Hierarchical inference, frequency, context, and frequency in context

Vsevolod Kapatsinski[1]

[1]*Department of Linguistics, University of Oregon (USA)*

Sound change does not proceed at an equal rate in all contexts. Rather, any innovative pronunciation is favored by some lexical, phonological, and social context compared to others. In sociolinguistics, social and grammatical influences on sound change have generally been modeled using logistic regression [3]. Extending this framework, [7] and [13] argued that lexical influences on sound change should be treated as random effects in the same framework. This amounts to a claim that the phonological grammar is a regression model in which effects of larger linguistic units like words, collocations, and morphemes on phonology are inferred through hierarchical inference.

Hierarchical inference performs adaptive partial pooling of evidence [5]. When the learner has many observations of a particular unit, the phonological behavior of that unit is best inferred on the basis of its own tokens. In contrast, when the learner has few observations of a unit, s/he must rely to some extent on the behavior of other similar units. Hierarchical inference optimally trades off these sources of information such that credit for a pronunciation is optimally allocated between smaller and larger units.

[7] showed that adaptive partial pooling naturally predicts that words need to be frequent enough to withstand peer pressure, resulting in analogical leveling of low-frequency exceptions even in the absence of memory limitations (e.g., [5], [8]). Iterated learning via hierarchical inference also accounts for the emergence and maintenance of stable variation and makes predictions for the circumstances under which it is likely to emerge. Specifically, stable lexically-specific variation frequently emerges when words undergoing a change vary widely in frequency because the learner misattributes the frequency effect to lexical idiosyncrasy.

However, an important limitation of this model is that it does not account for the frequency-in-favorable-contexts effect (FFC; [1], [2], [4], [10], [11]). FFC refers to the finding that words that frequently occur in change-favoring contexts are more likely to undergo the change in other contexts as well, compared to words that rarely occur in change-favoring contexts. Although hierarchical inference is compatible with this finding, [7] did not explore what happens when some phonological contexts favor change and others disfavor it. Because frequency in any context favored reduction, all exceptions in the model were exceptionally conservative – no words were exceptionally innovative.

The present paper extends the model of sound change developed in [7] by incorporating favorable and disfavorable effects of phonological context. It is shown that the model predicts that high-frequency words will drift apart, with some becoming exceptionally conservative and others exceptionally innovative depending on the contexts in which they tend to occur (replicating the FFC effect and resulting in polarized variation). In contrast, rare words remain similar, following the central tendency of the lexicon.

The model further shows that, whereas the effect of frequency in change-favoring contexts [1] always arises, an overall effect of token frequency may or may not arise, depending on how strongly change-disfavoring contexts favor conservative pronunciation variants. This has the potential of resolving conflicting findings in the literature (e.g., [4], vs. [11]).

An interesting difference from exemplar models of change (e.g., [9]) is that an increase in the word frequency effect is not predicted to automatically occur when a change accelerates (Fig. 1; and indeed it is often not observed, e.g., [12]). That is, a change can be driven by word frequency and yet the token frequency effect will stay constant. The increases in frequency effects that sometimes accompanying the acceleration of change are argued to be due to an increase in social marking of the competing variants instead of lexical storage.

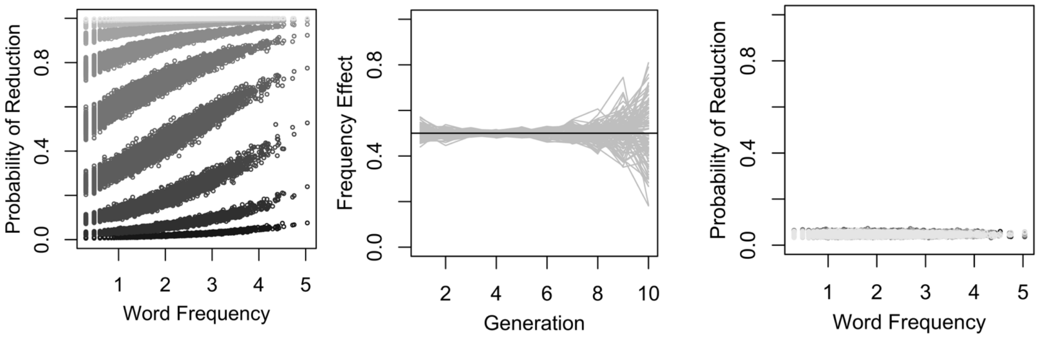

Fig. 1. An illustration of how the frequency effect can remain constant in the model even though change is driven by frequency. Left: The relationship between word frequency and probability of producing the reduced variant of the changing sound for each word across 10 generations of speakers (earliest = darkest). The change progresses through the lexicon. Middle : the effect of frequency on reduction (in log odds) stays constant across the generations. The first generation is seeded with frequent words being more reduced than rare words due to online reduction. Subsequent generations learn the overall probability of reduction and word-specific deviations from this overall probability, and also reduce on top of this target probability. Right : what happens if there is no online effect of frequency on reduction with the same overall probability of reduction in Generation 1 as on the left: the change does not progress, showing that it is driven by frequency.

References (Use a 9 or 10 point font size with single line spacing.)

[1] Brown, E. L. (2004). The reduction of syllable-initial /s/ in the Spanish of New Mexico and Southern Colorado: A usage-based approach. PhD diss., University of New Mexico.

[2] Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*, 261–290.

[3] Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, *50*, 333–355.

[4] Forrest, J. (2017). The dynamic interaction between lexical and contextual frequency: A case Study of (ING). *Language Variation and Change*, *29*, 129–156.

[5] Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

[6] Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83-91.

[7] Kapatsinski, V. (2021). Hierarchical inference in sound change: Words, sounds, and frequency of use. *Frontiers in Psychology*, *12*, 652664.

[8] Lieberman, E., Michel, J. B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, *449*, 713-716.

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

[10] Raymond, W. D., & Brown, E. L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In S. Th. Gries & D. Divjak (Eds.) *Frequency effects in language learning and processing* (pp.35-52). Berlin: Mouton de Gruyter.

[11] Raymond, W. D., Brown, E. L., & Healy, A. F. (2016). Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. *Language Variation and Change*, *28*, 175-202.

[12] Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, *47*, 18-35.

[13] Zymet, J. (2018). Learning a frequency-matching grammar together with lexical idiosyncrasy: Maxent versus hierarchical regression. In *Proceedings of the annual meetings on phonology*.