Usage-based phonology and simulations as means to investigate unintuitive voicing behavior[*]

Jagoda Bruni, Daniel Duran and Grzegorz Dogil
Institute for Natural Language Processing, University of Stuttgart

Abstract

According to studies conducted by Coetzee & Pretorius (2010) and Rothenberg (1968), languages from the Sotho-Tswana group of Bantu languages demonstrate unintuitive voicing behavior in devoicing of post-nasal voiced plosives (/mb/ → [mp]) – unintuitive in that greater articulatory effort is required to terminate voicing than to maintain it (Westbury & Keating, 1986). Nasals preceding stop consonants are said to have appeared in Bantu languages in order to facilitate production of voicing during the stop segment and were lost later during language evolutionary changes in languages like Swahili, Sotho or Duala (Meinhof, 1932). Current studies on Tswana and Shekgalagari (Coetzee & Pretorius, 2010; Hyman, 2001; Solé et al., 2010), however, demonstrate that nasal segments remained in those languages – surprisingly not only before voiced stops but also before voiceless ones.

We present an attempt at using computational simulations on voicing behavior of Tswana post-nasal stops. Previous approaches to phonological simulations (e.g. Boersma & Hamann, 2008) put a strong emphasis on the functional bias and its role in language change. We base our investigations on the assumption that the role of social biases might play even a higher role in the formation and change of phonologically and phonetically driven sociolinguistic processes (Nettle, 1999; Coetzee & Pretorius, 2010).

1 Introduction

The phenomenon of post-nasal voicing behavior has been investigated from different perspectives. In their acoustic study involving Tswana native speakers, Coetzee & Pretorius (2010) present experimental data which provide evidence of active post-nasal devoicing. The authors describe measurements of Tswana post-nasal stops and report devoicing of these, arguing that one group of speakers applied aerodynamic and mechanical forces during the closure voicing, without employing any phonological rule.

Pater (1999) accounts for the *NT constraint, claiming that many languages demonstrate existence of prenasalized voiced stops but lack prenasalized voiceless stops. The rule penalizes consonantal sequences of [+nasal] followed by [−voice] and Pater (1999) claims that NC̥ clusters seem to be uncommon in a variety of languages. He states that typological data, as well as phonetic evidence argue for a universal but violable *NC̥ constraint.

Many African languages cope with this requirement in several ways. In Venda, Swahili or Maore, the nasal in NC̥ has been deleted or, like in OshiKwanyama, the post-nasal obstruent has become voiced (Pater, 1999; Meinhof, 1932). In another study, Hayes (1997) claims that an *NT constraint is phonetically driven, contrary to the corresponding *ND, which rules out sequences of a nasal followed by a voiced stop. Coetzee & Pretorius (2010) point out that given the phonetic naturalness of post-nasal voicing and phonetic unnaturalness of post-nasal devoicing, phonetic grounding of phonology would assume no language could exist with the phonological rule of post-nasal devoicing. Still, the phenomenon of post-nasal devoicing is clearly measurable and its diachronic spread in languages like Tswana has to be accounted for.

The approach presented in this paper is grounded in usage-based language accounts. As pointed out by Boersma & Hamann (2008), an exemplar-theoretic approach as proposed by Wedel (2006), for example, provides a link between simulated language inventories and their set of rules or phonological constraints. Usage-based grammatical structures are formed by experience through which particular instances

of constructions are categorized in memory based on their degree of similarities (Bybee, 2008). These units/exemplars depend on frequency of exposure and degree of usage. The more frequent ones form centers of categories and are easily accessible during language production, while the least frequent ones undergo memory decay (Bybee, 2008; Pierrehumbert, 2001). It has also been demonstrated (Pierrehumbert, 2001) that learning and remembering many exemplars enables better recognition of fine-grained phonetic processes. We incorporate the exemplar-theoretic approach in which categories of exemplars (in our case post-nasal stops) are composed of different voicing profiles and compete based on their scoring weights to become production targets.

The phonetic process of post-nasal devoicing has been analyzed also by Hayes & Stivers (2000). The authors implemented the computational model of Westbury & Keating (1986) based on previous work by Rothenberg (1968) and tested the hypothesis that part or all of the stop closure after a nasal is realized with vocal fold vibration. The results by Hayes & Stivers (2000) demonstrate that a post-nasal position of a stop facilitates its voicing. It confirms the hypothesis of Westbury & Keating (1986) that voicelessness requires additional articulatory cost, whereas voicing reflects a neutral state in post-nasal position. In that sense post-nasal devoicing is an unexpected or unintuitive process from the phonetic perspective, hence it appears unlikely that it can be described by a production bias.

The experiment of Coetzee & Pretorius (2010) demonstrated that the devoiced variant of post-nasal stop is not the only possible, although it is the dominant one (more than 80% of /m+bV/ and /m+pV/ sequences was realized as [m+pV]). The authors claim that this process is not categorical and that it is active phonologically, at least for part of the Tswana speakers. It is thus argued, that the devoicing tendency in Tswana post-nasal stops might result from historical language changes (described in more detail by Hyman, 2001) during which a general stop-devoicing process happened at a history stage where stops were observed only in the post-nasal environment. Coetzee & Pretorius (2010) argue that possible Tswana sound change is phonetic in origin but once it is phonologized, it turns out to be independent from phonetics. The phenomenon of post-nasal devoicing demonstrated in this study is argued to result from a function of the mechanical and aerodynamic forces for some speakers (where the closure of voicing has a limited duration and rarely matched with half of the total of the consonantal closure).

In our approach we consider various possibilities for the sound change currently occurring in Tswana. Our investigations are based on the data described above (Coetzee & Pretorius, 2010), as well as on documentation of historical changes in the Sotho-Tswana group proposed by Meinhof (1932) and Hyman (2001). It has been suggested by Brown et al. (2013) that some languages might demonstrate genetic relatedness by building sound correspondences across them. The authors suggest that an outgoing point in the research on genetically related languages should always be based on the consideration of word forms stemming from proto-languages. Another approach of investigating phonological sound change proposed by Johnson (2011) demonstrates the necessity to model language-internal processes through the observation of whole communities of speakers. Johnson suggests application of multi-agent simulations which replicate various social phenomena. He implements phonetic bias factors defined by elements like motor planning, gestural mechanics, speech aerodynamics and speech perception. The exemplar-theoretic background suggested by Johnson (2011) points out that phonetically biased variants and representations are re-used for production interacting with socially biased exemplars leading altogether to a sound change.

The investigation described in this paper is based on the notion of a usage-based approach to phonology. It can serve as an explanation for various sound changes. We simulate both social and functional biases and we propose an exemplar-based category formation (Pierrehumbert, 2001; Wedel, 2004).

Computer simulation studies provide a means of investigating models and testing hypotheses (cf. Liljencrants & Lindblom, 1972; Albright & Hayes, 2003; Duran, 2013). In particular, they can directly address empirically inaccessible phenomena. Moreover, only an implemented model (i.e. a simulation based on a computational model) can be tested and analyzed in all due detail. Boersma & Hamann (2008), for example, use computer simulations to investigate diachronic development of sibilant inventories. They successfully replicated the diachronic development of the Polish three sibilant system from a medieval state to its present-day configuration. Boersma & Hamann (2008) used exclusively functional phonetic bias to account for these developments.

Nettle (1999) uses computer simulations of Social Impact Theory (Cavalli-Sforza & Feldman, 1981; Boyd & Richerson, 1985) to investigate language change under varying conditions of social interaction and acquisition biases within a population. Wedel (2004) presents computer simulations of category competition
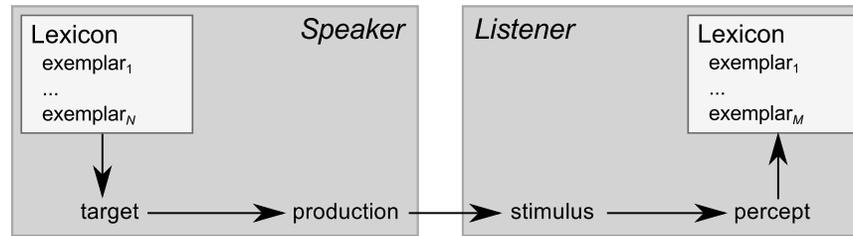
Figure 1: Distinction between different kinds of exemplars according to their place in the production–perception chain from a speaker to a listener.

in an exemplar-theoretic framework. He argues that contrast is not "a property of forms" but that it can be described rather in an exemplar-theoretic framework as being implicitly driven by the statistical association of forms to categories.

In our work we adapt and combine the methods proposed Nettle (1999) and Wedel and his colleagues (Wedel, 2004, 2006, 2012; Blevins & Wedel, 2009; Wedel & Van Volkinburg, unpublished) by modeling competition between variants undergoing functional and social selection during language acquisition over many generations. We show that modeling voicing profiles (Möbius, 2004; Bruni, 2011), which can be extracted from labeled data bases, can be achieved by assigning *functional* and/or *social biases* to such processes as sonorant devoicing in obstruent context.

With our simulation experiments we investigate the influence of various parameters and compare the results against the currently reported voicing behavior in Tswana.

The aim of this work is to apply exemplar-based phonetic simulations in order to investigate factors influencing post-nasal devoicing and its evolution over time. Exemplar Theory (Goldinger, 1997; Lacerda, 1995; Pierrehumbert, 2001) assumes that language production and perception are tightly linked. Percepts of linguistic experiences are stored in the mental lexicon with their concrete forms, including for example phonetic detail. It is claimed that language use plays a crucial role in the formation of the sound system. In this sense, phonological rules stem from generalizations of representations of directly used forms (Bybee, 2001). Such a usage-based approach to language analysis presumes also categorical storage of exemplars, where frequency of occurrence and activation determines successful storage of a phonological item and its role in speech production/perception and language acquisition (Feldman et al., 2013).

2   Model & Parameters

Our hybrid modeling framework combines an exemplar-theoretic model developed by Wedel and colleagues[1] and a sociolinguistic model developed by Nettle (1999). In this section, we introduce this proposed modeling framework with its relevant terminology and describe the different parts of the model. In the following section we present a first simulation study investigating the influence of various model parameters and discuss the results in comparison to the reported "unintuitive" voicing behavior in Tswana.

In our discussion of the modeling framework, there are two competing forms representing free variants of the same phonemic category. According to the discussion in the previous section, we refer to one of these forms as the *intuitive* variant and to the other form as the *unintuitive* variant.

2.1   *A population of agents.*   The framework presented in this paper is based on an *agent-based model*, where individual members of a language community, or *population*, are represented by *agents*. These are independent, autonomous entities within the model which are defined by internal states and methods of interacting with other agents[2].

The population of agents is embedded within a *social network*. This network defines social relations

---

[1]   In particular, we follow the model description from Wedel & Van Volkinburg (unpublished).
[2]   In our case, the agents are embedded within a social network but do not explicitly interact with their environment as it is sometimes the case in agent-based models.

between the members of the community. Formally, it can be represented as a graph, where nodes correspond to individual agents and edges between the nodes correspond to direct social relations between two agents. A *social distance* between two given agents may then be defined as the minimal path length (i.e. the minimum number of edges) between the two corresponding nodes. Note that such a definition does not correspond to geographical distance. The particular arrangement of agents within the social network, i.e. the *network topology*, is not determined by the model. Thus, network topology represents another model parameter which needs to be investigated by studying its effects on the behavior of the system.

In addition to social relations between agents, we assume that individuals are characterized by a *social status*[3]. Different status levels correspond to different degrees of influence which individuals exert on others within their community. Some individuals in our model have a disproportionately higher status than the average individual. Adopting terminology by Nettle (1999), we call these *hyper-influential* individuals. For the sake of simplicity, we assume that the social status is an inherent property of an individual which does not change over time.

2.2 *Social interactions.*   The agents of the population act both as *speakers* and *listeners*[4]. The scheme, according to which agents interact, constitutes yet another model parameter. From the point of view of a listener, the interaction scheme represents a *sampling* of input speech items (or, implicitly, of speakers) from the collective productions of the population. The influence of different interaction schemes on the evolution of the system may be studied by defining appropriate rules of interaction.

2.3 *Aging.*   In our model, agents are born, they go through a number of *life stages* and finally die. This is in keeping with the model proposed by Nettle (1999). Note, however, that Nettle (1999:97) implicitly adopts the "critical period" hypothesis by assuming that individuals may acquire a particular variant only during their first two life stages. After initial learning and re-evaluation, the acquired variant is fixed and no further change occurs for that agent. We do not assume the existence of such a critical period of language acquisition. In our model, we assume life-long learning and accommodation to the linguistic environment.

In Nettle's model, learners are influenced by all other members of the population except for individuals in their first life stage. We adopt this view and assume that agents in their first life stage do not speak (in the sense that they do not produce relevant linguistic input for other members of the population). The model proposed by Wedel & Van Volkinburg (unpublished) does not incorporate explicit aging of its individual agents. Each agent learns in each epoch of the simulation by updating its lexicon. We adopt this usage-based approach and assume that each agent is a listener in every life stage. Within our exemplar-theoretic framework, this corresponds to continuous learning by adding new linguistic experience to the lexicon. However, at some point, each agent is removed from the population as it finishes its final life stage.

2.4 *Exemplars and Lexicons.*   We assume an exemplar-theoretic organization of an individual's mental lexicon, i.e. categories are represented by collections of remembered speech items (*exemplars*). Our model implements a strict interpretation of Exemplar Theory: Speech production and perception are modeled at the level of individual exemplars. Each agent has an individual internal memory of previously perceived exemplars which is referred to as its *lexicon*. This is in keeping with Wedel's approach but in contrast with Nettle's model which computes averages over the entire population.

Within an exemplar-theoretic model, the term *exemplar* may refer to different instances of linguistic objects. Figure 1 shows the different terms we use to refer to various kinds of exemplars and it visualizes their relations within the production–perception chain from speaker to listener. Note that each agent has its own lexicon with potentially different numbers of exemplars stored in it.

Pierrehumbert (2001:140) argues that labels may "constitute a level of representation in their own right" and she points out that exemplars may be "subject to more than one categorization scheme". We adopt this view and represent exemplars in our model by *sets* containing various types of information. We assume that exemplars contain social information like the social status of the originating speaker and the corresponding social closeness (see below). We also assume a phonetically rich representation of exemplars retaining

---

[3] Johnson (2011) speaks of "social empowerment" and includes a social factor in his simulations which he refers to as "social identity".

[4] Often, the terms 'teachers' and 'learners' are used to distinguish these different roles of agents (cf. Kirby & Sonderegger, 2013).

perceptual detail. This is similar to the approaches by Nettle (1999), Wedel & Van Volkinburg (unpublished) or Kirby & Sonderegger (2013), who also incorporate *continuous phonetic cues*. The continuous phonetic information may be formally represented by n-dimensional real vectors (where n is the number of phonetic cues) and the social information may be formally represented by real numbers indicating social status and social closeness. For the current implementation of the framework presented here, we assume that the intuitive and the unintuitive variants belong to the same phonemic category. Therefore, there is no category label attached to exemplars, and there is no categorization involved in perception. This, however, is not a general restriction of the model. It might be easily adjusted to incorporate a categorization procedure, in which case various linguistic labels may the attached to individual exemplars.

The lexicon *capacity* is one of the model parameters that needs to be systematically investigated in simulation studies. Two fundamental distinctions regarding the lexicon capacity can be made: (1) having a lexicon with a limited capacity and (2) having an unlimited capacity. The apparent enormous amount of stored exemplars postulated by Exemplar Theory often faces criticism. This *head-filling-up problem* is addressed by Johnson (1997). The term refers to the seeming impossibility of an ideal exemplar model which would require storage of all perceived items. This problem is partially mitigated by reference to the observed human ability of remembering very large numbers of pictures. Johnson (1997) assumes additionally that the perceptual space may be quantized based on just noticeable differences, thus reflecting the fact that humans cannot perceive arbitrarily small differences along auditory dimensions. In addition to quantized representations, Pierrehumbert (2001) assumes that memories decay over time which leads to the loss of old or non-activated exemplars. The simulations presented by Wedel (2004) avoid this memory problem by deleting one exemplar from the memory for each new addition. Likewise, we assume that old exemplars are removed for new additions, once the number of exemplars in a lexicon reaches its capacity.

2.5 *The production–perception chain.* One central aspect of production in our hybrid model is the process of *target selection* by which one particular exemplar is selected from the lexicon as a production target (cf. Fig. 1). Target selection in our model is formally based on a *scoring function* over the set of stored exemplars. The score is a weighted sum with three components. For an exemplar x_i it is defined as follows:

$$\text{score}(x_i) = \frac{\alpha \, \text{sim}(x_i, x_z) + \beta \, \text{status}(x_i) + \gamma \, \text{closeness}(x_i)}{\alpha + \beta + \gamma} \tag{1}$$

where:

- $\text{sim}(x_i, x_z)$ is the phonetic similarity of the i-th exemplar to the centroid x_z of the lexicon (in a multi-category model, the centroid of the corresponding cluster may be used instead). We define phonetic similarity according to Equation (4a) defined by Nosofsky (1986:40) (cf. Wedel, 2006:266; and Johnson, 1997:147):

$$\text{sim}(x_i, x_z) = \mathrm{e}^{-d_{iz}} \tag{2}$$

where d_{iz} is the Euclidean distance between two exemplars x_i and x_z (treated as n-dimensional vectors):

$$d_{iz} = \sqrt{\sum_{k=1}^{n} |x_{ik} - x_{zk}|^2} \tag{3}$$

The values of the phonetic similarity are limited to the interval $[0.0, 1.0]$ where a value of 1.0 means that the two items are maximally similar, i.e. the same. This component of the scoring function favors *prototypical* representatives of the lexicon by assigning high scores to exemplars which are close to the center of the set of exemplars.

- $\text{status}(x_i)$ is the social status attached to the i-th exemplar. This is the (perceived[5]) social status of the original speaker of that exemplar. If social status levels are limited to a fixed range, the values can

---

[5] Note that social status might plausibly be modeled as a *subjective* property of an individual. For the sake of simplicity, the perceived social status of a given speaker is the same for all listeners in our framework.

network topology	interaction full	by status	by closeness
regular grid	Simulation 1	Simulation 3	Simulation 5
small world	Simulation 2	Simulation 4	Simulation 6

Table 1: Simulations according to social network topology and interaction schemes.

be limited to the interval $[0.0, 1.0]$, without loss of generality. A value of 1.0 is the highest possible status assigned only to hyper-influential agents. This component of the scoring function favors forms originally produced by influential speakers with a high social status.

- closeness(x_i) is the social closeness of the originating speaker of the i-th exemplar to the listener. It is computed according to the following formula:

$$\text{closeness}(x_i) = 1 - \frac{d_{\text{speaker,listener}}}{d_{\text{max}}} \quad (4)$$

where $d_{\text{speaker,listener}}$ is the social distance between the speaker who produced exemplar x_i and the listener who stored exemplar x_i in her lexicon (i.e. the minimum number of edges between the two nodes representing the individuals within the social network graph). Distance is always ≥ 0, with 0 only for the case where $\text{speaker} = \text{listener}$ (corresponding to self-edges in the graph, which in our current implementation are forbidden). The normalizing quantity d_{max} is the maximum distance within the network (which is always a finite value for a finite number of agents). The closeness values are limited to the interval $[0.0, 1.0]$ where 1.0 is the highest possible value corresponding to the closest individual (i.e. the ego). Note also, that this formula gives $\text{closeness} = 0$ for the maximum distance.

The overall score assigned to an exemplar x_i is thus also limited to a value in the interval $[0.0, 1.0]$. The weights α, β and γ are model parameters which need to be set. Based on this scoring function, one exemplar is selected probabilistically from the lexicon as a production target. A small amount of random noise is added to the phonetic cues of the selected target and the produced exemplar is then transmitted from the speaker to the listener representing its input stimulus.

On the perception side, there is another transformation from stimulus to the actual percept (cf. Fig. 1). Pierrehumbert (2001) points out that a model with production noise leads to a steady increase in variance over time. However, *entrenchment* rather than continued spreading of categories can be observed in actual language use. Entrenchment in our model is promoted by the incorporation of a *perceptual bias* corresponding to the *perceptual magnet effect* (Kuhl, 1991). This perceptual bias warps stimuli slightly towards local maxima of exemplar distributions in the lexicon. Following Wedel (2006)[6] this transformation is applied as a low-level auditory process prior to (potential) categorization. Note, that this contrasts with the exemplar-theoretic model of the perceptual magnet effect proposed by Lacerda (1995) who takes explicit category information into account.

3   Simulations

We investigate the phenomenon of unintuitive voicing behavior in Tswana employing our novel hybrid modeling framework. In this section, we discuss a preliminary simulation study and present some observations. Considering different combinations of network topology and interaction schemes, we present in total six different simulations. Table 1 gives an overview of these different simulation types. All six simulations are modifications of the basic setup with a modified network topology or interaction scheme.

---

[6] Wedel (2006) refers to the neural network model proposed by Guenther & Gjaja (1996) who claim that the effect is a consequence of the formation of non-uniform neural maps in the auditory system based on language experience.

3.1 *Population and lexicon settings.*   The population consists of a fixed number of agents each of which has the following properties: (1) an age which ranges between 0 and 4, (2) a social status between 0 and 1, and (3) a lexicon. Each agent gets an initial age (uniformly distributed over the population) and a randomly assigned social status. Initially, we assume a variant distribution such that the unintuitive variant represents a rare variant and the majority of the population has the intuitive variant. Agents of age 0 start with an empty lexicon. Others are initially seeded with a number of noisy copies of the intuitive variant. The initial phonetic cue values are taken from real data (voicing profiles) reported by Coetzee & Pretorius (2010). A small number of agents is assigned a much higher social status than the average agent. These are the hyper-influential individuals in the population. Their lexicons are seeded with a certain percentage of noisy copies of the unintuitive variant (such that this represents their majority variant).

In the social impact study by Nettle (1999), agents learn in their first two life stages. We adopt this premise by setting the lexicon capacity to 800. This has the effect that during the first two epochs (at least in Simulation 1 and Simulation 2), all perceived exemplars are stored in the lexicon, and no decay takes place. Once the lexicon size (i.e. the total number of stored exemplars) reaches capacity, *lexicon decay* comes into play. Every time a new exemplar is added beyond that point, the oldest exemplar is removed from the lexicon. Thus, in contrast to Nettle's model, learning does not stop after an initial acquisition phase. In order to investigate the effect of an unlimited lexicon capacity, the limit parameter has been set to a large value > 2000 in modification of the basic simulations. In Simulation 1 and Simulation 2 this has the effect that during the five life stages of the agents, lexicon capacity is never reached, and thus, no decay occurs.

3.2 *Simulation 1.*   In this simulation, we follow Nettle (1999) and arrange the agents along a regular grid with 20 columns and 20 rows with each node in the network representing one agent. In order to avoid edge-effects, the network takes on a closed toroidal topology. Thus, each agent has exactly four direct neighbors. The network is connected, i.e. there is a path between any two agents in the network.

The simulation runs for a specified number of epochs. In each epoch, each agent interacts with everybody else in the population. Speaker and listener orders are randomized each time. We refer to this interaction scheme as *full interaction*. At the end of an epoch, the age of each agent is increased by one. Agents who reach the age of five are replaced by new agents with age zero. If the removed agent was a hyper-influential individual, the new one will also have a hyper-influential status (however, all new agents start with an empty lexicon).

3.3 *Simulation 2.*   The social network in this simulation is not a regular grid but a *small world* network (Milgram, 1967) which is build according to the random rewiring procedure proposed by Watts & Strogatz (1998). A 20×20 toroidal network is initialized as in Simulation 1 and then transformed such that it constitutes a small world network. The effect of this network is that there is a smaller average distance between any pair of agents, while the average number of direct neighbors for each agent is still the same. This network topology is closer to real social networks.

3.4 *Simulation 3 and 4.*   In these simulations, interaction between agents is based on social status. In each epoch, everybody interacts with other agents depending on her social status with uniformly distributed random listeners. The total number of listeners for a given speaker depends on its social status. We multiply the status by the size of the population, such that hyper-influentials with a status of 1.0 will effectively speak to everybody else and other agents will speak to only a fraction of the population. This interaction scheme implements a social bias towards individuals with a high social status.

3.5 *Simulation 5 and 6.*   In these simulations, interaction between agents is based on social closeness. In each epoch, everybody interacts with other agents depending on their social closeness. A number of listeners is probabilistically selected according to the their social closeness to the speaker (with some agents potentially being spoken to multiple times by the same speaker in one epoch). This interaction scheme implements a social bias towards individuals within the close neighborhood around an individual.

3.6 *Evaluation.*   As the employed hybrid modeling framework belongs to the family of multi-agent simulations, the observed dynamics of the modeled system are not the result of globally defined rules but are an emergent phenomenon based on the multitude of agent interactions.
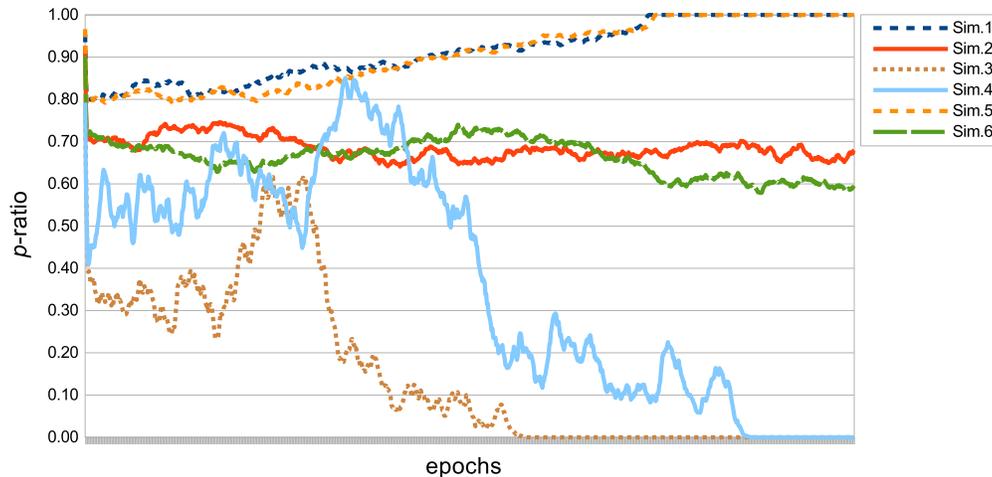
Figure 2: Proportion of plain variant productions for different network topologies and interaction schemes with equal scoring weights and a limited lexicon capacity over 1,000 epochs.

Adding auditory warping results in much more effective entrenchment than relying on the selection bias towards more central exemplars alone. This conclusion is based on statistics about the individuals' lexicon contents, where a low standard deviation from the mean can be observed with perceptual magnets, as well as "tight" minimum and maximum values of the phonetic dimensions (not shown). This confirms Pierrehumbert's (2001) argument on entrenchment. Based on this initial finding, we incorporate auditory warping as described in the previous section to all simulations reported here.

One illustrative way of evaluating the performance of the simulations is to look at the totality of productions in each epoch and record its evolution. Since the intuitive variant is initially the majority form in the population, we will refer to it as the "plain" variant, for simplicity. We determine the ratio of productions of plain variants out of all produced exemplars over the entire population per epoch and refer to this quantity as the *p-ratio*. Thus, a p-ratio of 1.0 corresponds to the case where all produced exemplars in one epoch (across all agents) are instances of the plain, i.e. phonetically intuitive, voiced variant.

Once all produced exemplars are instances of one variant, we refer to this case as the *loss* of one variant. Note that in our implementation of the framework, the first epoch is detected where such a loss occurs. At this point, however, the simulation is not interrupted but it continues for several additional epochs. Only if the apparently lost variant does not re-surface for a number of epochs, the simulation is aborted and the loss is recorded as the outcome of that run[7].

Figure 2 shows individual results for all six simulations[8]. Three qualitatively different outcomes can be observed: The model may result in an apparently stable state where both variants co-exist in the population. This result is observed for the small world networks with full and closeness-based interactions. The other two outcomes correspond to the loss of either variant, as indicated by a p-ratio of exactly 0 or 1, respectively.

Figure 3 shows a comparison of different scoring weights in Simulation 1. Only four qualitatively different settings are shown: First, having equal weights on the three components of the scoring function as in figure 2. Additionally, we set the parameters α, β and γ in turn to a high value while keeping the remaining two parameters at an equally low value. The results indicate that a large α, i.e. a large weight on phonetic similarity, appears initially very similar to the case with equal weights on all three exemplar

---

[7] Another way of evaluating the simulations would be to compute statistics over the contents of agents' lexicons. For the present study we restrict ourselves to the population-level evaluation of productions. It should be kept in mind, however, that an observed "loss" in the population's collective productions does not necessarily correspond to a complete loss of memorized exemplars of that variant.

[8] Note that due to the random factor in the simulation, consecutive runs with the same settings will never be identical.
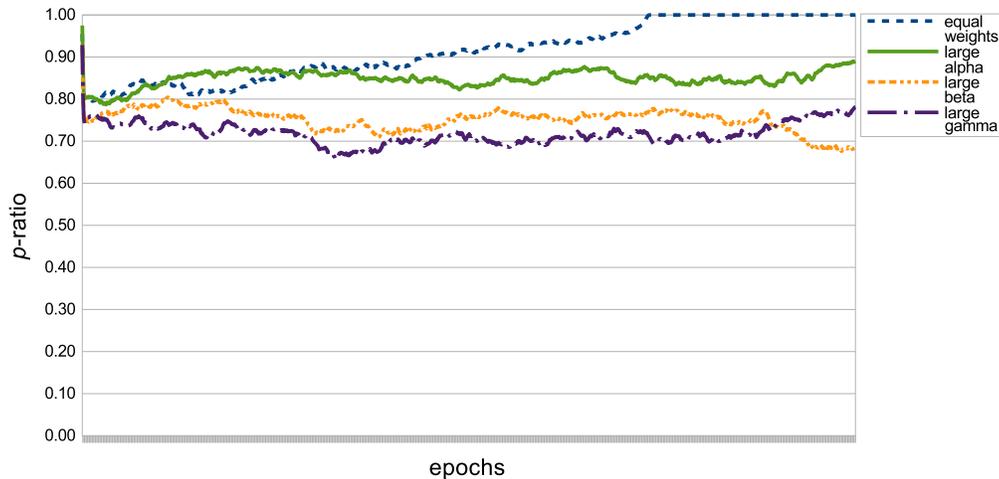
Figure 3: Proportion of plain variant productions for different scoring weights with a regular grid topology, full interaction and a limited lexicon capacity over 1,000 epochs.

components but does not lead to a loss of the unintuitive variant. A large β or γ, i.e. a large weight on status or closeness, respectively, leads to a faster spread and stabilization of the unintuitive variant within the population, indicated by lower p-ratios.

Figure 4 shows a comparison between Simulations 1 and 2, both with limited and unlimited lexicon capacities. These results indicate that a limited lexicon enhances the population dynamics, while the two simulation runs with an unlimited lexicon fall between the extremes with p-ratios in-between the two simulation runs with a limited lexicon capacity. It seems that forgetting promotes change in the lexicon.

4   Future Work

Future work will require training of the various parameters fitting the system's behavior to the empirical data – in our Tswana case, for example, the goal might be to achieve a proportion of intuitive vs. unintuitive forms according to the empirically observed proportions of these variants (Coetzee & Pretorius, 2010). This optimization problem will be addressed by applying stochastic gradient descent to learn the weights. We expect to obtain more realistic results for the synchronic data and we will attempt to find the modeling framework for the diachronic developments.

Moreover, as a further step we would like to address categorization processes subject to conflicting biases. According to Coetzee et al. (2007), nasals and stops in the NC clusters are always separated by a morpheme boundary which indicates morphological number. It is suggested that post nasal devoicing could occur in order to add perceptual saliency between the two segments and make the distinction between clitics and verb roots more prominent. Knowing that the voiced stop post-nasally is a favored voicing profile from the articulatory point of view and that the voiceless stop is generally more natural and preferred from a perceptual view, it would be worth simulating how these two conflicting profiles evolve over time.

5   Conclusion

Our study describes usage-based simulations of phonetically unexpected post-nasal stop devoicing in Tswana. Westbury & Keating (1986) claim that production of voicing in stops is a paradox. Despite the higher articulatory effort required to produce them (voicelessness seems more "natural"), voiced stops are widely spread among many languages. Such phonetically unintuitive distributions in the world's languages are hard to explain by means of natural phonological rules (Keating, 1988; Yip, 1988), or in case of Exemplar
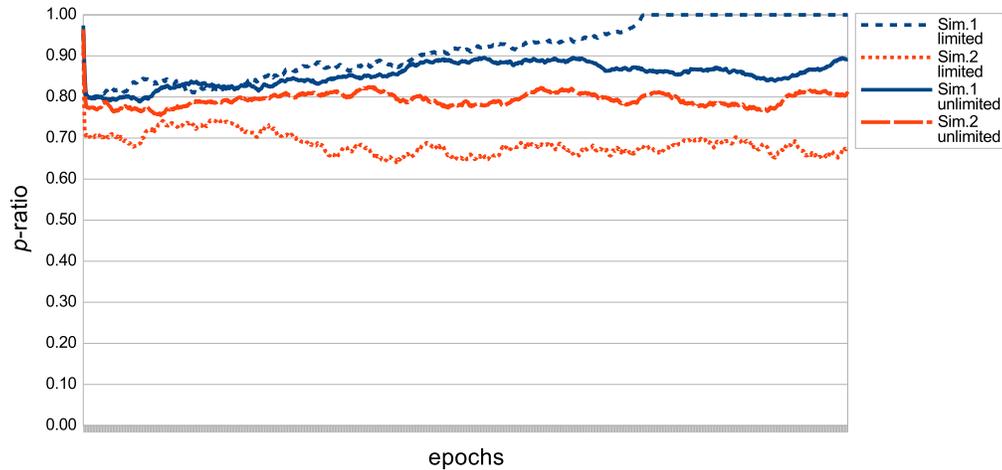
Figure 4: Proportion of plain variant productions for different lexicon capacities with full interaction and equal scoring weights over 1,000 epochs.

Theory, by means of a functional phonological bias. In the case of Tswana post-nasal clusters, the unintuitive voicing behavior might be grounded in sociolinguistic reason like a prominence bias for social status of individuals using rare devoiced variants. In our implementation of the sociolinguistics we use the well-understood Social Impact Theory (Cavalli-Sforza & Feldman, 1981; Boyd & Richerson, 1985; Nettle, 1999) which clearly models various social statuses of individuals.

This present approach is limited to two network models of a *regular grid* and a *small world* structure, which simulate closeness vs. social distance of speakers within a community. The implementation of learning in our sociolinguistic model consists of five life stages (infancy and four stages of adolescence and adulthood). In Nettle's (1999) model, learning is limited to just a few stages. However, from the general point of view, a limitation in the learning phonological contrast does not have to be true, at least not for all learners, since some speakers/listeners are more talented than others (Dogil & Reiterer, 2009). It has also been shown by Harrington (2006) that even speakers with an extremely high social impact may adapt their phonetic forms during their life time. Thus, the existence of a particular critical period for phonetic/phonological learning is controversial.

As for the functional bias, we use perceptual warping of generated voicing profiles in order to model the bias in distribution of voiced and devoiced post-nasal stops. On the other hand, it might be an interplay of the morphophonological boundary which favored the post-nasal stops to be more perceptually salient.

As proposed by several previous accounts (Pater, 1999; Hayes, 1997; Hayes & Stivers, 2000), the phenomenon of post-nasal devoicing cannot be generalized to a phonological rule but rather it appears in numerous languages as a contrast to the *NT constraint. The exemplar-theoretic usage-based approach (Bybee, 2008; Pierrehumbert, 2001) which we employ in our simulations generates a link between phonological analysis and phonetic grounding. Thus, on one hand speaking agents use the capacity of their mental lexicon and availability of exemplar categories during speech production, and on the other hand they deal with functional, phonetically-driven biases as well as incoming interplay of social factors which are employed within the network interactions.

This novel hybrid framework combines social modeling principles proposed by Nettle (1999) and self-organizing exemplar-theoretic dynamics proposed by Wedel (2006) in a framework facilitating agent-based simulations of language change. To our knowledge, this is the first model that combines these approaches into a single investigative framework. Our preliminary study already highlights the potential benefits of the hybrid model.

References

Albright, Adam & Bruce Hayes (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90:2, 119–161.

Blevins, Juliette & Andrew Wedel (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronia* 26:2, 143–183.

Boersma, Paul & Silke Hamann (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25, 217–270.

Boyd, Robert & Peter J. Richerson (1985). *Culture and the evolutionary process*. University of Chicago Press, Chicago, IL.

Brown, Cecil H., Eric W. Holman & Søren Wichmann (2013). Sound correspondences in the world's languages. *Language* 89:1, 4–29.

Bruni, Jagoda (2011). *Sonorant voicing specification in phonetic, phonological and articulatory context*. Doctoral dissertation, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Bybee, Joan (2008). Usage-based grammar and second language acquisition. Robinson, Peter & Nick C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, Routledge, NY, 216–236.

Bybee, Joan L. (2001). *Phonology and Language Use*. Cambridge University Press, Cambridge.

Cavalli-Sforza, Luigi Luca & Marcus W. Feldman (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press, Princeton, NJ.

Coetzee, Andries W. & Rigardt Pretorius (2010). Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38:3, 404–421.

Coetzee, Andries W., Susan Lin & Rigardt Pretorius (2007). Post-nasal devoicing in Tswana. *Proceesdings of ICPhS XVI*, Saarbrücken, 861–864.

Dogil, Grzegorz & Susanne Reiterer (eds.) (2009). *Language Talent and Brain Activity*. Mouton de Gruyter.

Duran, Daniel (2013). *Computer simulation experiments in phonetics and phonology: simulation technology in linguistic research on human speech*. Doctoral dissertation, Universität Stuttgart, URL http://elib.uni-stuttgart.de/opus/volltexte/2013/8789.

Feldman, Naomi H., Thomas L. Griffiths, Sharon Goldwater & James L. Morgan (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review* 120:4, 751–778.

Goldinger, Stephen D. (1997). Words and voices – perception and production in an episodic lexicon. Johnson & Mullennix (1997), chap. 3, 33–66.

Guenther, Frank H. & Marin N. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100:2, 1111–1121.

Harrington, Jonathan (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34:4, 439–457.

Hayes, Bruce (1997). Phonetically driven phonology. The role of optimality theory and inductive grounding. Manuscript written for the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics.

Hayes, Bruce & Tanya Stivers (2000). Postnasal voicing, URL http://www.linguistics.ucla.edu/people/hayes/Phonet/NCPhonet.pdf. Manuscript.

Hyman, Larry M. (2001). The limits of phonetic determinism in phonology: *NC revisited. Hume, Elizabeth & Keith Johnson (eds.), *The role of speech perception in phonology*, Academic Press, New York, 141–185.

Johnson, Keith (1997). Speech perception without speaker normalization: An exemplar model. Johnson & Mullennix (1997), 145–165.

Johnson, Keith (2011). Modeling phonology in time. *2011 Annual Report*, UC Berkeley Phonology Lab, 183–188, URL http://linguistics.berkeley.edu/phonlab/annual_report/annual_report_2011.html.

Johnson, Keith & John Mullennix (eds.) (1997). *Talker Variability in Speech Processing*. Academic Press.

Keating, Patricia A. (1988). Underspecification in phonetics. *Phonology* 5, 275–292.

Kirby, James P. & Morgan Sonderegger (2013). A model of population dynamics applied to phonetic change. Knauff, M., M. Pauen, N. Sebanz & I. Wachsmuth (eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, TX, 776–781.

Kuhl, Patricia K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50:2, 93–107.

Lacerda, Francisco (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. Elenius, K. & P. Branderyd (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences*, Stockholm, vol. 2, 140–147.

Liljencrants, Johan & Björn Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48:4, 839–862.

Meinhof, Carl (1932). *Introduction to the phonology of the Bantu languages: being the English version of "Grundriss einer Lautlehre der Bantusprachen"*. Dietrich Reimer (Ernst Vohsen); Williams & Norgate, Ltd., Berlin; London. Ed. and trans. by Nicolaas Jacobus van Warmelo.

Milgram, Stanley (1967). The small-world problem. *Psychology Today* 1:1, 61–67.

Möbius, Bernd (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica* 38:1-2.

Nettle, Daniel (1999). Using social impact theory to simulate language change. *Lingua* 108:2-3, 95–117.

Nosofsky, Robert M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115:1, 39–57.

Pater, Joe (1999). Austronesian nasal substitution and other NÇ effects. Kager, René, Harry van der Hulst & Wim Zonnenveld (eds.), *The Prosody-Morphology interface*, Cambridge University Press, Cambridge, 310–343.

Pierrehumbert, Janet B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. Bybee, Joan L. & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, John Benjamins Publishing, 137–157.

Rothenberg, Martin (1968). *The breath-stream dynamics of simple-released-plosive production.* No. 6 in Biblioteca Phonetica, Karger, Basel.

Solé, Maria-Josep, Larry M. Hyman & Kemmonye C. Monaka (2010). More on post-nasal devoicing: The case of shekgalagari. *Journal of Phonetics* 38:4, 604–615.

Watts, Duncan J. & Steven H. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature* 393:6684, 440–442.

Wedel, Andrew (2004). Category competition drives contrast maintenance within an exemplar-based production/perception loop. *Proceedings of the Seventh Meeting of the ACL Special Interest Group in Computational Phonology*, Association for Computational Linguistics, Barcelona, Spain, 1–10.

Wedel, Andrew (2006). Exemplar models, evolution and language change. *The Linguistic Review* 23, 247–274.

Wedel, Andrew (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition* 4:4, 319–355.

Wedel, Andrew & Heather Van Volkinburg (unpublished). Modeling simultaneous convergence and divergence of linguistic features between differently-identifying groups in contact, URL `http://dingo.sbs.arizona.edu/~wedel/publications/PDF/Wedel_VanVolkinburgSneetches.pdf`. Manuscript.

Westbury, John R. & Patricia A. Keating (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics* 22:01, p. 145.

Yip, Moira (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19:1, 65–100.