.

KE, J., OGURA, M., AND WANG, W. S.-Y. 2003. Modeling evolution of sound systems with genetic algorithm. *Computational Linguistics 29,* 1, 1–18.

KE, J. AND YAO, Y. 2006. Analyzing language development from a network approach.

KERNIGHAN, M. D., CHURCH, K. W., AND GALE, W. A. 1990. A spelling correction program based on a noisy channel model. In *Proceedings of COLING.* ACL, NJ, USA, 205–210.

KIPARSKY, P. 1988. Phonological change. In *Cambridge Survey of Linguistics*, F. Newmeyer, Ed. Vol. 1. Cambridge University Press, 363–410.

KIPARSKY, P. 1996. The shift to head-initial vp in germanic. In *Studies in Comparative Germanic Syntax*, S. Epstein, H. Thrainsson, and S. Peters, Eds. Vol. 2. Kluwer, Dordrecht.

KIRBY, S. 1999. *Function, Selection and Innateness: the Emergence of Language Universals.* Oxford University Press. The full-text is only a sample (chapter 1: A Puzzle of Fit).

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.

KIRBY, S. 2002. Learning, bottlenecks and the evolution of recursive syntax. In *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, T. Briscoe, Ed. Cambridge University Press, Chapter 6.

KIRBY, S. AND HURFORD, J. 1997. The evolution of incremental learning: language, development and critical periods. Tech. rep., Language Evolution and Computation Research Unit, University of Edinburgh.

Kishore, S. P. and Black, A. 2003. Unit size in unit selection speech synthesis. In *Proceedings of Eurospeech 2003*.

Kishore, S. P., Kumar, R., and Sangal, R. 2002. A data driven synthesis approach for indian languages using syllable as basic unit. In *Proceedings of International Conference on Natural Language Processing*, R. Sangal and S. Bendre, Eds. Vikas Publishing House Pvt. Ltd., New Delhi, India, 311–316.

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 University Press, 276–290.

Klein, S., Kuppin, M. A., and Kirby, A. 1969. Monte carlo simulation of language change in tikopia and maori. In *Proceedings of COLING*.

Knuth, D. 1997. *The art of computer programming, Volume 3: Sorting and Searching*. Addison-Wesley.

Kohler, K. L. 1966. Is syllable a phonetic universal? *Journal of Linguistics 2*, 207–208.

Komarova, N. L. and Niyogi, P. 2004. Optimizing the mutual intelligibility of linguistic agents in a shared world. *Artificial Intelligence 154,* 1-2 (April), 1–42.

Komarova, N. L., Niyogi, P., and Nowak, M. A. 2001. The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology 209,* 1, 43–59.

Komarova, N. L. and Nowak, M. A. 2003. Language dynamics in finite populations. *Journal of Theoretical Biology* 3, 445–457.

Kosmidis, K., Halley, J. M., and Argyrakis, P. 2005. Language evolution and population dynamics in a system of two interacting species. *Physica A: Statistical Mechanics and its Applications 353*, 595–612.

Kosmidis, K., Kalampokis, A., and Argyrakis, P. 2006a. Language time series analysis. *Physica A: Statistical Mechanics and its Applications 370,* 2 (October), 808–816.

Kosmidis, K., Kalampokis, A., and Argyrakis, P. 2006b. Statistical mechanical approach to human language. *Physica A: Statistical Mechanics and its Applications 366*, 495–502.

Kroch, A. 1989a. Reflexes of grammar in patterns of language change. *Language Variation and Change 1*, 199–244.

Kroch, A. 1989b. Reflexes of grammar in patterns of language change. *Language Variation and Change 1*, 199–244.

Kroch, A. 2001. Syntactic change. In *Handbook of Syntax*, M. Baltin and C. Collins, Eds. Blackwell, 699–729.

Kroch, A. and Taylor, A. 1997a. The middle english verb-second constraint: A case study in language contact and language change. In *textual parameters in older language*, S. Herring, P. van Reenen, and lene Schoesler, Eds. John Benjamins, Philadelphia.

Kroch, A. and Taylor, A. 1997b. Verb movement in old and middle english: Dialect variation and language contact. In *Parameters and morphosyntatic change*. Cambridge University Press, New York, 297–325.

Kukich, K. 1992. Technique for automatically correcting words in text. *ACM Computing Surveys 24*, 377 – 439.

Labov, W. 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.

Labov, W. 1975a. The use of the present to explain the past. In *Proceedings of the 11th International Congress of Linguists*, L. Heilmann, Ed. Il Molino, 824–851.

Labov, W. 1975b. *What is a Linguistic Fact*. Peter de Ridder Press.

Ladefoged, P. and Maddieson, I. 1996. *Sounds of the World's Languages*. Blackwell.

Landau, L. D. and Lifshitz, E. M. 1994. *Statistical Physics Part 1*. Course of Theoretical Physics, vol. 5. Pergamon.

Lass, R. 1980. *On Explaining Language Change*. Cambridge University Press.

Lass, R. 1997. *Historical Linguistics and Language Change*. Cambridge University Press.

Lebrun, Y. 1966. Sur la syllable, somet de sonorité. *Phonetica 14*, 1–15.

Lee, Y., Collier, T. C., Taylor, C. E., and Stabler, E. P. 2005. The role of population structure in language evolution. In *Proceedings of the 10th International Symposium on Artificial Life and Robotics*.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics-Doklady 10*, 707–710.

Lightfoot, D. 1991. *How to Set Parameters: Arguments from Language Change*. MIT Press/Bradford Books, Cambridge, MA.

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

Liljencrants, L. and Lindlom, B. 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language 48*, 839–862.

Lindblom, B. 1986. Phonetic universals in vowel systems. In *Experimental Phonology*, J. J. Ohala and J. J. Jaear, Eds. Academic Press, 13–44.

Lindblom, B. 1998. Systemic constraints and adaptive change in the formation of sound structure. In *Approaches to the Evolution of Language: Social and Cognitive Bases*, J. R. Hurford, M. Studdert-Kennedy, and K. C., Eds. Cambridge University Press, Cambridge.

Livingstone, D. 2002. The evolution of dialect diversity. In *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds. Springer Verlag, London, Chapter 5, 99–118.

Livingstone, D. 2003. Computer models of the evolution of language and languages. Ph.D. thesis, University of Paisley.

Lowe, J. B. and Mazaudon, M. 1994. The reconstruction engine: A computer implementation of the comparative method. *Computational Linguistics 20,* 3, 381 – 417.

Luce, P. A. and Pisoni, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear & Hearing 19*, 1 – 36.

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge University Press, Cambridge.

Maniruzzaman. 1974. Dhakai upabhasha: Adiyabad pratyanchal. In *Bidyasagar Smarak Grantha*, A. Khan and U. Chattopadhyay, Eds. Bidyasagar Saraswat Samaj, 427–456.

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., and Older, L. 1994. Morphology and meaning in the english mental lexicon. *Psychological Review 101*, 3–33.

Meillet, A. 1967. *The comparative method in historical linguistics*. Champion, Paris.

Mihov, S. and Schulz, K. U. 2004. Fast approximate search in large dictionaries. *Computational Linguistics 30,* 4, 451–477.

Milroy, J. 1993. On the social origins of language change. In *Historical Linguistics: Problems and Perspectives*, C. Jones, Ed. Longman, London, New York.

Minett, J. W., Gong, T., and Wang, W. S.-Y. 2006. A language emergence model predicts word order bias. In *Proceedings of the 6th International Conference on the Evolution of Language*. 206–213.

Mira, J. and Paredes, A. 2005. Interlinguistic similarity and language death dynamics. *Europhysics Letters 69,* 6, 1031–1034.

Mishra, B. G. 1967. Historical phonology of standard hindi: Proto indo european to the present. Ph.D. thesis, Cornell University.

Mitchell, T. 1997. *Machine Learning*. McGraw Hill.

Modis, T. 2002. *Predictions: 10 years later*. Growth Dynamics.

Montana, D. J. and Davies, L. D. 1989. Training feedforward networks using genetic algorithms. In *Proceedings of the International Conference on Genetic Algorithms*. Morgan Kaufmann.

Motter, A. E., de Moura, A. P. S., Lai, Y.-C., and Dasgupta, P. 2002. Topology of the conceptual network of language. *Physical Review E 65,* 065102.

Nakhleh, L., Warnow, T., Ringe, D., and Evans, S. 2005. A comparison of phylogenetic reconstruction methods on an ie dataset. *Transactions of the Philological Society 3,* 2, 171–192. Under review.

Narasimhan, B., Sproat, R., and Kiraz, G. A. 2004. Schwa-deletion in hindi text-to-speech synthesis. *International Journal of Speech Technology 7,* 4, 319–333.

Nettle, D. 1999. Using social impact theory to simulate language change. *Lingua 108*, 95–117.

Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Review 45*, 167–256.

Niyogi, P. 1997. Modeling the dynamics of historical linguistics. In *New England Conference on Complex Systems*. Nashua, NH.

Niyogi, P. 2002. The computational study of diachronic linguistics. In *Syntactic Effects of Morphological Change*, D. Lightfoot, Ed. Cambridge University Press.

Niyogi, P. 2006. *The Computational Nature of Language Learning and Evolution*. MIT Press, Cambridge, MA.

Niyogi, P. and Berwick, R. C. 1996. A language learning model for finite parameter spaces. *Cognition 61,* 1-2, 161–193.

Niyogi, P. and Berwick, R. C. 1997a. A dynamical systems model for language change. *Complex Systems 11*, 161–204.

Niyogi, P. and Berwick, R. C. 1997b. Evolutionary consequences of language learning. *Linguistics and Philosophy 20,* 6, 697–719.

Niyogi, P. and Berwick, R. C. 1997c. Populations of learners: The case of european portuguese. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Stanford, CA.

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*.

Nowak, M. A., Komarova, N. L., and Niyogi, P. 2001. Evolution of universal grammar. *Science 291*, 114–118.

Odell, M. K. and Russell, R. C. 1918. U.s. patent number 1,261,167.

Odell, M. K. and Russell, R. C. 1922. U.s. patent number 1,435,663.

Ohala, J. 1974. Phonetic explanations in phonology. In *Papers from parasession on natural phonology*, A. Bruck, R. A. Fox, and M. W. LaGaly, Eds. Chicago Linguistics Society, Chicago, 251–274.

Ohala, J. 1983a. The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, P. F. MacNeilage, Ed. Springer, 189–216.

Ohala, J. 1987. Explanations in phonology opinions and examples. In *Phonologica 1984: Proceedings of the Fifth International Phonology Meeting*. Cambridge University Press, London, 215–225.

Ohala, J. 1989. Sound change is drawn from a pool of synchronic variation. In *Language change: Contributions to the study of its causes*, L. Breivik and E. Jahr, Eds. Mouton de Gruyter, Berlin, 173–198.

Ohala, J. 1990a. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *The Parasession on the Syllable in Phonetics and Phonology, Vol. 2: Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, Ziolkowski, Noske, and Deaton, Eds. CLS, Chicago, 319–338.

Ohala, J. 1990b. The phonetics and phonology of aspects of assimilation. In *Papers in Laboratory Phonology, vol 1: Between the grammar and physics of speech*. Cambridge University Press, 258–275.

Ohala, J. 1993. The phonetics of sound change. In *Historical linguistics: Problems and perspectives*, C. Jones, Ed. Longman, London, 237278.

Ohala, M. 1977. The treatment of phonological variation: An example from hindi. *Lingua 42*, 161–76.

Ohala, M. 1983b. *Aspects of Hindi Phonology*. MLBD Series in Linguistics, vol. II. Motilal Banarsidass, New Delhi, India.

Oliphant, M. 1996. The dilemma of saussurean communication. *Biosystems 1,* 37, 31–38.

Oliphant, M. and Batali, J. 1997. Learning and the emergence of coordinated communication. *The newsletter of the Center for Research in Language 11,* 1.

Ott, E. 2002. *Chaos in Dynamical Systems*. Cambridge University Press.

Oudeyer, P.-Y. 1999. Self-organization of a lexicon in a structured society of agents. In *ECAL99*, D. Floreano, J.-D. Nicoud, and F. Mondada, Eds. Springer-Verlag, 726–729.

Oudeyer, P.-Y. 2005a. How phonological structures can be culturally selected for learnability. *Adaptive Behavior 13,* 4, 269–280.

Oudeyer, P.-Y. 2005b. The self-organization of speech sounds. *Journal of Theoretical Biology 233,* 3, 435–449.

Pardo, T. A. S., Antiquiera, L., Nunes, M. G. V., Oliveira, O. N., and da F. Costa, L. 2006. Modeling and evaluating summaries using complex networks. *VII PROPOR*.

Parisi, D. and Cangelosi, A. 2002. A unified simulation scenario for language development, evolution, and historical change. In *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds. Springer Verlag, London, Chapter 12, 255–276.

Patriarca, M. and Leppanen, T. 2004. Modeling language competition. *Physica A: Statistical Mechanics and its Applications 338,* 1-2 (July), 296–299.

Perfors, A. 2002. Simulated evolution of language: a review of the field. *Journal of Artificial Societies and Social Simulation 5,* 2.

Pinasco, J. and Romanelli, L. 2006. Coexistence of languages is possible. *Physica A: Statistical Mechanics and its Applications 361,* 1 (February), 355–360.

Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. HarperCollins, New York.

Pinker, S. 2000. Survival of the clearest. *Nature 404*, 441–442.

Pinker, S. and Bloom, P. 1990. Natural language and natural selection. *Behavioral and Brain Sciences 13,* 4, 707–784.

Poibeau, T. 2006. Linguistically grounded models of language change. In *28th annual conference of the Cognitive Science Society (CogSci 2006)*. 255–276.

Pray, B. R. 1970. *Topics in Hindi - Urdu grammar*. Center for South and Southeast Asia Studies, University of California, Berkeley.

Prince, A. and Smolensky, P. 1993. Optimality theory: Constraint interaction in generative grammar. Tech. Rep. 2.

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

Pulgram, E. 1970. *Syllable, Word, Nexus, Cursus*. Mouton, The Hague.

Redford, M. A. 1999. An articulatory basis for the syllable. Ph.D. thesis, Psychology, University of Texas, Austin.

Redford, M. A., Chen, C. C., and Miikkulainen, R. 1998. Modeling the emergence of syllable systems. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, M. Gernsbacher and S. Derry, Eds. Lawrence Erlbaum Associates, Mahwah, NJ, 882–886.

Redford, M. A., Chen, C. C., and Miikkulainen, R. 2001. Constrained emergence of universals and variation in syllable systems. *Language and Speech 44*, 27–56.

Redford, M. A. and Diehl, R. L. 1999. The relative perceptibility of syllable-initial and syllable-final consonants. *Journal of Acoustic Society of America 106*, 1555–1565.

Roberts, I. 2001. Language change and learnability. In *Language Acquisition and Learnability*, M. Bertolo, Ed. Cambridge University Press, Cambridge, UK, 81–125.

Rudin, W. 1987. *Real and Complex Analysis*. McGraw-Hill.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol 1: Foundations)*, D. E. Rumelhart and J. L. McClelland, Eds. MIT Press.

Russell, S. J. and Norvig, P. 2002. *Artificial Intelligence: A Modern Approach*. Prentice Hall.

Sampson, G. 1970. On the need of phonological base. *Language 46*, 586–626.

Santorini, B. 1989. The generalization of the verb-second constraint in the history of yiddish. Ph.D. thesis, University of Pennsylvania.

Schwartz, J.-L., Boë, L.-J., Vallée, M., and Abry, C. 1997. Dispersion-focalization theory of vowel systems. *Journal of Phonetics 25*, 255–286.

Selkirk, E. O. 1982. The syllable. In *The structure of phonological representations*, H. van der Hulst and N. Smith, Eds. Vol. 2. Foris, Dordrecht, The Netherlands, 337 – 383.

Sen, A. and Samudravijaya, K. 2002. Indian accent text-to-speech system for web browsing. *Sādhanā 27,* 1, 113–126.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell system technical journal 27*, 379–423, 623–656.

Sigman, M. and Cecchi, G. 2002. Global organization of the wordnet lexicon. *PNAS 99,* 3, 1742–1747.

Slobin, D. I. 1979. *Psycholinguistics*. Scott, Foresman and Company, Gelnview, Illinois.

Smith, A. D. M. 2005. Mutual exclusivity: Communicative success despite conceptual divergence. In *Language Origins: Perspectives on Evolution*, M. Tallerman, Ed. Oxford University Press, 507 – 516.

Smith, K. 2002. The cultural evolution of communication in a population of neural networks. *Connection Science 14,* 1, 65–84.

Smith, K. and Hurford, J. 2003. Language evolution in populations: extending the iterated learning model. In *ECAL03*. 507–516.

Soares, M. M., Corso, C., and Lucena, L. S. 2005. Network of syllables in portuguese. *Physica A: Statistical Mechanics and its Applications 355,* 2-4 (September), 678–684.

Sole, R. V., Corominas, B., Valverde, S., and Steels, L. 2005. Language networks: their structure, function and evolution. *Trends in Cognitive Sciences*.

Stauffer, D. and Schulze, C. 2005. Microscopic and macroscopic simulation of competition between languages. *Physics of Life Reviews 2,* 2 (June), 89–116.

Steels, L. 1995. A self-organizing spatial vocabulary. *Artificial Life 2,* 3, 319–332.

STEELS, L. 1996. Emergent adaptive lexicons. In *SAB96*, P. Maes, Ed. MIT Press, Cambridge, MA.

STEELS, L. 1997a. Language learning and language contact. In *Proceedings of the workshop on Empirical Approaches to Language Aquisition*, W. Daelemans, A. V. den Bosch, and A. Weijters, Eds. Prague, 11–24.

STEELS, L. 1997b. The synthetic modeling of language origins. *Evolution of Communication 1,* 1, 1–34.

STEELS, L. 1998. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence 103,* 1-2, 133–156.

STEELS, L. 2000. Language as a complex adaptive system. In *Proceedings of PPSN VI*, M. Schoenauer, Ed. Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany.

STEELS, L. 2001. Language games for autonomous robots. *IEEE Intelligent systems*, 16–22.

STEELS, L. AND KAPLAN, F. 1998. Spontaneous lexicon change. In *COLING-ACL98*. ACL, Montreal, 1243–1249.

STERIADE, D. 2001. *The Phonology of perceptibility effects: the P-map and its consequences for constraint organization.* http://web.mit.edu/linguistics/people/faculty/steriade/papers/P-map_for_phonology.doc.

SUGITA, Y. AND TANI, J. 2005. Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adaptive Behavior 13,* 1 (March), 33–52.

TAMARIZ, M. 2005. Exploring the adaptive structure of the mental lexicon. Ph.D. thesis, Department of theoretical and applied linguistics, Univerisity of Edinburgh.

TESILEANU, T. AND MEYER-ORTMANNS, H. 2006. Competition of languages and their hamming distance. *International Journal of Modern Physics C 17,* 2, 259–278.

TOMASELLO, M. 2002. Some facts about primate (including human) communication and social learning. In *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds. Springer Verlag, London, Chapter 15, 327–340.

TOMASELLO, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Harvard University Press.

TRANEL, B. 1999. Optional schwa deletion: on syllable economy in french. In *Formal Perspectives on Romance Linguistics*, J. M. A. et al, Ed.

TURNER, H. 2002. An introduction to methods for simulating the evolution of language. In *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds. Springer Verlag, London, Chapter 2, 29–50.

UKKONEN, E. 1985. Finding approximate patterns in strings. *Journal of Algorithms 6,* 1-3, 132–137.

VENNEMANN, T. 1988. *Preference laws for syllable structure and explanation of sound change.* Mouton de Gruyter, Berlin.

VINCENT, N. 1976. Perceptual factors in word order change in latin. In *Romance syntax*, M. Harris, Ed. University of Salford, Salford, UK, 54–68.

VITEVITCH, M. S. 2005. Phonological neighbors in a small world: What can graph theory tell us about word learning? Spring 2005 Talk Series on Networks and Complex Systems, Indiana University.

WAGNER, K., REGGIA, J., URIAGEREKA, J., AND WILKINSON, G. 2003. Progress in the simulation of emergent communication and language. *Adaptive Behavior 11,* 1, 37–69.

WANG, W. S.-Y. 1991. The three scales of diachrony. In *Explorations in Language*. Pyramid Press, Taipei, Taiwan, 60–71.

WANG, W. S.-Y., KE, J., AND MINETT, J. W. 2005. Computational studies of language evolution. In *Computational Linguistics and Beyond: Perspectives at the beginning of the 21st Century, Frontiers in Linguistics 1. Language and Linguistics*, C. Huang and W. Lenders, Eds.

WATERMAN, J. C. 1966. *A History of the German Language*. Waveland Press Inc.

WEINREICH, U., LABOV, W., AND HERZOG, M. I. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics, a Symposium*, W. P. Lehmann and Y. Malkiel, Eds. University of Texas Press, Austin, TX, 95–195.

WIKIPEDIA. 2007. Ising model. *http://en.wikipedia.org/wiki/Ising_model*.

WU, S. AND MANBER, U. 1992. Fast text searching allowing errors. *Communications of the ACM 35,* 10, 83 – 91.

XIAO, Z., MCENERY, A., BAKER, P., AND HARDIE, A. 2004. Developing asian language corpora: standards and practice. In *Proceedings of the Fourth Workshop on Asian Language Resources*, V. Somlertlamvanich, T. Tokunaga, and C. Huang, Eds. Blackwell, 1–8.

YANG, C. D. 2000. Internal and external forces in language change. *Language Variation and Change 12,* 3 (October), 231–250.

ZOBEL, J. AND DART, P. 1995. Finding approximate matches in large lexicons. *Software – Practice and Experience 25,* 3, 331–345.

Appendix A

Publications from the Thesis

Publications on the work of the thesis : In chronological order with the comments within parenthesis.

CHOUDHURY, M. AND BASU, A. 2002. A rule-based schwa deletion algorithm for Hindi. In *Proceedings of the International Conference on Knowledge-based Computer Systems (KBCS)*. Navi Mumbai, India, 343–353. (The algorithm for SDH, Chapter 4)

CHOUDHURY, M., BASU, A. AND SARKAR, S. 2004. A diachronic approach for schwa deletion in Indo-Aryan languages. In *Proceedings of the Seventh Meeting of the ACL Special Interest Group in Computational Phonology (SIGPHON)*. ACL, Barcelona, Spain, 20–26. (A preliminary version of the constrained optimization framework for SDH, Chapter 4)

CHOUDHURY, M., BASU, A. AND SARKAR, S. 2006. Multi-agent simulation of emergence of the schwa deletion pattern in Hindi. *Journal of Artificial Societies and Social Simulation (JASSS)*. 9(2). http://jasss.soc.surrey.ac.uk/9/2/2.html (MAS Model I, Chapter 5)

CHOUDHURY, M., ALAM, M., SARKAR, S. AND BASU A. 2006. A rewrite rule based model of Bangla morpho-phonological change. In *Proceedings of the International Conference on Computer Processing of Bangla (ICCPB)*. Dhaka, Bangladesh, 64–71. (Application of computational model of language change in natural language generation, Chapter 7)

Choudhury, M., Jalan, V., Sarkar, S. and Basu, A. 2007. Evolution, optimization and language change: the case of Bengali verb inflections. To be published in the *Proceedings of ACL SIGMORPHON9*. Prague, Czech Republic. (MOGA Model, Chapter 7)

Other related publications : In chronological order with comments in parenthesis.

Choudhury, M. and Basu, A. 2002. A rule-based grapheme to phoneme mapping for Hindi. In *Abstracts of the Presentations by the Young Scientists: 90th Indian Science Congress of ISCA*. Bangalore, India. (Algorithms for SDH, syllabification in Hindi and other related issues. The paper won the ISCA Young Scientist Award)

Bansal, B., Choudhury, M., Ray, P. R., Sarkar, S. and Basu, A. 2004. Isolated-word error correction for partially phonemic languages using phonetic cues. In *Proceedings of the International Conference on Knowledge-based Computer Systems (KBCS)*. Allied Publishers, Hyderabad, India, 509–519 (Application of grapheme-to-phoneme converter for Bengali/Hindi in spell-checking.)

Bhattacharya, B., Choudhury, M., Sarkar, S. and Basu, A. 2005. Inflectional morphology synthesis for Bengali noun, pronoun and verb systems. In *Proceedings of the National Conference on Computer Processing of Bangla (NCCPB)*. Dhaka, Bangladesh, 34–43. (Describes in detail the morphological characteristics of Bengali, including verb inflections and their synthesis.)

Choudhury, M., Sarkar, S. and Basu, A. 2005. Affects of inflectional and derivational morphology in Bengali phonology and its implications in speech technology. In *Proceedings of the Workshop on Morphology*. Mumbai, India. (Describes the nature of schwa deletion in Bengali)

Mukhopadhyay, A., Chakraborty, S., Choudhury, M., Lahiri, A., Dey, S. and Basu, A. 2005. Shruti – an embedded text-to-speech system for Indian languages. *IEE Proceedings on Software Engineering*, 153, 2, 75–79. (An application of the grapheme-to–phoneme converters for Hindi and Bengali)

Banerjee, A., Choudhury, M., Sarkar, S. and Basu, A. 2005. Learning schwa pronounceability rules in Bengali compound words using decision trees. In *Proceedings of the Second Symposium on Indian Morphology, Phonology and Language Engineering (SIMPLE).* Mysore:CIIL, Kharagpur, India, 49–55. (Interaction of schwa deletion and morphology in Bengali)

Choudhury, M., Mukherjee, A., Basu, A. and Ganguly, N. 2006. Analysis and synthesis of the distribution of consonants over languages: a complex network approach. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Session.* ACL, Sydney, Australia, 128–135 (Modeling language change in the context of sound inventories using complex networks.)

Choudhury, M., Saraf, R., Jain, V., Sarkar, S. and Basu, A. 2007. Investigation and modeling of the structure of texting language. In *Proceedings of the IJCAI-07 workshop on Analytics of Noisy Unstructured Text Data (AND07).* Hyderabad, India, 63–70. (Model of language change over written medium from the perspective of application to SMS text normalization. The paper won the best student's paper award.)

Choudhury, M., Thomas, M., Mukherjee, A., Basu, A. and Ganguly, N. 2007. How difficult is it to develop a perfect spell-checker? a cross-linguistic analysis through complex network approach. In *Proceedings of HLT-NAACL Workshop – TextGraphs 2.* Rochester, NY, USA, 81–88.

Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N. 2007. Emergence of community structures in vowel inventories: an analysis based on complex networks. To be published in the *Proceedings of ACL SIGMORPHON9.* Prague, Czech Republic.

Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N. 2007. Redundancy ratio: an invariant property of the consonant inventories of the world's languages. To be published in the *Proceedings of ACL'07*. Prague, Czech Republic.

Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N. Forthcoming. Modeling the co-occurrence principles of consonant inventories: a complex network approach. Accepted for publication in *International Journal of Modern Physics C*.

Appendix B

List of all Publications by the Candidate

This is a list of all the publications by the candidate, including those on and related to the work of this thesis. The publications are presented in reverse chronological order under three sections – journals, refereed conferences and workshops, and other un-refereed publications. The articles are available online at

`http://cel.iitkgp.ernet.in/~monojit/publications.html`

B.1 Journal Publications

Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N. In press. Modeling the co-occurrence principles of consonant inventories: a complex network approach. Accepted for publication in *International Journal of Modern Physics C*.

Choudhury, M., Basu, A. and Sarkar, S. 2006. Multi-agent simulation of emergence of the schwa deletion pattern in Hindi. *Journal of Artificial Societies and Social Simulation (JASSS)*. 9(2). http://jasss.soc.surrey.ac.uk/9/2/2.html

Lahiri, A., Basu, A., Choudhury, M., and Mitra S. 2006. Battery aware power savings. *Electronics Systems and Software Magazine, IET*. IEE, April/May 2006.

Mukhopadhyay, A., Chakraborty, S., Choudhury, M., Lahiri, A., Dey, S. and Basu, A. 2006. Shruti – an embedded text-to-speech system for Indian languages. *IEE Proceedings on Software Engineering*, 153, 2, 75–79.

Choudhury, M. and Ray, P. R. 2003.Measuring similarities across musical compositions: an approach based on the raga paradigm. *Journal of ITC Sangeet Research Academy*. ITC-SRA, 17, 1–10.

B.2 Refereed Conferences and Workshops

Choudhury, M., Jalan, V., Sarkar, S. and Basu, A. 2007. Evolution, optimization and language change: the case of Bengali verb inflections. To be published in the *Proceedings of ACL SIGMORPHON9*. Prague, Czech Republic.

Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N. 2007. Emergence of community structures in vowel inventories: an analysis based on complex networks. To be published in the *Proceedings of ACL SIGMORPHON9*. Prague, Czech Republic.

Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N. 2007. Redundancy ratio: an invariant property of the consonant inventories of the world's languages. To be published in the *Proceedings of ACL'07*. Prague, Czech Republic.

Choudhury, M., Thomas, M., Mukherjee, A., Basu, A. and Ganguly, N. 2007. How difficult is it to develop a perfect spell-checker? a cross-linguistic analysis through complex network approach. In *Proceedings of HLT-NAACL Workshop – TextGraphs 2*. Rochester, NY, USA, 81–88.

Choudhury, M., Saraf, R., Jain, V., Sarkar, S. and Basu, A. 2007. Investigation and modeling of the structure of texting language. In *Proceedings of the IJCAI-07 workshop on Analytics of Noisy Unstructured Text Data (AND07)*. Hyderabad, India, 63–70.

Choudhury, M., Murguia, E., Sarkar, S., Moriceau, V., Kawtrakul, P. and Saint-Dizier, P. 2007. Generating instrumental expressions in a multilingual question-answering system. In *Proceedings of the IJCAI-07 workshop on Cross Lingual Information Access (CLIA 07)*. Hyderabad, India, 63–70.

Choudhury, M., Mukherjee, A., Basu, A. and Ganguly, N. 2006. Analysis and synthesis of the distribution of consonants over languages: a complex network approach. In *Proceedings of the COLING/ACL 2006 Main Conference poster Session*. ACL, Sydney, Australia, 128–135.

Lahiri, A., Basu, A., Choudhury, M., and Mitra S. 2006. Battery aware code partitioning for a text to speech system. In *Proceedings of Design Automation and Test in Europe Conference*. Munich, Germany, 672–677.

Choudhury, M., Alam, M., Sarkar, S. and Basu A. 2006. A rewrite rule based model of Bangla morpho-phonological change. In *Proceedings of the International Conference on Computer Processing of Bangla (ICCPB)*. Dhaka, Bangladesh, 64–71.

Das, D., Choudhury, M., Sarkar, S. and Basu A. 2005. An affinity based greedy approach towards chunking of Indian languages. In *Proceedings of the International Conference on Natural Language Processing (ICON)*. Kanpur, India, 55–62.

Lahiri, A., Choudhury, M., Chowdhury, S., Mukhopadhyay, A., Dey, S. and Basu A. 2005. A windows based file reader for Indian languages. In *Proceedings of the International Conference on Virtual Environments, Human-computer Interfaces and Measurement Systems (VECIMS)*. IEEE, Sicily, Italy.

Banerjee, A., Choudhury, M., Sarkar, S. and Basu, A. 2005. Learning schwa pronunceability rules in Bengali compound words using decision trees. In *Proceedings of the Second Symposium on Indian Morphology, Phonology and Language Engineering (SIMPLE)*. Mysore:CIIL, Kharagpur, India, 49–55.

Bhatt, A., Choudhury, M., Sarkar, S. and Basu, A. 2005. Exploring the limits of spellcheckers: a comparative study in Bengali and English. In *Proceedings of the Second Symposium on Indian Morphology, Phonology and Language Engineering (SIMPLE)*. Mysore:CIIL, Kharagpur, India, 60–65.

Das, D. and Choudhury, M. 2005. Finite state models for generation of Hindustani classical music. In *Proceedings of the International Symposium on Frontiers of Research in Speech and Music (FRSM)*. Bhubaneswar, India, 59–64.

Choudhury, M. and Ray, P. R. 2004. In search of musicality: can fractals show the way? In *Proceedings of the First World Conference on Lateral Computing (WCLC)*. WFLC, Bangalore, India.

Agarwal, A., Ray, B., Choudhury, M., Basu, A. and Sarkar, S. 2004. Automatic extraction of multiword expressions in Bengali: an approach for miserly resources scenario. In *Proceedings of the International Conference on Natural Language Processing (ICON)*. Allied Publishers, Hyderabad, India, 165–172.

Bansal, B., Choudhury, M., Ray, P. R., Sarkar, S. and Basu, A. 2004. Isolated-word error correction for partially phonemic languages using phonetic cues. In *Proceedings of the International Conference on Knowledge-based Computer Systems (KBCS)*. Allied Publishers, Hyderabad, India, 509–519.

Choudhury, M., Basu, A. and Sarkar, S. 2004. A diachronic approach for schwa deletion in Indo-Aryan languages. In *Proceedings of the Seventh Meeting of the ACL Special Interest Group in Comutational Phonology (SIGPHON)*. ACL, Barcelona, Spain, 20–26.

Annam, S., Choudhury, M., Basu, A. and Sarkar, S. 2003. ABHIDHA: an extended LexicalNet for Indo-Aryan languages. In *Proceedings of the 13th Meeting International Workshop on Research Issues on Data Engineering: Multi-lingual Information Management (RIDE-MLIM)*. IEEE, Hyderabad, India.

Choudhury, M. and Basu, A. 2002. A rule-based schwa deletion algorithm for Hindi. In *Proceedings of the International Conference on Knowledge-based Computer Systems (KBCS)*. Navi Mumbai, India, 343–353.

Basu, A., Sarkar, S., Chakraborty, K., Bhattacharya, S., Choudhury, M. and Patel, R. 2002. Vernacular education and communication tool for the people with multiple disabilities. In *Proceedings of Development by Design Conference (DYD)*. Bangalore, India.

B.3 Other Publications

Choudhury, M., Sarkar, S. and Basu, A. 2005. NLP activities at IIT Kharagpur. *Asia-Pacific Association for Machine Translation (AAMT) Journal*. September 2005, 22–23.

Choudhury, M., Sarkar, S. and Basu, A. 2005. Affects of inflectional and derivational morphology in Bengali phonology and its implications in speech technology. In *Proceedings of the Workshop on Morphology*. Mumbai, India.

Bhattacharya, B., Choudhury, M., Sarkar, S. and Basu, A. 2005. Inflectional morphology synthesis for Bengali noun, pronoun and verb systems. In *Proceedings of the National Conference on Computer Processing of Bangla (NCCPB)*. Dhaka, Bangladesh, 34–43.

Choudhury, M., Basu, A. and Sarkar, S. 2004. Schwa deletion and syllable economy in Indo-Aryan languages. In *Lecture Compendium of the first Symposium on Indian Morphology, Phonology and Language Engineering (SIMPLE)*. Kharagpur, India, 35–38.

Sarkar, S., Basu, A., Ray, P. R., Choudhury, M. and Bhattacharya, S. 2004. Machine translation and natural language generation framework using a semantic frame-based Interlingua. In *Lecture Compendium of the first Symposium on Indian Morphology, Phonology and Language Engineering (SIMPLE)*. Kharagpur, India, 95–98.

Choudhury, M. and Basu, A. 2002. A rule-based grapheme to phoneme mapping for Hindi. In *Abstracts of the Presentations by the Young Scientists: 90th Indian Science Congress of ISCA*. Bangalore, India.

Appendix C

The ITRANS Notation

The ITRANS notation has been designed to represent Indian language graphemes in ROman script (Chopde 2001). Since Indian languages are partially phonemic in nature, ITRANS can be used for transcribing the phonemes as well.

We provide the reference table for ITRANS and the corresponding default phonemes for each sympol in International Phonetic Alphabet (IPA)[1]. The first two columns of the table shows the graphemes of Hindi and Bengali in the Devanagari and Bengali scripts respectively. The graphemes are arranged according to the Indian language alphabet. The third column shows the ITRANS options for the grapheme. Note that there are multiple possibilities for representing some graphemes. In the last column, the default pronunciation of the graphemes in the two languages are shown in IPA.

Note that since the basic pronunciation of *a* is different in Bengali and Hindi, in the phonetic representation, the phonetic value of the symbol *a* has to be interpreted in the context of the language.

[1] http://www.arts.gla.ac.uk/IPA/

Script		ITRANS Transcoding	Default Pronunciation	
Devanagari	Bengali		Hindi	Bengali
Vowels				
अ	অ	a	ə, ʌ	ɔ, o
आ	আ	aa or A	a	a
इ	ই	i	i	i
ई	ঈ	ii *or* I	ɪ	i
उ	উ	u	u	u
ऊ	ঊ	uu *or* U	u	u
ऋ	ঋ	RRi *or* R^i	ri	ri
ॠ		RRI *or* R^I	ri	
ए	এ	e	e	e, æ
ऐ	ঐ	ai	æ	oi
ओ	ও	o	o	o
औ	ঔ	au	ɔ	ou
Consonants				
क	ক	k	k	k
ख	খ	kh	k^h	k^h
ग	গ	g	g	g
घ	ঘ	gh	g^h	g^h
ङ	ঙ	~N *or* N^	ŋ	ŋ
च	চ	ch	ç	ç
छ	ছ	chh	ç^h	ç^h
ज	জ	j	ɟ	ɟ
झ	ঝ	jh	ɟ^h	ɟ^h
ञ	ঞ	~n or JN	ɲ	ɲ
ट	ট	T	ʈ	ʈ
ठ	ঠ	Th	ʈ^h	ʈ^h
ड	ড	D	ɖ	ɖ
ढ	ঢ	Dh	ɖ^h	ɖ^h
ण	ণ	N	ɳ	n

Script		ITRANS Transcoding	Default Pronunciation	
Devanagari	Bengali		Hindi	Bengali
त	ত	t	t	t
थ	থ	th	t^h	t^h
द	দ	d	d	d
ध	ধ	dh	d^h	d^h
न	ন	n	n	n
प	প	p	p	p
फ	ফ	ph	p^h	p^h
ब	ব	b	b	b
भ	ভ	bh	b^h	b^h
म	ম	m	m	m
य	য	y	y	ɟ
	য়	Y		y
र	র	r	r	r
ल	ল	l	l	l
व		v *or* w	v	
श	শ	sh	ʃ	ʃ, s
ष	ষ	Sh *or* shh	ʃ, ʂ	ʃ, ʂ
स	স	s	s	ʃ, s
ह	হ	h	h	h
ड़	ড়	.D	ɽ	ɽ
ढ़	ঢ়	.Dh	$ɽ^h$	ɽ
Accents and Special Symbols				
अं	অং	a.n	represents homo-organic nasal	ŋ
अँ	অঁ	a.N	ə̃ (nasalized)	ɔ̃ (nasalized)
अः	অঃ	aH	ah	ah
क्	ক্	k.h	absence of schwa	absence of schwa

Appendix D

Bengali Verb Morphology

Bengali is an agglutinative language, and nouns and verbs inflect for a large number of attributes. The inflections are in the form of *suffixes*; suffixation usually does not change the structure of the root. In Bengali, a verb inflects for the following attributes.

- **Tense:** There are three basic tenses – *past*, *present* and *future*.
- **Aspect:** In past and present tenses, there are three aspects – *simple*, *continuous* and *perfect*. In future tense, only the *simple* aspect is realized by suffixation, whereas the other two are realized by auxilaries. There is also a specific *habitual* marker for the past tense, which we treat as an aspect of the past tense.
- **Modality:** Modality is realized by auxilaries and modal verbs. However, the *imperative* mood has dedicated inflections.
- **Person:** For every combination of tense, aspect and mood, Bengali has five different suffixes to mark for the person. There are three basic persons – *first*, *second* and *third*. However, there are three subcases of the *second* person, traditionally referred to as *familiar*, *normal* and *formal*. *Third person* also has two subcases – *honorofic* or *formal* and *non-honorofic* or *normal*. Nevertheless, the suffixes marking the *formal* cases for third and

second persons are identical, and need not be treated separately. Thus, we are left with five cases of persons.

- **Polarity:** The *negative polarity* is usually realized by the use of the *particle* "*nA*". However, in the cases of present perfect and past perfect, the negativizer is realized by the suffix "*ni*".

Thus, there are $10 \times 5 = 50$ inflectional suffixes for the finite forms of a verb. There are also four non-finite forms of a verb – *conjunctive*, *conditional*, *infinitive* and *gerund*. Furthermore, the nominal form of a verb or the gerund also takes the noun inflections. There are two emphasizer suffixes "*i*" (only) and "*o*" (also), that can be appended at the end of any inflected or uninflected noun, pronoun or verbs. In the two subsequent sections, we enlist the verb forms for the verb root "*kar*" (to do) in SCB and Classical Bengali. The last section discusses the morphological paradigms of Bengali verbs.

D.1 Verb Inflections in SCB

TAM,	Person				
Pol.,	1^{st}	2^{nd} Fam.	2^{nd} Norm.	2^{nd} Form.	3^{rd}
Pr-S.	*kor-i*	*kor-ish*	*kar-o*	*kar-en*	*kar-e*
Pr-C.	*kor-Chi*	*kor-Chish*	*kor-Cho*	*kor-Chen*	*kor-Che*
Pr-Pf.	*kor-eChi*	*kor-eChish*	*kor-eCho*	*kor-eChen*	*kor-eChe*
Pa-S.	*kor-lAm*	*kor-li*	*kor-le*	*kor-len*	*kor-lo*
Pa-C.	*kor-ChilAm*	*kor-Chili*	*kor-Chile*	*kor-Chilen*	*kor-Chila*
Pa-Pf.	*kor-eChilAm*	*kor-eChili*	*kor-eChile*	*kor-eChilen*	*kor-eChilo*
Pa-H.	*kor-tAm*	*kor-tish*	*kor-te*	*kor-ten*	*kor-to*
Fut.	*kor-bo*	*kor-bi*	*kor-be*	*kor-ben*	*kor-be*
Imp.	–	*kar*	*karo*	*kor-un*	*kor-uk*
Pf-N.	*kor-ini*	*kor-ishni*	*kar-oni*	*kar-enni*	*kar-eni*

Legends: Pol: polarity, Pr: present tense, Pa: past tense, Fut: future tense, S: simple, C: continuous, Pf: perfect, H: habitual, Imp: Imperative, N: negative, Fam: familiar, Norm: normal, Form: formal. Note that except for the case of imperative mood, all other inflections are for the indicative mood. Fut. is for simple aspect, and Pf-N. is for the past and present tenses. '-' indicates morpheme boundary.

All the forms are in phonetic representation. Note that although the root form is *kar*, and is represented as *kar* in orthography, in most of the cases, the root is modified as *kor*, which is the reminiscent of vowel height assimilation that took place during the evolution of these suffixes from classical Bengali forms. The non-finite forms of the verb are mentioned below.

- Conjunctive: *kor-e*
- Conditional: *kor-le*
- Infinitive: *kor-te*
- Nominal forms: *kar-A*, *kar-A-r* (oblique or possesive form), *kar-A-ke* (accusative form), *kar-A-te* (locative form)

D.2 Verb Inflections in Classical Bengali

TAM,	Person			
Pol.,	1^{st}	2^{nd} Norm.	2^{nd} Form.	3^{rd}
Pr-S.	*kar-i*	*kar-a*	*kar-ena*	*kar-e*
Pr-C.	*kar-iteChi*	*kar-iteCha*	*kar-iteChena*	*kar-iteChe*
Pr-Pf.	*kar-iYAChi*	*kar-iYACha*	*kar-iYAChena*	*kar-iYAChe*
Pa-S.	*kar-ilAma*	*kar-ile*	*kar-ilena*	*kar-ila*
Pa-C.	*kar-iteChilAm*	*kar-iteChile*	*kar-iteChilena*	*kar-iteChila*
Pa-Pf.	*kar-iYAChilAma*	*kar-iYAChile*	*kar-iYAChilena*	*kar-iYAChila*
Pa-H.	*kar-itAma*	*kar-ite*	*kar-itena*	*kar-ita*
Fut.	*kar-iba*	*kar-ibA*	*kar-ibena*	*kar-ibe*
Imp.	–	*kar-aha*	*kar-auna*	*kar-auka*
Pf-N.	*kar-i nAi*	*kar-a nAi*	*kar-ena nAi*	*kar-e nAi*

Legends: As above. Note that Classical Bengali makes a distinction between four persons. The second familiar form is absent. Moreover, since the exact pronunciations of these forms are debatable and variable over time and place, we have provided the orthographic forms used during the 19^{th} century. Also note that there was no negative marking suffix in Classical Bengali; rather the negative particle "*nAi*" was used to denote negative polarity.

The non-finite forms of the verb are as follows.

- Conjunctive: *kar-iYA*
- Conditional: *kar-ile*
- Infinitive: *kar-ite*
- Nominal forms: *kar-A*, *kar-ibAr* (oblique or possesive form), *kar-A-ke* (accusative form), *kar-A-te* (locative form)

D.3 Verb Paradigms in SCB

Although all the verbs in Classical Bengali inflected uniformly for all the suffixes, the verb roots of SCB undergoes differential orthographic and phonological changes during suffixation. On the basis of these changes, the verb roots of SCB can be classified into 19 morphological paradigms. In fact, the verbs having similar syllable structure belong to the same paradigm.

We can classify a verb root into a paradigm based on the final and penultimate vowels (if any) in its orthographic form. The classes for the roots with at least two vowels are shown below. (All the roots cited as examples are in orthographic forms.)

Penultimate	Final vowel(s)		
vowel	*a*	*A*	*oYA*
a	*kara* (do)	*karA* (do, *caus.*)	*saoYA* (bear, *caus.*)
A	*jAna* (know)	*jAnA* (inform)	*khAoYA* (feed)
i	*likha* (write)	*ni∼NrA* (twist)	–
e	*dekha* (see)	*dekhA* (show)	*deoYA* (give, *caus.*)
o	*tola* (pick)	*tolA* (pick, *caus.*)	*so;oYA* (lie, *caus.*)
u/au	–	*ghumA* (sleep)	–

'–' indicates that there is no known verb root pertaining to the paradigm. Apart from the 15 paradigms shown above, there are monosyllabic verb roots that can be classified into four paradigms based on the vowel. The representative roots for these paradigms are *sa* (bear), *khA* (eat), *de* (give), and *so* (lie down). There are also a few irregular verbs, which cannot be categorized in any of the 19-classes. These are *ha* (happen), *jA* (go) and *Asa* (come).

Appendix E

Data related to the MOGA Models

E.1 The Real Dialects

The inflected forms for the verb root *kar* is presented below for three modern dialects of Bengali along with the corresponding classical forms. Legends: Pr – present, Pa – Past, S – Simple, C – Continuous, P – Perfect, H – Habitual, 1 – first person, 2 – second normal person, 3 – third person, F – formal (in second and third persons)

Attributes	Classical (Λ_0)	SCB	ACB	Sylheti
PrS1	*kari*	*kori*	*kori*	*kori*
PrS2	*kara*	*karo*	*kara*	*kara*
PrS3	*kare*	*kare*	*kare*	*kare*
PrSF	*karen*	*karen*	*karen*	*karoin*
PrC1	*kariteChi*	*korChi*	*kartAsi*	*koirtrAm*
PrC2	*kariteCha*	*korCho*	*kartAsa*	*koirtrAe*
PrC3	*kariteChe*	*korChe*	*kartAse*	*koirtAse*
PrCF	*kariteChen*	*korChen*	*kartAsen*	*koirtAsoin*

Attributes	Classical (Λ_0)	SCB	ACB	Sylheti
PrP1	*kariAChi*	*koreChi*	*korsi*	*koirsi*
PrP2	*kariACha*	*koreCho*	*karsa*	*koirsa*
PrP3	*kariAChe*	*koreChe*	*karse*	*koirse*
PrPF	*kariAChen*	*koreChen*	*karsen*	*koirsoin*
PaS1	*karilAm*	*korlAm*	*karlAm*	*koirlAm*
PaS2	*karile*	*korle*	*karlA*	*koirlAe*
PaS3	*karila*	*korlo*	*karla*	*koirlA*
PaSF	*karilen*	*korlen*	*karlen*	*koirlAin*
PaC1	*kariteChilAm*	*korChilAm*	*kartAslAm*	*koirtesilAm*
PaC2	*kariteChile*	*korChile*	*kartAslA*	*koirtesilAe*
PaC3	*kariteChila*	*korChilo*	*kartAsla*	*koirtesilA*
PaCF	*kariteChilen*	*korChilen*	*kartAslen*	*koirtesilAin*
PaP1	*kariAChilAm*	*koreChilAm*	*korsilAm*	*koirsilAm*
PaP2	*kariAChile*	*koreChile*	*korsilA*	*koirsilAe*
PaP3	*kariAChila*	*koreChilo*	*korsila*	*koirsilA*
PaPF	*kariAChilen*	*koreChilen*	*korsilen*	*koirsilAin*
PaH1	*karitAm*	*kortAm*	*kartAm*	*koirtAm*
PaH2	*karite*	*korte*	*kartA*	*koirtAe*
PaH3	*karita*	*korto*	*karta*	*koirtA*
PaHF	*kariten*	*korten*	*karten*	*koirtAin*

E.2 Emergent Dialects

In this section, we provide examples of emergent dialects for the different MOGA models. Two sets of examples are selected for every model: the set **A** that best resembles the real dialects in terms of $\overline{sim_p}$ metric, and the set of **B** that seems linguistically implausible or are not known to exist. The fitness values for the dialects are also mentioned.

Note that for Models 1 to 4, the lexicon of each dialect consists of 18 forms, which are, in sequence, the corresponding forms for the following classical Bengali words.

> *kari, kara, kare, kariteChi, kariteCha, kariteChe, kariAChi, kariACha, kariAChe, karilAm, karile, karila, kariteChilAm, kariteChile, kariteChila, kariAChilAm, kariAChile, kariAChila*

The lexicon for Models 5 and 6 consists of 28 forms that are counterparts of the following Classical Bengali words (again, in sequence).

> *kari, kara, kare, karen, kariteChi, kariteCha, kariteChe, kariteChen, kariAChi, kariACha, kariAChe, kariAChen, karilAm, karile, karila, karilen, kariteChilAm, kariteChile, kariteChila, kariteChilen, kariAChilAm, kariAChile, kariAChila, kariAChilen, karitAm, karite, karita, kariten*

Also note that in all the experiments, we do not distinguish between the phonemes $/Ch/$ and $/s/$. This is because, a functional explanation of the change from $/Ch/$ (an affricate) to $/s/$ (a fricative) is based on the minimization of the number of gestures during articulation, modeling which is beyond the scope of this work.

Λ_1, Λ_2 and Λ_3 stand for the lexica of SCB, ACB and Sylheti respectively.

E.2.1 MOGA Model 1

Set A

Example A1.1
Closest to SCB, $\overline{sim}_p(\Lambda_1, \Lambda) = 2.44$
$f_e(\Lambda) = 104$, and $f_d(\Lambda) = 57.1$

> *kor karo krae koreCh koreCho kore karACh karACho karAe korAm karle karlo koreChAm koreChle koreChlo karAChAm karAChle karAChlo*

Example A1.2
Closest to ACB, $\overline{sim}_p(\Lambda_2, \Lambda) = 2.61$
$f_e(\Lambda) = 116$, and $f_d(\Lambda) = 47.6$

kar kara krae karetCh karetCha karete korChA korACha korAe karAlm karle karla karetChAlm karetChle karetChla korChAlm korChAle korChAla

Example A1.3
Closest to Sylheti, $\overline{sim}_p(\Lambda_3, \Lambda) = 3.72$
$f_e(\Lambda) = 141$, and $f_d(\Lambda) = 38.1$

koir kara krae koirteChi koirtChea koirteChe karChAi karACha karAChe koirAlm koirle koira koirteChiAlm koirteChile koirteChia karChAiAlm karChAile karChAia

Set B

Example B1.1
$f_e(\Lambda) = 80$, and $f_d(\Lambda) = 418.9$

kor kara kare kore korea kore kor kora kore korAm kore kora koreAm kore korea korAm kore kora

Example B1.2
$f_e(\Lambda) = 95$, and $f_d(\Lambda) = 65.2$

kor kara kar kore korea kore karA karAa karA korAlm korl korla koreAlm korel korela karAlm karAl karAla

Example B1.3
$f_e(\Lambda) = 141$, and $f_d(\Lambda) = 38.1$

kori kara krae koirteChi koirtChea koirteChe karChAi karACha karAChe koirAlm koirle koira koirteChiAlm koirteChile koirteChia kraChAiam karChAile karChAia

E.2.2 MOGA Model 2

Set A

Example A2.1
Closest to SCB, $\overline{sim}_p(\Lambda_1, \Lambda) = 2.78$
$f_e(\Lambda) = 102$, and $f_d(\Lambda) = 60.4$

> *kor kara kare koreCh koreCha koreChe karA karACha karAChe korAm korle korla koreChAm koreChle koreChla karAm karAle karAla*

Example A2.2
Closest to ACB, $\overline{sim}_p(\Lambda_2, \Lambda) = 3.83$
$f_e(\Lambda) = 128$, and $f_d(\Lambda) = 43.2$

> *kor karo kare karteCh karteo karteChe koriCh koriCho koriAe korilAm korile korio karteChilAm karteChile karteChio koriChilAm koriChile koriChio*

Example A2.3
Closest to Sylheti, $\overline{sim}_p(\Lambda_3, \Lambda) = 3.72$
$f_e(\Lambda) = 111$, and $f_d(\Lambda) = 51.7$

> *kor kara kare korteCh korteCha korte karChA karACha karAe korAm korle korla korteChAm korteChle korteChla karChAm karChAle karChAla*

Set B

Example B2.1
$f_e(\Lambda) = 99$, and $f_d(\Lambda) = 62.9$

> *kor kara kare korte kortea korte karA karAa karAe korAm korle korla korteAm kortele kortela karAAm karAle karAla*

Example B2.2
$f_e(\Lambda) = 96$, and $f_d(\Lambda) = 70.6$

> *kor kara kare kore korea kore karA karAa karAe korlAm korle korla korelAm korele korela karAlAm karAle karAla*

Example B2.3
$f_e(\Lambda) = 145$, and $f_d(\Lambda) = 38.0$

> *kor karo kare koritChe korietCho korietChe kairACh kairACho kairAe koriAm korile korilo koritCheiAm koritCheile koritCheilo kairAChiAm kairAChile kairAChilo*

E.2.3 MOGA Model 3

Set A

Example A3.1
Closest to SCB, $\overline{sim}_p(\Lambda_1, \Lambda) = 1.83$
$f_e(\Lambda) = 108$, $f_d(\Lambda) = 59.1$ and $f_r(\Lambda) = -0.674$

> *kor karo kare koreCh koreCho koreChe karACh karACho karAChe korlAm korle korlo koreChlAm koreChle koreChlo karAChlAm karAChle karAChlo*

Example A3.2
Closest to Sylheti, $\overline{sim}_p(\Lambda_2, \Lambda) = 2.78$, $\overline{sim}_p(\Lambda_3, \Lambda) = 2.56$
$f_e(\Lambda) = 132$, $f_d(\Lambda) = 47.5$ and $f_r(\Lambda) = -0.808$

> *kari kara kare kairtei kairteCha kairteChe koirChi koirCha koir-ChAe karilAm karile karila kairteilAm kairteile kairteila koirChilAm koirChile koirChila*

Set B

Example B3.1
$f_e(\Lambda) = 87,\ f_d(\Lambda) = 118.4,\ f_r(\Lambda) = -0.389$

> *kor kara kare kore korea kore kar kara kare korAm korle korla koreAm korele korela karAm karle karla*

Example B3.2
$f_e(\Lambda) = 98,\ f_d(\Lambda) = 63.5,\ f_r(\Lambda) = -0.427$

> *kor kara kare kore koreCha koreChe karACh karACha karAe korAm korle korla koreAm korele korela karAChAm karAChle karAChla*

Example B3.3
$f_e(\Lambda) = 124,\ f_d(\Lambda) = 51.1,\ f_r(\Lambda) = -0.810$

> *kar kara kare kairt kairta kairte karA karAa karAe korilAm korile korila kairtilAm kairtile kairtila karAilAm karAile karAila*

E.2.4 MOGA Model 4

Set A

Example A4.1
Closest to SCB, $\overline{sim}_p(\Lambda_1, \Lambda) = 2.44$, $\overline{sim}_p(\Lambda_2, \Lambda) = 3.16$, $\overline{sim}_p(\Lambda_3, \Lambda) = 3.61$
$f_e(\Lambda) = 124$, and $f_d(\Lambda) = 45.5$

> *kor kara kare korteCh kortea korteChe karChA karACha karAChe korilAm korile korila korteChilAm korteChile korteChila karChAilAm karChAile karChAila*

Set B

Example B4.1
$f_e(\Lambda) = 109$, and $f_d(\Lambda) = 61.3$

> *kori kara kare kairti kairta kairte kari kara kare korim korile korial kairtim kairtile kairtial karim karile karial*

Example B4.2
$f_e(\Lambda) = 131$, and $f_d(\Lambda) = 42.9$

> *kor kar kare koirte koirteCh koirteChe karACh karACh karAChe korilAm korile koril koirteilAm koirteile koirteil karAChilAm karAChile karAChil*

Example B4.3
$f_e(\Lambda) = 143$, and $f_d(\Lambda) = 39.3$

> *kor kara kare koirte koirteCha koirte koriACh koriChAa koriAe karliAm karlie karlia koirteilAm koirteile koirteila koriAChilAm koriAChile koriAChila*

E.2.5 MOGA Model 5

Set A

Example A5.1
Closest to SCB, $\overline{sim}_p(\Lambda_1, \Lambda) = 1.96$
$f_e(\Lambda) = 192$, and $f_d(\Lambda) = 89.4$

> *kori kora kore koren karChi karCha karChe karChen koriChi koriCha koriChAe koriChAen korilAm korile korila korilen karChilAm karChile karChila karChilen koriChilAm koriChile koriChila koriChilen kartAm karte karta karten*

Example A5.2
Closest to ACB, $\overline{sim}_p(\Lambda_2, \Lambda) = 2.43$
$f_e(\Lambda) = 181$, and $f_d(\Lambda) = 94.6$

kor kora kore koren karte karteCha karteChe karteChen karACh karACha karChAe karChAen korlAm korle korla korlen kartelAm kartele kartela kartelen karAChlAm karAChle karAChla karAChlen kartAm karte karta karten

Example A5.3
Closest to Sylheti, $\overline{sim}_p(\Lambda_3, \Lambda) = 4.07$
$f_e(\Lambda) = 209$, and $f_d(\Lambda) = 80.5$

koir kora kore koren kairChi kairCha kairChe kairChen koriAi koriChAa koriAChe koriAChen korilAm korile korila korilen kairChilAm kairChile kairChila kairChilen koriAilAm koriAile koriAila koriAilen kairtAm kairte kairta kairten

Set B

Example B5.1
$f_e(\Lambda) = 228$, and $f_d(\Lambda) = 75.9$

koir kora kore koren kairteChi kairteCha kairteChe kairteChen koriAi koriChAa koriAChe koriAChen korliAm korlie korlia korlien kairteChilAm kairteChile kairteChila kairteChilen koriAilAm koriAile koriAila koriAilen kairtAm kairte kairta kairten

Example B5.2
$f_e(\Lambda) = 163$, and $f_d(\Lambda) = 118.5$

kari kara kare karen kori kora kore koren karAi karAa karAe karAen karlAm karle karla karlen korlAm korle korla korlen karAlAm karAle karAla karAlen kortAm korte korta korten

E.2.6 MOGA Model 6

Set A

Example A6.1
Closest to SCB, $\overline{sim}_p(\Lambda_1, \Lambda) = 1.21$
$f_e(\Lambda) = 165$, and $f_d(\Lambda) = 114.3$

> *kor kora kore koren karChi karCha karChe karChen korChi korCha korChe korChen korlAm korle korla korlen karChilAm karChile karChila karChilen korChilAm korChile korChila korChilen kartAm karte karta karten*

Example A6.2
Closest to ACB, $\overline{sim}_p(\Lambda_2, \Lambda) = 1.71$
$f_e(\Lambda) = 191$, and $f_d(\Lambda) = 89.7$

> *kori kora kore koren karteChi karteCha karteChe karteChen koriChi koriCha koriChe koriChen karlAm karle karla karlen karteChilAm karteChile karteChila karteChilen koriChilAm koriChile koriChila koriChilen kartAm karte karta karten*

Example A6.3
Closest to Sylheti, $\overline{sim}_p(\Lambda_3, \Lambda) = 2.46$
$f_e(\Lambda) = 224$, and $f_d(\Lambda) = 76.8$

> *kori kora korA koren kairteChi kairteCha kairteChA kairteChen koriChAi koriACha koriAChA koriAChen kairlAm kairlA kairla kairlen kairteChilAm kairteChilA kairteChila kairteChilen koriChAilAm koriChAilA koriChAila koriChAilen kairtAm kairtA kairta kairten*

Set B

Example B6.1
$f_e(\Lambda) = 160$, and $f_d(\Lambda) = 128.9$

kar kora kore koren kari kara kare karen karAi karAa karAe karAen karlAm karle karla karlen karilAm karile karila karilen karAilAm karAile karAila karAilen kartAm karte karta karten

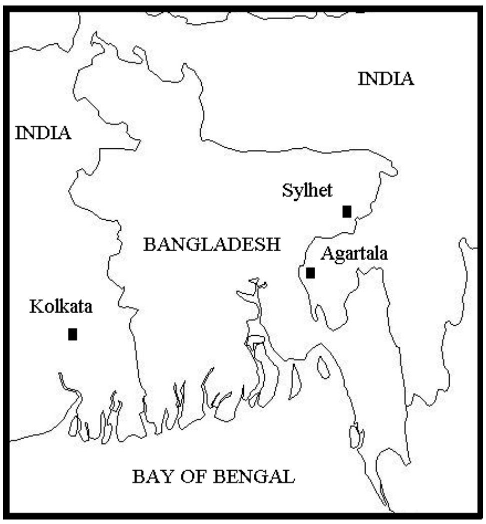
Example B6.2
$f_e(\Lambda) = 157$, and $f_d(\Lambda) = 206.3$

kar kara kare karen kartA kartAa kartAe kartAen karA karAa karAe karAen karA karle karla karlen kartAA kartAle kartAla kartAlen karAA karAle karAla karAlen kartA karte karta karten

Example B6.3
$f_e(\Lambda) = 197$, and $f_d(\Lambda) = 85.9$

kari kara karA karen korteChi korteCha korteChA korteChen kariChi kariCha kariChA kariChen karliem karliA karlia karlien korteChilem korteChilA korteChila korteChilen kariChilem kariChilA kariChila kariChilen kortem kortA korta korten

E.3 Map Showing the Geographical Locations of the Real Dialects



Index