Unintuitive phonetic behavior in Tswana post-nasal stops

Jagoda Bruni[1], *Daniel Duran*[1], *Grzegorz Dogil*[1]

[1] Institute for Natural Language Processing, University of Stuttgart

Jagoda.bruni; Daniel.duran; Grzegorz.dogil@ims.uni-stuttgart.de

Abstract

This article describes the phonetic process of post-nasal devoicing in Tswana. We propose a multi-agent exemplar model with various interaction schemes which include factors like functional and social biases in order to account for this counter-intuitive phenomenon. Our novel hybrid multi-agent modeling framework facilitates investigation of sound change by combining the sociophonetic model of Nettle [22] and the exemplar-based model of Wedel [26] into a single unified model.

Index Terms: Tswana, devoicing, computer simulations, Exemplar Theory.

1. Introduction

According to studies conducted by [7] and [24], languages from the Sotho-Tswana group of Bantu languages demonstrate unintuitive voicing behavior in devoicing of postnasal voiced plosives (/mb/→[mp]) – unintuitive in that greater articulatory effort is required to terminate voicing than to maintain it [30]. 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 [20]. Current studies on Tswana and Shekgalagari [7, 25] however, demonstrate that nasal segments remained in those languages – surprisingly not only before voiced stops but also before voiceless ones.

[7] present experimental acoustic data which provide evidence of active post-nasal devoicing involving Tswana native speakers. The authors describe measurements of 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. It is pointed out [7] 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.

In our approach we consider various possibilities for the sound change currently occurring in Tswana. The main idea is that the interaction between different biases in speech production and perception along with an exemplar-theoretic organization of linguistic knowledge may trigger sound change and contribute to the development of unintuitive phonetic behavior. Our investigations are based on the data described above [7], as well as on documentation of historical changes in the Sotho-Tswana group proposed by [20] and [25]. It has been suggested by [3] that some languages might demonstrate genetic relatedness by building sound correspondences across them. The authors propose 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 [16] demonstrates the necessity to model language-internal processes through the observation of whole communities of speakers. [16] 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 framework suggested by [16] 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 phonetics and phonology incorporating sociophonetic aspects. It can serve as an explanatory framework for various sound changes. We simulate both social and functional biases and we propose an exemplar-based category formation [23] by implementing simulation schemes proposed by [22] and [26].

2. Unintuitive devoicing and Exemplar Theory

The phonetic process of post-nasal devoicing has been analyzed (among others) by [14]. The authors implemented the computational model of [30] based on previous work by [24] and tested the hypothesis that part or all of the stop closure after a nasal is realized with vocal fold vibration. The results by [14] demonstrate that a post-nasal position of a stop facilitates its voicing. It confirms the hypothesis of [30] 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 [7] demonstrated that the devoiced variant of post-nasal stop is not the only possible, still being the dominant one (more than 80% of /m+bV/ and /m+pV/ sequences was realized as [m+pV]). For most of the participants this process seemed not to be categorical. It is thus argued, that the devoicing tendency in Tswana post-nasal stops might result from historical language changes (described in more detail by [25]). During these changes a general stop-devoicing process applied at a history stage where stops were observed only in the post-nasal environment. [7] argue that possible Tswana sound change is phonetic in origin but once it is phonologized, it turns out to be independent from phonetics.

We assume an exemplar-theoretic organization of an individual's mental lexicon, i.e. categories are represented by collections of remembered speech items. Our assumption is that speech production and perception are tightly linked. Percepts of linguistic experiences are stored in the mental lexicon rich in detail, including phonetic and indexical information [11, 15, 23]. Our model implements a strict interpretation of Exemplar Theory: Speech production and perception are modeled at the level of individual exemplars (see *Fig.1*) – this is in keeping with Wedel's [26] approach but in contrast with Nettle's model which computes averages over the entire population [22]. Moreover, each individual agent has its own private lexicon containing previously perceived exemplars.

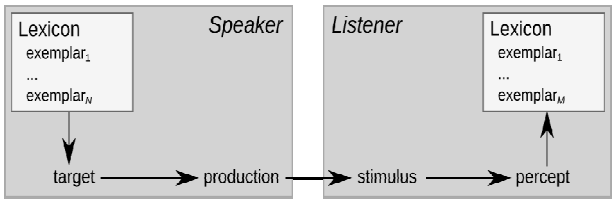

Figure 1: *Distinction between different kinds of exemplars according to their place in the production-perception chain from a speaker to a listener.*

3. Simulation approach

3.1 Related work

Exemplar Theory [11, 19, 23] 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. 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 phonetic item and its role in speech production/perception and language acquisition [10]. The more frequent exemplars form centers of categories and are easily accessible during language production, while the least frequent ones suffer from memory decay [5, 23]. It has also been demonstrated [23] 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.

Wedel and colleagues [26, 27, 28, 29] present computer simulations of category competition based on Exemplar Theory [11, 19, 23]. His multi-agent model assumes that each agent has its own lexicon consisting of a collection of previously encountered exemplars. These exemplars are rich in detail and represented by continuous phonetic cues. Production of exemplars is based on the items stored in the lexicon. Perception involves the comparison of the input form against the contents of the lexicon. Categories are implicitly represented by the distribution of exemplars within the feature space of an agent's lexicon. Various biases in exemplar selection influence the self-organizing evolution of this system. Moreover, in Wedel's [26] model agents do not age (i.e. they do not go through different life stages). The population is fixed throughout the simulations. No agents are removed from the population or introduced to it. The social structure in the model of [29] is limited to two distinct groups. Apart from its group identity, every agent is equal in terms of social standing or impact.

On the other hand, Nettle [22] uses computer simulations based on Social Impact Theory [6] where one of the important factors of the dynamics within the speech community is the social status associated with individuals. The learning process involves competition between two competing variants. Nettle [22] incorporates social distances, status and functional selections. When a "hyper-influential" individual happens to have a rare variant, it quickly spreads in the individual's immediate neighborhood. If it spreads fast enough, the rare variant attains critical mass and replaces the dominant form in a rapidly rising curve similar to that of real linguistic change. The population then stabilizes at near homogeneity with the previously rare variant as the norm and remains in that state for a long equilibrium period until another change is triggered.

3.2 Multi-agent exemplar model

Usage-based methodology is grounded in language accounts which, in our opinion, provide a link between simulated inventory of a language and its set of phonetic categories. Similarly, [2] in their simulations claim that the exemplar-theoretic approach serves better understanding in describing phonological constraints of a language.

We adapt and combine the methods proposed by Nettle [22] and Wedel and his colleagues [26, 27, 28, 29] by modeling competition between variants undergoing functional and social selection during language acquisition over many generations. We show that modeling voicing profiles [21, 4] 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 [7].

Our approach is currently 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 *small world* structure is represented by a 20 × 20 toroidal network and is initialized 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 as compared to a regular grid. This network topology is closer to real social networks.

The implementation of learning in our sociolinguistic model consists of five life stages (infancy and four stages of adolescence and adulthood), contrary to Nettle's [22] method, where learning is limited to just a few stages. From the general point of view, a limitation in the learning of phonological contrast does not have to be true, at least not for all learners, since some speakers/listeners are more talented than others [8]. [13] has also demonstrated 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.

Speech production in our model is based on a target selection procedure. The speaker selects one exemplar from her lexicon which serves as a production target. This selection is performed probabilistically based on a scoring of exemplars, as shown in equation 1 for an exemplar x_i. On the other hand perception involves a transformation of the actual stimulus exemplar into a percept. This is motivated by findings about experience-based language specific speech perception. The stimulus is warped slightly toward local maxima in the exemplar distribution of the listener's lexicon and it introduces a functional bias in perception which facilitates entrenchment. This process corresponds to the so-called perceptual magnet effect [19, 12, 18].

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 (Fig. 1). The score is a weighted sum with three components. For an exemplar x_i it is defined as follows:

$$score_i = \frac{\alpha\, sim(x_i, x_z) + \beta\, status(x_i) + \gamma\, closeness(x_i)}{\alpha + \beta + \gamma} \quad (1)$$

where $$sim(x_i, x_z) = e^{-d_{iz}} \quad (2)$$

is the phonetic similarity of the i-th exemplar to the centroid x_z of the lexicon [15, 22, 27], and d_z is the Euclidean distance between x_i and x_z. The similarity is positive with a value of 1.0 when two items are maximally similar, i.e. the same.

$$status(x_i) \quad (3)$$

is the social status attached to the i-th exemplar. This is the (perceived) social status of the original speaker of that exemplar. The values are limited to the range [0.0 1.0] where 1.0 is the highest possible status only assigned to hyper-influential agents.

$$closeness(x_i) = 1 - \frac{d_{s,l}}{d_{max}} \quad (4)$$

is the social *closeness* of the original speaker of exemplar x_i to the listener, where $d_{s,l}$ is the social distance between the speaker who produced exemplar x_i and the listener who stored 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 positive and the normalizing factor d_{max} is the maximum distance within the network. In our case of a 20x20 toroidal regular grid (Fig.2), d_{max}= 20. The closeness is thus limited to positive values where 1.0 is the maximum (corresponding to the individual itself). The overall score of an exemplar x_i is thus limited to a value between 0 and 1. The weights α, β and γ are model parameters which need to be set (or, which need to be learned). Here we assume that these weights are the same for all agents. The actual production target is selected probabilistically according to the scores assigned to the lexical items. Once selected, a specified amount of Gaussian noise is added to the phonetic values of an exemplar to approximate articulatory production noise. This noisy copy is then transmitted from the speaker to the listener.

The agents of the population act both as *speakers* and *listeners*. The scheme, according to which agents interact, constitutes a further 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:

full interaction: In each epoch, each agent interacts with everybody else in the population;

social status: In each epoch, everybody interacts with other agents depending on their social status with uniformly distributed random listeners. The total number of listeners for a given speaker depends on the respective social status;

closeness: In each epoch, everybody interacts with other agents depending on their social closeness. A number of listeners is probabilistically selected according to 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.

The influence of different interaction schemes on the evolution of the system may be studied by defining appropriate rules of interaction.

The simulation framework is implemented in Java. All model parameters are easily adaptable by the experimenter. Due to a modular program design, it is possible to adapt certain aspects of the simulation according to whatever hypotheses are of interest in a given study. The network topologies or the interaction schemes presented here could easily be adapted to describe a language contact situation with two distinct (sub) populations, for example. Various statistics are produced for individual agents as well as for the population as a whole – e.g. the exemplar distributions of an agent's lexicon or the record of all produced utterances in each epoch.

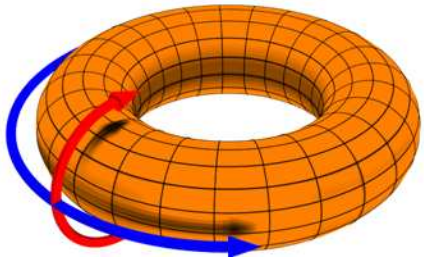

Figure 2: *Illustration* of *closed toroidal topology of agent network: the underlying regular grid network is connected such that there are no edges. [32]*

3.3 Results

The capabilities of our hybrid model are illustrated on the outcomes of preliminary experimental trials involving comparison of the two network topologies and various interaction schemes.

We determine the ratio of productions of plain variants (i.e. phonetically intuitive, voiced 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., the phonetically intuitive variant.

In Figure 3, 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 4 shows a comparison of different scoring weights. Four different parameter settings are shown: First, having equal weights on the three components of the scoring function as in Figure 3. Additionally, the parameters α, β and γ are set in turn to a high value while keeping the remaining two parameters at an equally low value.

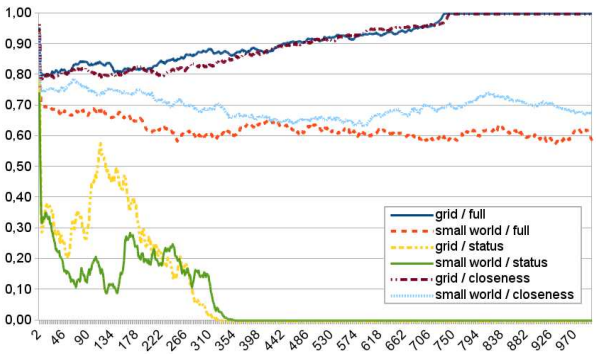

Figure 3: *Proportion of plain variant productions for different network topologies and interaction schemes with equal scoring weights and a limited lexicon capacity over 1,000 epochs.*

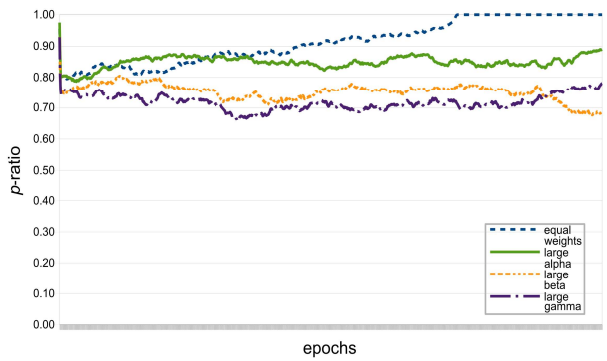

Figure 4: *Proportion of plain variant productions for different scoring weights with a regular grid topology, full interaction and a limited lexicon capacity over 1,000 epochs.*

These 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 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.

4. Conclusions

Our study describes a usage-based approach with a multi-agent exemplar model of phonetically unexpected post-nasal stop devoicing in Tswana. It has been claimed [30] 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. It is also claimed that such phonetically unintuitive distributions in the world's languages are hard to explain by means of natural phonological rules [17, 31]. In case of Exemplar Theory, they are also complex to describe by means of functional phonological biases. In 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. Thus our implementation has a sociolinguistic character where we use Social Impact Theory [6, 22] which clearly models various social statuses of individuals.

5. Acknowledgements

This work is funded by the German Research Foundation (DFG) within the Collaborative Research Center SFB 732 /A2.

6. References

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

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

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

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

[5] 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.

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

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

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

[9] 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.

[10] 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.

[11] Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22, 1166-1183.

[12] 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.

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

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

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

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

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

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

[23] 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.

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

[25] 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.

[26] 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.

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

[28] 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.

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

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

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

[32] Image source: DaveBurke (Own work) [GFDL (http://www.gnu.org/copyleft/fdl.html), CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/) or CC BY 2.5 (http://creativecommons.org/licenses/by/2.5)], via Wikimedia Commons.