A HYBRID MODEL TO INVESTIGATE LANGUAGE CHANGE

Daniel Duran, Jagoda Bruni, Michael Walsh and Grzegorz Dogil

Institute for Natural Language Processing, University of Stuttgart
<firstname.lastname>@ims.uni-stuttgart.de

ABSTRACT

In this paper we propose a hybrid multi-agent modelling framework which facilitates investigation into sound change by combining the sociophonetic model of Nettle [14] and the exemplar-based model of Wedel [21] into a single unified model. The framework facilitates simulation scenarios of different social networks with varying interaction schemes and social distances between the agents. Additionally, the structure of an individual agent's mental lexicon is embedded in an exemplar-theoretic setting. The goal of this modelling framework is to enable examination of competition between different phonetic forms. Though in no way limited to this particular example, we illustrate our new hybrid framework with the case of competition between phonetically intuitive /mb/ and unintuitive [mp] voicing variants of post-nasal stops in Tswana.

Keywords: phonetic simulations, language change, Tswana, Exemplar Theory.

1. INTRODUCTION

Dynamics of language use and historical change are empirically difficult to study on a population level. We suggest to model *experience*, which is based on fieldwork results providing machine readable speech corpora, by computational simulations. Simulations have proven useful in testing the effects of various model parameters in the evolution of entire language systems [1, 11]. Applying computer simulations provides a means of investigating models and testing hypotheses, as well as directly addressing empirically inaccessible phenomena, such as sound changes over many generations. Today, simulations are a well-established method in the study of sound change [1, 2].

A diverse range of computational approaches has been proposed over the past 5–6 decades. Early studies employed, for example, phonological "rule-testing" systems (e.g. aiming to generate contemporary Russian forms from a set of assumed Proto-Indo-European forms, as presented by [18]). This branch of research is, for example, reflected in contemporary computer simulation experiments within the framework of (Stochastic) Optimality Theory (e.g. modelling the emergence of dispersion effects in sound systems after a number of generations [4]). Recently, agent-based simulations have often been used to investigate language change under varying conditions of social structure and acquisition biases or other factors. These models often consist of pairs of speaker–hearer agents which repeatedly exchange some abstract speech items [3, 23, 21].

Nettle [14] uses computer simulations based on Social Impact Theory where one important factor of the dynamics of the speech community is the *social status* associated with individuals. The learning process involves competition between two variants. The model incorporates social distances, status and functional selections. When a "hyper-influential" individual happens to have a rare variant, the variant quickly spreads in the individual's immediate neighbourhood. 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 stabilises 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.

Wedel and colleagues [21, 2, 23] present computer simulations of category competition based on Exemplar Theory. His multi-agent model assumes that each agent has its own lexicon consisting of a collection of previously encountered exemplars. These exemplars are detail rich, represented by continuous phonetic cues. Production of an exemplar 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. Different biases on exemplar selection influence the self-organising evolution of the system.

We adapt and combine the methods proposed by Nettle and Wedel, yielding a modelling framework capable of investigating competition between variants undergoing functional and social selection during language acquisition over many generations, bringing together the sociophonetic and exemplar-theoretic viewpoints.

2. THE FRAMEWORK

We assume an exemplar-theoretic organisation 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 [6, 9, 16]. Our model implements a strict interpretation of Exemplar Theory: Speech production and perception are modelled at the level of individual exemplars – this is in keeping with Wedel's [21] approach but in contrast with Nettle's model which computes averages over the entire population [14]. Moreover, each individual agent has its own lexicon containing previously perceived exemplars.

The simulation framework is implemented in Java. All model parameters are easily adaptable by the experimenter. Due to a modular programme 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.

2.1. Social structure and interactions

The topology of the social network is one modelling decision that affects the dynamics of the system. We investigate the effects of two different network topologies (represented as formal graphs): a *regular grid* and a *small world* network. Within the former, agents are arranged in a social network on a regular grid with 20 columns and 20 rows [14], where each node in the network represents one agent and each edge represents a social connection between two agents. The minimal path length within the graph represents the social distance between two given agents. The network is connected, i.e. there is a path between any two agents in the network, and it takes on a closed toroidal topology (doughnut shaped). Each agent has exactly four direct neighbours who are other agents with a social distance of 1 (see below). In addition to this *regular grid* topology, we investigate a *small world* network created according to the random rewiring procedure proposed in [20]. The structural effect of this network is that there is a smaller average distance between any pair of agents, while the average number of direct neighbours for each agent is still the same. It has been argued that this network topology is closer to real social networks. Note that the total number of agents as well as their topological arrangements are not inherent to our model and can be adjusted as required.

Exemplars are represented by sets containing various types of information. Here, we assume exemplars consist of continuous phonetic information (implemented as N-dimensional real vectors) and social information (real numbers indicating social status and social closeness, see below).

Initially each agent gets an age between 0 and 4 (inclusive) and a randomly assigned social status according to a specified distribution. Agents of age 0 start with an empty lexicon. All other agents are initially seeded with a number of exemplars of the "plain" variant. A small number of agents are 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 exemplars of the "unintuitive" variant. The two variants are distinguished by different phonetic prototypes, the values of which must be specified in advance. Our current assumption is that the competing variants belong to the same phonemic category. Thus, no category label is attached to exemplars, and there is no categorisation involved in perception, i.e. accurate categorisation of percepts is assumed in the simulations.

The simulation runs for a specified number of epochs. In each epoch, each agent acts as a listener. The speakers for each listener are selected from the population according to some sampling procedure which we refer to as *interaction scheme*. The precise definition of such an interaction scheme is another modelling decision. We assume that only agents above age zero are potential speakers. At the end of each 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 (starting with an empty lexicon).

2.2. Production target selection and auditory biases

Selection of a production target is based on a scoring procedure. The score of an exemplar x_i is a weighted sum with three components defined as follows:

(1)

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

where:

- $\text{sim}(x_i, x_z) = \text{e}^{-d_{iz}}$ is the phonetic similarity of the i-th exemplar to the centroid x_z of the lexicon [15, 22, 9], and d_{iz} is the Euclidean distance between x_i and x_z. The similarity is positive with a value of 1.0 when the two items are maximally similar, i.e. the same.
- $\text{status}(x_i)$ is the social status attached to exemplar x_i. This is the (perceived) social status of the original speaker of that exemplar. The values are positive, with 1.0 being the highest possible status (only assigned to hyper-influential agents).
- $\text{closeness}(x_i) = 1 - \frac{d_{s,l}}{d_{\max}}$ 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. The normalising factor $d_{\max}$ is the maximum distance within the network. In our case of a 20×20 toroidal regular grid, $d_{\max} = 20$. 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.

Additionally, we incorporate an auditory bias which draws percepts towards local maxima of the exemplar distribution within the lexicon. This auditory warping corresponds to the *perceptual magnet effect* [12]. The effect is that of entrenchment which counters the effect of production noise. In general, it will also draw outliers closer to the dense regions of the lexicon in the phonetic space. The influence of such auditory biases can be studied by deactivating this processing module and analysing the resulting behaviour of the system. In our discussion below we employ auditory warping for all simulations.

3. MODEL APPLICATION

By way of illustration, we discuss the framework presented in this paper in the context of investigating voicing behaviour in post-nasal stops in Tswana. According to [5, 17], languages from the Sotho-Tswana group of Bantu languages demonstrate *unintuitive* voicing behaviour in devoicing of post-nasal voiced plosives /mb/ → [mp] – unintuitive in that greater articulatory effort is required to terminate voicing than to maintain it [24]. 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 [13]. Current studies on Tswana and Shekgalagari [5, 8, 19], however, demonstrate that nasal segments remained in those languages – surprisingly not only before voiced stops but also before voiceless ones. The phenomenon of post-nasal behaviour has been investigated from different perspectives. Hayes & Stivers [7] implemented the computational model of Westbury & Keating [24] based on the previous work of Rothenberg [17] and tested the hypothesis that part or all of the stop closure after a nasal, is realised with vocal fold vibration. Their results demonstrate that post-nasal position of a stop facilitates its voicing. It confirms the hypothesis [24] that voicelessness requires additional articulatory cost, whereas voicing reflects a neutral state in post-nasal position. Furthermore, Coetzee & Pretorius [5] pointed 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. The authors describe acoustic 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. We apply our hybrid model in order to investigate factors influencing post-nasal devoicing and its evolution over time.

4. PILOT STUDY AND DISCUSSION

We illustrate the capabilities of the hybrid model on the outcomes of experimental runs comparing the two network topologies and different interaction schemes. The phonetic features for the two competing variants are modelling the Tswana case. The initial phonetic cue values are taken from [5]. Three interaction schemes are employed: (1) "full" interaction, where the listener gets input from all other members of the population (above age zero), (2) "status" interaction, where speakers are selected probabilistically for each listener based on their social status (with higher status having a higher probability of speaking), and (3) “closeness” interaction where speakers are selected probabilistically based on their social distance to the listener.

The following figures show the development of the linguistic exchange over 1,000 epochs. In each epoch, the total numbers of produced exemplars of both variants are recorded. The proportion of all produced exemplars of the “plain” (intuitive) variant in one epoch is referred to as the *p-ratio*. A p-ratio of 1.0 means that all of the productions are of the plain variant. As the population is initialised with the plain variant being the dominant one, the simulations all start with a high p-ratio. Figure 1 shows three examples for configurations with equal target scoring weights α, β and γ and a regular grid network topology. The three curves show the effect of the three different interaction schemes. Figure 2 shows three examples for the same configurations except for the network topology which in this case has *small world* properties. It can be observed that two of the interaction schemes lead to a stable state where both variants are used.

Figure 1: Different interaction schemes on regular grid network with equal weights $\alpha = \beta = \gamma$

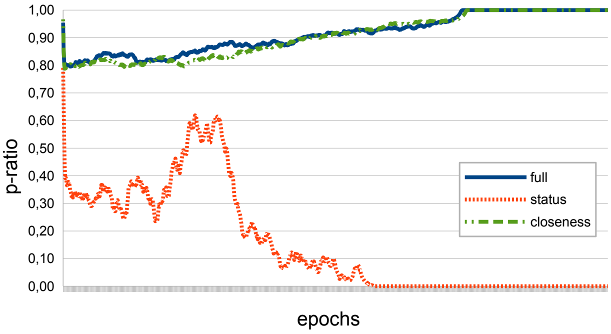
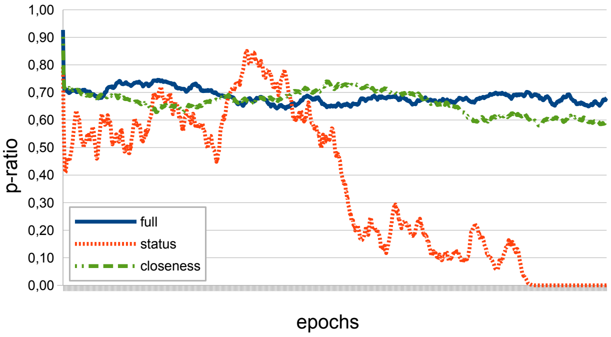


Figure 2: Different interaction schemes on small world network with equal weights $\alpha = \beta = \gamma$



Comparing figures 1 and 2 shows that in these cases the regular grid in figure 1 does not seem to favour a state in which both competing variants can be observed at the population level as in the Tswana case. Both figures illustrate that learning from all other members from the population does not appear to be qualitatively different from learning from socially close individuals (the ‘full’ and ‘closeness’ schemes). Learning from individuals according to their status, on the other hand, leads to a very different evolution of the system (the ‘status’ scheme).

According to Westbury & Keating [24], 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 in a particular phonetic environment. In the case of Tswana post-nasal clusters, the unintuitive *de*voicing behaviour might be grounded in sociolinguistic reason like a prominence bias for social status of individuals using rare devoiced variants. On the other hand, it might be an interplay connected with the functional biases which favour devoiced stops in order to make the morphological boundary more perceptually salient. The results presented here are based on a specific parameter configuration, with the assumption of equal target scoring weights being undoubtedly an unrealistic one. Future work will require training of the various parameters fitting the systems behaviour to the empirical data – in the Tswana case, for example, the goal might be to achieve a p-ratio of 20% according to the observed proportion of the variants [5]. In further work, this optimisation problem will be addressed by applying stochastic gradient descent to learn the weights.

5. CONCLUSIONS

Our novel hybrid framework combines social modelling principles proposed by Nettle [14] and self-organizing exemplar-theoretic dynamics proposed by Wedel [21] in a fully functional framework facilitating agent-based simulations of language change. This is the first model that combines these approaches into a single investigative framework. Our preliminary study, though still based on simplified assumptions, already highlights the potential benefits of the hybrid model presented here.

The implementation of the framework will be made available to the research community in the future in order to allow reproduction of any reported results as well as to provide a tool for further studies.

6. ACKNOWLEDGEMENTS

This work is funded by the German Research Foundation (DFG), grant SFB 732 / A2.

7. REFERENCES

[1] Baker, A. 2008. Computational approaches to the study of language change. *Language and Linguistics Compass* 2(2), 289–307.

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

[3] de Boer, B. 1998. A realistic model of emergent phonology. AI-memo 98-04 Artificial Intelligence Laboratory at the Vrije Universiteit Brussel.

[4] Boersma, P., Hamann, S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25, 217–270.

[5] Coetzee, A. W., Pretorius, R. July 2010. Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38(3), 404–421.

[6] Goldinger, S. D. 1997. Words and voices – perception and production in an episodic lexicon. In Johnson and Mullennix [10] chapter 3, 33–66.

[7] Hayes, B., Stivers, T. June 2000. Postnasal voicing. Manuscript.

[8] Hyman, L. M. 2001. The limits of phonetic determinism in phonology: *NC revisited. In: Hume, E., Johnson, K., (eds), *The role of speech perception in phonology*. New York: Academic Press 141–185.

[9] Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In Johnson and Mullennix [10] 145–165.

[10] Johnson, K., Mullennix, J., (eds) 1997. *Talker Variability in Speech Processing*. Academic Press.

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

[12] Kuhl, P. 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.

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

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

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

[16] Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J. L., Hopper, P., (eds), *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing 137–157.

[17] Rothenberg, M. 1968. *The breath-stream dynamics of simple-released-plosive production*. Number 6 in Biblioteca Phonetica. Basel: Karger.

[18] Smith, R. N. 1969. Automatic simulation of historical change. *International Conference on Computational Linguistics COLING 1969: Reprint No. 9*. Association for Computational Linguistics (ACL).

[19] Solé, M.-J., Hyman, L. M., Monaka, K. C. Oct. 2010. More on post-nasal devoicing: The case of shekgalagari. *Journal of Phonetics* 38(4), 604–615.

[20] Watts, D. J., Strogatz, S. H. June 1998. Collective dynamics of ‘small-world’ networks. *Nature* 393(6684), 440–442.

[21] Wedel, A. July 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* Barcelona, Spain. Association for Computational Linguistics 1–10.

[22] Wedel, A. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23, 247–274.

[23] Wedel, A., Van Volkinburg, H. unpublished. Modeling simultaneous convergence and divergence of linguistic features between differently-identifying groups in contact. Manuscript.

[24] Westbury, J. R., Keating, P. A. Mar. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22(01), 145.