Contributed to *DougSchrift: A Collection of Squibs and Puzzles presented to Doug Pulleyblank*, 2025.
https://arts-pulleyblank-2024.sites.olt.ubc.ca/

Modeling Japanese voiced velar nasalization in Emergent Phonology: An embodied approach to frequency effects

Connor Mayer (University of California, Irvine) and Bryan Gick (UBC)

1. Introduction

The goal of this squib is to demonstrate how Emergent Phonology (henceforth EP; Archangeli and Pulleyblank 2015, 2017, 2022) can be augmented to model phonological frequency effects. We will demonstrate the efficacy of this method with an analysis of experimental data looking at the phonology of compound formation in Japanese. We will close by outlining some connections between this perspective and the Embodied Speech (ES) approach described in Gick and Mayer (in prep), and how EP can serve as a bridge between embodied approaches to speech and theoretical phonology.

2. Background on Emergent Phonology

While traditional generative phonology treats phonological knowledge as a distinct cognitive subsystem, EP starts from the assumption that it is not: rather, the kinds of phenomena we see in phonological systems can be explained on the basis of domain-general cognitive mechanisms such as memory, attention, sequence processing, sensitivity to frequency, the ability to generalize, and so on. EP is an attempt to formalize some of the consequences and predictions that arise from this perspective. It's beyond the scope of this humble squib to describe EP in all its glory and to justify its formal properties, but we will quickly summarize the parts necessary to understand the analysis to come.

In EP, the fundamental unit of organization is the morpheme. The role of the phonological grammar is to select from a set of surface morphs corresponding to that morpheme, called a *morph set*. In this respect, EP is a surface-oriented model of phonology: it does not posit abstract underlying forms, but rather maps directly from morphosyntactic representations to surface morphs. The choice between different morphs is governed by Well-Formedness Conditions (WFCs), which are analogous to constraints in Optimality Theory (OT; Prince & Smolensky 1993/2004). WFCs can encode both paradigmatic (e.g. no round front vowels) and syntagmatic constraints (e.g. no sequences of a round vowel follow by a non-round vowel), and can make reference to both phonological and morphosyntactic properties. Archangeli and Pulleyblank suggest that WFCs are ultimately derived from the frequencies with which a language permits certain structures to occur (for a similar perspective, see Hayes & Wilson 2008). To the extent that both are evident in the phonological patterns of a language, WFCs could be the result of general articulatory or perceptual difficulty (in which case similar WFCs might emerge across languages), or the result of language-specific, idiosyncratic patterns.

Let's look at a quick example using the English regular plural, which has the morph set {z, s, ɨz}$_{PL}$. The distribution of these morphs is as follows: [-ɨz] occurs following a strident, as in [kɪs-ɨz] "kisses", [-s] occurs following a voiceless non-strident, as in [kæts] "cats", and [-z] occurs elsewhere, as in [dɑgz] "dogs". We'll use the following WFCs to account for this distribution:

- **AGREEVOICE**: obstruent clusters must agree in voicing.
- ***[strident][strident]**: Don't have adjacent strident sounds.
- ***[nonstrident]{ɨz}$_{PL}$**: Don't use the {ɨz} morph of the plural after a non-strident sound.

The first two WFCs are fairly typical for an analysis of this pattern, and represent general restrictions on surface structures in English (corresponding to markedness constraints in OT). These two WFCs get us most of the way to an analysis of this pattern, but do not impose a preference between [kæts] and *[kætɨz] nor [dɑgz] and *[dɑgɨz], since none of these forms violate either WFC. In OT, we rely on the notion of an underlying form to rule out forms like *[kætɨz] and *[dɑgɨz]: the underlying form of the plural morpheme is analyzed as /z/, and the realization of /z/ as [ɨz] violates the DEP constraint, which penalizes inserting material (in this case a vowel) that wasn't present in the underlying form. Because inserting the vowel doesn't eliminate any other constraint violations, we take the path of least resistance and use either the [s] or [z] morphs. However, because EP does not have underlying forms, we cannot make use of faithfulness constraints to rule out these candidates, and so we need a different strategy.

The fact that *[kætɨz] and *[dɑgɨz] are ill-formed does not seem related to their phonological properties (e.g., there are similarly formed words like "lattice" or "togas"). Instead, we might treat the distribution of [ɨz] as derived from a *morphophonological* restriction: the [ɨz] form of the plural morpheme is only used following roots that end in a strident fricative. This restriction derives from the distribution of this plural morph, and is without exception in English. We can encode this as *[nonstrident]{ɨz}$_{PL}$, which reads "don't use the [ɨz] morph of the plural suffix following a non-strident sound."

We demonstrate that these three WFCs produce the correct results using an "assessment table", which is the EP equivalent of OT's tableaux. EP uses the same strict ranking evaluation process as OT to select the winning candidate (though in this particular case the ranking of the WFCs is unimportant). Below we show the assessments for the three forms of the plural suffix.

Assessment for [dɑgz]$_{\text{DOG-PL}}$
Morph sets: {dɑg}$_{\text{DOG}}$; {z, s, ɨz}$_{\text{PL}}$

{dɑg}$_{\text{DOG}}$ - {z, s, ɨz}$_{\text{PL}}$	***[strident][strident]**	**AgreeVoice**	***[nonstrident]{ɨz}$_{\text{PL}}$**
a. ☞ [dɑgz]			
b. [dɑgs]		*!	
c. [dɑgɨz]			*!

Assessment for [kæts]$_{\text{CAT-PL}}$
Morph sets: {kæt}$_{\text{CAT}}$; {z, s, ɨz}$_{\text{PL}}$

{kæt}$_{\text{CAT}}$ - {z, s, ɨz}$_{\text{PL}}$	***[strident][strident]**	**AgreeVoice**	***[nonstrident]{ɨz}$_{\text{PL}}$**
a. [kætz]		*!	
b. ☞ [kæts]			
c. [kætɨz]			*!

Assessment for [kɪsɨz]$_{\text{KISS-PL}}$
Morph sets: {kɪs}$_{\text{KISS}}$; {z, s, ɨz}$_{\text{PL}}$

{kɪs}$_{\text{KISS}}$ - {z, s, ɨz}$_{\text{PL}}$	***[strident][strident]**	**AgreeVoice**	***[nonstrident]{ɨz}$_{\text{PL}}$**
a. [kɪsz]	*!	*	
b. [kɪss]	*!		
c. ☞ [kɪsɨz]			

3. Frequency effects in Japanese compound formation

Frequency effects refer to cases where usage frequency influences phonological behavior (see Coetzee and Pater 2016 for an overview of some of these). Archangeli and Pulleyblank (2022, p. 42-43) touch on frequency effects only briefly, making use of a WFC to enforce a limited frequency effect:

***{morph$_\beta$}**: Assign a violation to each morph$_\beta$ which is not the most frequently occurring morph in its morph set.

This WFC imposes a preference for only the most frequent form, but they note that a more realistic definition would be "given a choice between two morphs from the same morph set, choose the one that is most frequent" (p. 43). In other words, in the absence of other WFCs that guide morph selection, we should prefer to use morphs that are more frequent. In the remainder of the squib, we will present a preliminary attempt to integrate frequency effects into an EP analysis. The analysis will show that similar morphs associated with different morphemes might behave in dramatically different ways depending on frequency of use.

The specific phenomenon we will look at is Voiced Velar Nasalization (VVN) in certain phonologically conservative dialects of Japanese. This feature is often associated with the (now largely extinct) Yamanote dialect in central Tokyo, but the data we will look at here, from Breiss et al. (in press), comes from the Tōhoku dialect, spoken in northern parts of *Honshū*, the main Japanese island. See references in Breiss et al. (2021a, 2021b, in press) for more detail.

The broad pattern in VVN is that the surface sounds [g] and [ŋ] occur in complementary distribution. Word-initially, we get [g], as in [gama] 'toad', while word-medially, we get [ŋ], as in [kaŋami] 'mirror' (cf. [kagami] in dialects of Japanese without VVN). In these monomorphemic words, the sounds of interest are always realized as [g] and [ŋ] in these positions respectively. It is also possible to observe *alternations* between [g] and [ŋ] within the same morpheme by looking at compounds. For example, consider a pair of words like [doku] 'poison' and [ga] 'moth', which can both serve as independent words. Combining them into the compound [doku-ga] 'poison moth' generates a word-internal [g], which is in principle subject to VVN. What is interesting about such compounds is that they display *variability* in whether VVN applies: 'poison moth' can be realized as either [doku-ga] or [doku-ŋa].

This variability is conditioned by a number of different factors, including the frequencies of the individual morphemes in isolation as well as the frequency of the compounds as a whole (Breiss et al. 2021a, 2021b, in press). We will focus on one specific frequency effect here: the more frequently that N2 (the second morpheme in the compound) is used in isolation, the lower the rate of VVN. An example of this is the morpheme [ga] 'fang', which is segmentally identical to the morpheme [ga] 'moth', but cannot be used as an independent word *[ga]. Accordingly, though 'fang' is realized with a [g] when it is the first element in a compound, as in [ga-ʒoː] 'main castle (literally 'fang castle'), it is always realized with [ŋ] when it occurs as the second element in a compound, as in [doku-ŋa] 'poison fang' (*[doku-ga]). This is an extreme example because the N2 in this compound has a standalone frequency of 0, and accordingly the compound has a VVN rate of 1. This effect also manifests more gradiently: as N2 standalone frequency goes up, the overall rate of VVN in its compound forms decreases. This has been shown to hold in both corpus data on existing forms (Breiss et al. 2021a, 2021b) as well as experimental data on novel compounds (Breiss et al. in press).

4. Modeling VVN

The question we'll attempt to address here is why this relationship between frequency and VVN rates should exist: specifically, what exactly links the two? We will show that an EP model augmented with frequency information predicts this relationship straightforwardly. In the discussion, we will suggest that the properties of this model align with proposals in the movement literature on how frequency of execution influences movement selection.

Our model will be set up as follows: we will assume that each N2 morpheme has two morphs in its associated morph set: one with an initial [g] and one with an initial [ŋ]. The outcome our model will try to predict is which morph gets selected in each case. We will use only two WFCs here:

***[son]g[son]**: Don't have a [g] between two sonorants (Ito and Mester 2003).

USEFREQUENT: Penalize the selection of low-frequency morphs.

The first WFC penalizes [g] in the contexts where we typically see [ŋ]. The second WFC imposes a pressure to use more frequent morphs (this is essentially the *{morph$_{\beta}$} WFC discussed above, but we have generalized it to be sensitive to relative differences in frequency and applied a more transparent name). We will assume that the number of violations of USEFREQUENT is proportional to the frequency with which each morph is produced. Specifically, we define the violations of USEFREQUENT to be the sum of the *cost* of each selected morph. We define cost as

$$\textit{cost}(\mathrm{x}) = 1 \,/\, \textit{asinh}(\textit{count}(\mathrm{x}))$$

where *asinh* is the inverse hyperbolic sine function and *count*(x) is the number of times the morph x occurs. We use *asinh* as opposed to the similar *log* function to avoid numerical issues when a morph has a count of 1, since *asinh* produces a small positive value for these values, while *log* produces 0. This function states that the cost of using a morph decreases as its frequency increases, with differences at lower frequencies producing larger changes than differences at higher frequencies. We will make the simplifying assumption that all ŋ-initial morphs have a frequency of 1, so that only the frequency of g-initial morphs affects model behavior.

The data we will apply this model to come from Experiment 1 in Breiss et al. (in press). In this study, eight speakers of Tohoku Japanese were asked to read 261 compound forms with a g-initial N2, 81 of which were existing forms and 180 of which were novel compounds. We will ignore the attested forms for the moment, which display rather more complex frequency dependencies, and focus only on the novel forms. Participants' responses were coded based on whether they produced the [g] or [ŋ] form of the N2 in each compound. The frequency for each N2 in isolation was estimated from the Balanced Corpus of Contemporary Written Japanese (Maekawa et al. 2014).

Because we are dealing with data that displays gradience (variability in the rate of VVN across different compounds), we will employ a Maximum Entropy model (henceforth MaxEnt; also known as a log-linear model). It's beyond the scope of this squib to provide a full definition of MaxEnt models, but they are commonly applied in phonological research as *Maximum Entropy Optimality Theory* models, which allow OT models to generate probability distributions over candidates rather than simply choosing the best candidate (see, e.g., Goldwater and Johnson 2003, Hayes and Wilson 2008, Mayer et al. 2024). We present here the world's first *Maximum Entropy Emergent Phonology* model, which has the delightful acronym MEEP.

In a MEEP model, WFC rankings are replaced with numeric WFC *weights*, with larger weights corresponding to higher rankings (stronger WFCs). The WFC weights and the violations of each individual candidate are used to compute a probability distribution over candidates, rather than simply choosing a single winner. Each candidate output *y* is assigned a *harmony score* based on its WFC violations:

$$H(y) \; = \; \sum_{i=1}^{K} w_i WFC_i(y),$$

where K is the number of WFCs, w_i is the weight of the i^{th} WFC, and $WFC_i(y)$ is the number of times candidate [y] violates the i^{th} WFC. A candidate that violates no WFCs receives a harmony score of 0, and a more positive score indicates a less preferred candidate.

The probability distribution over all possible candidates given a set of WFC weights is

$$P(y|w) \; = \frac{e^{-H(y)}}{\sum_{z \in \Omega} e^{-H(z)}} \; .$$

where Ω is the set of all candidates. In other words, the probability of a candidate is determined by its harmony (how many WFCs it violates) in proportion to the harmonies of other candidates.

Crucially, the weights assigned to each WFC can be learned from a dataset of observed outcomes by algorithmically choosing weights that assign the dataset the highest probability, allowing us to link our model directly to the quantitative patterns in the data. The weights in the model below were fit to the responses in the experiment using the maxent.ot R library (Mayer et al. 2024).

The weights of the two WFCs *[son]g[son] and UseFrequent in the fitted model were 0 and 1.14, respectively. The positive weight of UseFrequent indicates that, as the cost of the g-initial morph goes up (i.e., as its stand-alone frequency goes down), the likelihood of VVN occurring goes up. Surprisingly, the weight of zero on *[+son]g[+son] indicates that knowing whether the resulting compound violates the WFC against word-medial [g] does not add any additional predictive power after N2 frequency is taken into account. This is consistent with the results from Breiss et al. (in press).

Figure 1 shows the relationship between predicted VVN rates by the model and the VVN rates in the experimental data. While it is clear from this plot that there is much variability in the data that N2 frequency does not account for, the predictions of the model align fairly well in general with the observed VVN rates (r=0.76).

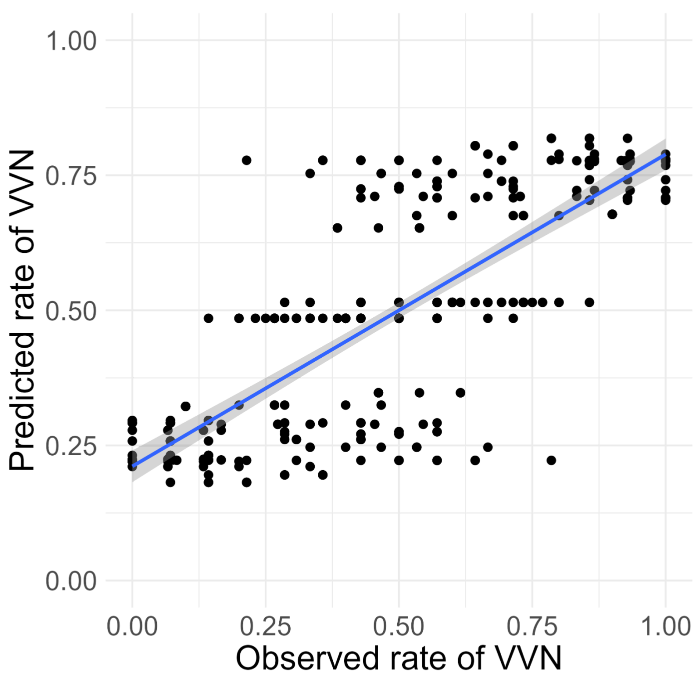

Figure 1. Observed rates of VVN in novel compounds from the experimental study in Breiss et al. (in press) plotted against rates of VVN predicted by our simple model (r=0.76).

The assessment tables below show the probability distributions over realizations of two compounds with the same N1: one with a high N2 frequency and one with a low N2 frequency.

{te:}$_{LOW}$ - {ge:, ŋe:}$_{ART}$	**Predicted Frequency**	**Observed Frequency**	**H**	***[son]g[son] w=0**	**UseListed w=1.14**
a. te:ge:	0.76	0.79	0.1596	1	0.14
b. te:ŋe:	0.24	0.21	1.29	0	1.13

{te:}$_{LOW}$ - {gai, ŋai}$_{APPEARANCE}$	**Predicted Frequency**	**Observed Frequency**	**H**	***[son]g[son] w=0**	**UseListed w=1.14**
a. te:gai	0.5	0.43	1.29	1	1.13
b. te:ŋai	0.5	0.57	1.29	0	1.13

In the former case, the pressure to use the high frequency g-initial N2 morph decreases the probability of VVN. In the latter, the low frequency of the g-initial morph N2 means that the form with VVN is chosen more frequently.

5. Discussion

The simple study above has demonstrated, consistent with Breiss et al. (2021a, 2021b, in press) that frequency effects can go at least part of the way towards explaining variability in rates of VVN. This is true even with a relatively impoverished picture of morph frequency that does not include information about ŋ-initial morph frequencies (because ŋ-initial morphs occur only in compounds and the [g]~[ŋ] distinction is not represented orthographically, it is difficult to estimate their frequency). The model predicts that the frequency of the ŋ-initial morphs should also influence VVN rates, and integrating this information could improve its predictions.

We noted earlier that the original analysis of VVN in Breiss et al. (2021a, 2021b, in press) found an additional effect of overall compound frequency: as whole compounds become more frequent, they are more likely to display VVN. This effect is not evident in novel compounds (since they have a frequency of 0), but appears in existing ones, both in corpus and experimental contexts. This indicates that usage frequencies at multiple levels of representation are relevant for understanding variability in VVN, and, accordingly, a phonological model must be able to refer to frequencies at multiple levels of structure. Archangeli & Pulleyblank (2022) anticipate this, proposing that "highly polymorphic forms may even be represented as part of a stem's morph set" (p. 148). They use the example of the inflected verbal form "looked", which could be generated by concatenating one morph each from the morph sets corresponding to LOOK and PAST, or by selecting the inflected form as a single stored, complex morph. The formal mechanisms of how stored structures at multiple levels interact in an EP model have not been fleshed out yet, but we suspect this approach holds much promise for modeling phenomena such as VVN.

We want to close by highlighting some connections between EP and our ongoing work in an Embodied Speech framework (ES; Gick 2019, Gick and Mayer in prep), which approaches speech from the perspective of a physical movement system and attempts to understand how the properties of movement systems are reflected in the higher level properties of speech. Our general approach aligns quite directly with the broad goals of EP, which similarly attempts to understand language as the product of domain-general mechanisms. The correspondences between EP and ES go deeper than this, however.

First, many researchers who study movement systems take a modular approach, suggesting that we develop a repertoire of stored movement routines of varying sizes (“chunks” or “motor programs”) that we draw on when we move, with selection between different chunks determined by task and environmental demands (e.g. Verwey 1996, Wolpert & Kawato 1998, d’Avella 2016, Schmidt et al. 2018). This corresponds with EP’s characterization of the phonological grammar as a mechanism for morph (chunk) selection. Second, the movement literature has also identified pressures to reuse more frequently executed chunks (e.g., Criscimagna-Hemminger and Shadmehr 2008, Loeb 2012, d’Avella 2016), which directly relates to the mechanism proposed here for encoding frequency effects.

These correspondences highlight the usefulness of an embodied approach in understanding why phonological systems should behave as they do. They also demonstrate how EP provides a bridge between embodied approaches to speech on the one hand, and traditional phonological approaches on the other. We look forward to continuing to explore these connections.

References

Archangeli, D., & Pulleyblank, D. (2015). Phonology without universal grammar. *Frontiers in psychology*, *6*, 1229.

Archangeli, D., & Pulleyblank, D. (2017). Phonology as an emergent system. In *The Routledge handbook of phonological theory* (pp. 476-503). Routledge.

Archangeli, D., & Pulleyblank, D. (2022). *Emergent phonology (Volume 7)*. Language Science Press.

Breiss, C., Katsuda, H., and Kawahara, S. (2021a). Paradigm uniformity is probabilistic: Evidence from velar nasalization in Japanese. In *Proceedings of WCCFL 39*. Cascadilla Press.

Breiss, C., Katsuda, H., and Kawahara, S. (2021b). A quantitative study of voiced velar nasalization in Japanese. In *University of Pennsylvania Working Papers in Linguistics*, volume 27.

Breiss, C., Katsuda, H., & Kawahara, S. (in press). Token frequency modulates optional paradigm uniformity in Japanese voiced velar nasalization. *Phonology*.

Coetzee, A. W., & Pater, J. (2011). The place of variation in phonological theory. *The handbook of phonological theory*, 401-434.

Criscimagna-Hemminger, S. E., & Shadmehr, R. (2008). Consolidation patterns of human motor memory. *Journal of Neuroscience*, *28*(39), 9610-9618.

d'Avella, A. (2016). Modularity for motor control and motor learning. *Progress in Motor Control: Theories and Translations*, 3-19.

Gick, B. (2019, August). How bodies talk. In *Proc. 19th Int. Congress of Phonetic Sciences, Melbourne, Australia* (pp. 5-9).

Gick, B., & Mayer, C. (in prep). *How Bodies Talk: Speech as an Embodied Movement System.*

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the workshop on variation within Optimality Theory* (pp. 111-120).

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379-440.

Loeb, G. E. (2012). Optimal isn't good enough. *Biological cybernetics*, *106*, 757-765.

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., ... & Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language resources and evaluation*, *48*, 345-371.

Mayer, C., Tan, A., & Zuraw, K. R. (2024). Introducing maxent. ot: an R package for Maximum Entropy constraint grammars. *Phonological Data and Analysis*, *6*(4), 1-44.

Prince, Alan & Paul Smolensky. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Cambridge, MA: Blackwell.

Schmidt, R. A., Lee, T. D., Winstein, C., Wulf, G., and Zelaznik, H. N. (2018). *Motor control and learning: A behavioral emphasis*. Human Kinetics.

Verwey, W. B. (1996). Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 544.

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural networks*, *11*(7-8), 1317-1329.