Modeling frequency-conditioned paradigm uniformity in Japanese voiced velar nasalization*

ANONYMOUS

Under review at *Phonology*

Abstract

Recent quantitative work on the variable [g]~[ŋ] alternation in compounds of certain dialects of Japanese has revealed *token frequency* of the compound as a whole, and of the compound's second-member (N2) in its freestanding form, to be important predictors of the alternation. In this paper, we propose a formal phonological analysis that integrates usage-based factors like frequency with the action of the phonological grammar, extending mechanisms of lexicon-grammar interaction previously proposed in the context of Lexical Conservatism. We demonstrate that our model fits the experimental data better than—or at least comparably to—a theoretically-naïve statistical model proposed in the previous work. Based on the success of our modeling, we discuss the role of token frequency in phonological patterning more broadly, and how the mechanism that we propose might be extended to unify a range of contradictory frequency-dependent processes that have been observed in the literature.

1 Introduction

This paper is about how to integrate information about *usage frequency*—here, the token frequency of morphemes in the language experience of an individual speaker—into a constraint-based phonological grammar formalism that characterizes that speaker's generative linguistic knowledge.

We take as our empirical case the frequency-conditioned variability in optional paradigm uniformity in voiced velar nasalization (henceforth "nasalization") in phonologically-conservative Japanese dialects, recently studied using corpus data by Breiss et al. (2021b) and experimentally

*Acknowledgments suppressed. Supplementary material can be accessed at *https://osf.io/wcvyx/?view_only=ca3fb8d413d04a76b60844fe78376594.*

verified in Breiss, Katsuda and Kawahara (*to appear*) — henceforth BKK. These studies are the latest in a long research tradition centered on the allophonic distribution of /g/ in conservative Japanese dialects, where a static phonotactic restriction enforces /g/ to be realized as [g] prosodic-word-initially and [ŋ] elsewhere (e.g. Kindaichi 1942; Trubetskoy 1969; Labrune 2012). This correspondence is disrupted in compounds with /g/-initial second member (N2) that can occur as a free morpheme: in compounds with N2s that do not occur as free-standing words, the /g/ $\rightarrow$ [ŋ] alternation is exceptionless, but in compounds where N2 may additionally occur as a free-standing word (that is, with initial [g]) the nasalization process is optional (Ito and Mester, 1996, 2003).

The contribution of recent work by BKK (reviewed in section 3) is to characterize this variation in quantitative detail, and in particular to highlight how the token frequency of both the compound and the free N2 impact the outcome of optional nasalization: higher frequency compounds encourage more nasalization of medial /g/ to [ŋ], while higher frequency free N2s encourage more retention of medial /g/ as [g], remaining uniform across the paradigm of their free-standing forms and compound forms (Steriade, 2000; Benua, 2000).

The novel contribution of this paper is to provide a formally explicit model of the experimental data. The model builds upon the Voting Bases model of lexicon-grammar interaction (Breiss, 2024), originally proposed to model Lexical Conservatism (Steriade, 1997). Lexical Conservatism is a type of paradigm uniformity where the distribution of stem allomorphs (referred to as "bases") in a paradigm influences the way that paradigm accommodates new members. The canonical example comes from Steriade (1997), who observed that the phonologically-similar forms *rémedy* and *párody* differ in their behavior when affixed with *-able*, yielding *remédiable* with shifted stress, but *párodiable*, with fixed stress. She argued that this difference stems not from the forms *remedy* and *parody* themselves, but from the fact that *remedy* has a stem allomorph *remédi-* in *remédial* that satisfies the marked lapse arising from affixation.

Breiss (2024) examined the same Lexical Conservatism dependency using novel derived forms (like *lábor* + *-able*, with related form *labórious*, and *pláster* + *-able* with no phonologically-advantageous related form), and found that in experimental settings, speakers are sensitive not only to the *presence* of the phonologically-beneficial stem allomorph (like *remédial* and *labórious*), but also to its salience in the lexicon as manipulated by priming. To account for these data, he proposed a formal phonological model that integrates the influence of the contents of the lexicon along with their resting activation, enabling the phonological grammar to be sensitive to the psycholinguistic properties of the morphemes which it manipulates. Breiss (2024) termed this formal model of lexicon-grammar interaction the *Voting Bases* model.

In this paper, we demonstrate that the Voting Bases model extends, without modification, to the separate case of lexicon-grammar interaction found in Japanese nasalization. The success of the model suggests that the foundational principles of the Voting Bases model may be a good candidate for a general theory of the way that the lexicon and grammar interact. This finding also underscores the explanatory value to be gained for phonological phenomena by adopting a more psycholinguistically-nuanced portrait of the lexicon as a dynamic substrate that can influence the computations of the grammar on the items which it contains. In section 6.3 we take up a series of questions which arise when adopting this boundary-blurring approach, in light of the traditional dichotomy between generative and usage-based perspectives on linguistic data.

The layout of the paper is as follows: the first two sections of the paper review in some depth basic facts about Japanese nasalization drawn from the literature (section 2), and then specifically reviews in detail Experiment 1 of BKK (section 3). Though this may not constitute new information, we hope the reader will find its inclusion helpful in contextualizing the theoretical analysis. The following section, 4, focuses on the Voting Bases model, and how we apply it to the context of optional paradigm uniformity. Section 5 then actually fits the model to the experimental results, and discusses relative and absolute model fit in comparison to minimally-different models that incorporate only some of the assumptions of the Voting Bases model. The paper closes in section 6 with a discussion of broader issues, touching on how such a system might come to be in the mind of the learner, on the merits of a joint model of psycholinguistic and grammatical influence on word formation, and on what a unified theory of token frequency effects on the phonological grammar might look like.

2 The traditional picture of Japanese nasalization

The data that we model in this paper comes from Experiment 1 of BKK, which investigated the variation between [g] and [ŋ] induced by the phonotactics of phonologically-conservative dialects of Japanese. The pattern, which has been well-studied in both descriptive (Kindaichi, 1942; Trubetskoy, 1969; Hibiya, 1995) and generative (Labrune, 2012; Ito and Mester, 1996, 2003) literature on Japanese linguistics, is exemplified in the complementary distribution of [g] and [ŋ] shown in the monomorphemic data in example (1) below, where the voiced oral velar stop is only permitted word-initially, and the velar nasal is only permitted word-medially.

(1) a. /kaŋami/ → [kaŋami]
"mirror"
b. /gimu/ → [gimu]
"obligation"

We assume throughout that non-alternating forms are stored surface-true as URs in the lexicon, in accordance with the phonological tradition of (Strong) Lexicon Optimization (Prince and

Smolensky, 1993; Sanders, 2003). This stance is supported by psycholinguistic research on the contents of lexical representations, reviewed in section 4.1.

Japanese's extensive use of compounding in word-formation gives the opportunity for the phonotactic restriction to drive alternations, seen in examples (2)-(5) below. Here we see that when a /g/-initial morpheme is word-initial (either as a prosodically-free word, in examples (2)–(4), or as the first member (N1) of a compound, in example (5)[1]), it is realized with an initial [g], while when it occurs as the second member of a compound (N2) it is realized with initial [ŋ]. Critically for the current study, Ito and Mester (1996) observed that although in all cases the /g/-initial N2 *may* be realized word-medially with initial [ŋ], nasalization is optional when the N2 can stand on its own as a prosodically-free form (cf. the "b" series in (2)–(4) vs. (5c))—a case of optional paradigm uniformity.

(2) a. /hai + gan/ → [hai-ŋan] ~ [hai-gan]
lung cancer
"lung cancer"

b. /gan/→ [gan]
cancer
"cancer"

(3) a. /noo + geka/ →[noo-ŋeka] ~ [noo-geka]
brain surgery
"brain surgery"

b. /geka/→ [geka]
surgery
"surgery"

(4) a. /doku + ga/ → [doku-ŋa] ~ [doku-ga]
poison moth
"poison moth"

b. /ga/ → [ga], "moth" (a free-standing morpheme)

(5) a. /doku + ga/ → [doku-ŋa], *[doku-ga]
poison fang
"poison fang"

b. /ga + ʒoo/→ [ga-ʒoo]
fang castle
"main castle"

[1]We temporarily adopt here for the traditional assumption that the [g]-initial form of a free N2 is underlying, for expository ease and continuity with the previous literature. Our own proposal is laid out in section 4.

c. /ga/→ *[ga] (a bound morpheme)
fang
“fang”

Breiss et al. (2021b) examined this variation in a corpus derived from a pronunciation dictionary (NHK, 1993) and found that among compounds with free N2s, the two most prominent predictors of whether an item would be nasalized was the frequency of the N2's free [g]-initial form, and the frequency of the whole compound. These effects ran in opposite directions: higher frequency compounds were more likely to be nasalized (the left facet of Figure 1); on the other hand, the more frequent the free N2, the less likely the nasalization was (the right facet of Figure 1).

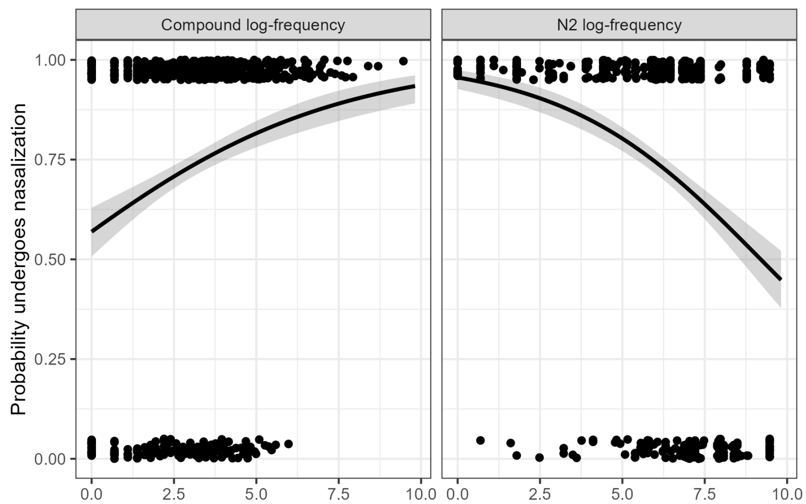

Figure 1: The effects of whole compound frequency (left facet) and N2 frequency (right facet) on the probability of nasalization (vertical axis), with binomial smooths in the corpus data. One dot represents one lexical item; vertical jitter has been added for readability. Figure and caption adapted with permission from Breiss et al. (*to appear*), data from Breiss et al. (2021b).

The corpus data was modeled as a case of probabilistic paradigm uniformity in Breiss et al. (2021a) using Output-Output Faithfulness constraints (Benua, 2000) indexed to items binned by the relative frequency of each compound and N2. The paper was limited, however, by the untested assumption of their model that the frequency-modulation of paradigm uniformity in their corpus data actually represents the synchronic knowledge of speakers. Additionally, their formal model was not explicitly informed by psycholinguistic considerations and thus its linking hypothesis between frequency (necessarily a lexical characteristic) and the phonological grammar had a problem of simply being stipulative—in other words, there was nothing in their model that prevented the opposite relation between frequency and paradigm uniformity from holding.

In this paper, we offer two improvements on the state of affairs in Breiss et al. (2021a). First, we model experimental data from Breiss et al. (*to appear*) (BKK) where the frequency-conditioning of the variable paradigm uniformity is reproduced in existing compounds and extended to novel ones. Second, we do this by extending the Voting Bases model of Breiss (2024) which is compatible with consensus understanding of the way lexical frequency is connected to the lexical representation and activation, and which offers an explicit linking hypothesis relating the real-time dynamics of the lexicon to the representation and computations of the phonological grammar.

3 BKK's Experiment 1

BKK carried out two experiments on Japanese nasalization, with the goal of seeing whether the corpus patterns were representative of speakers' generalizable knowledge, both in the aggregate as individuals. They found that both individually and in aggregate, speakers' propensity to nasalize displayed sensitivity to the frequency of the free N2 and compound, in existing and novel compounds. In this paper, we focus our modeling efforts on the results of their Experiment 1, which we describe in some detail below.[2]

BKK's stimuli were roughly balanced between existing Japanese compounds of varying frequencies, and novel (i.e., zero-frequency) semantically-compositional Sino-Japanese compounds. Both existing and novel stimuli had attested free N2s of a range of frequencies. Out of a desire to sample compounds with a wide range of frequencies that would likely be known to participants, existing compounds ranged from two to eight moras in length, while all novel compounds were four moras long. Complete details of the experimental materials are available from BKK's OSF repository[3]..

BKK recruited speakers of the phonologically-conservative Tōhoku dialect of Japanese, and used a short dialect questionnaire to ensure that their speech exhibited the allophonic distribution of word-initial [g] and word-medial [ŋ]. For the purposes of the model which we develop, we will see that these monomorphemic words provide crucial evidence for the lower bound of the weight of the markedness constraint driving nasalization, since with data from compounds alone, it is not uniquely identified against the background of faithfulness constraints that the Voting Bases model uses (see section 5 for further details).

The dialect questionnaire consisted of a production task where speakers were asked to read aloud 10 monomorphemic words with word-initial [g] of varying frequencies, and 10 monomor-

[2]They also sought to determine whether correlation between nasalization and the overall prosodic size of the compound, which is observed in the corpus (Breiss et al., 2021b) but is a typologically unusual pattern, was replicated in participants' online productions (Experiment 2). They actually found that there was no evidence of a direct relationship between nasalization and global prosodic length (cf. Jiang 2023). We therefore do not address this experimental data here, as our point is made in the simpler case of data from Experiment 1.

[3]`https://osf.io/avnpw/?view_only=cd2afdcc183f4de3ac1261b4af66f08d`

phemic words with word-medial [ŋ]. The stimuli were written with *kanji* orthography, which does not distinguish between [g] and [ŋ]—this is also true of the main production experiment described below, so we follow BKK in assuming that the participants' production was not influenced by orthographic factors. The twenty words were shown to the participant in a random order, and their productions were recorded; only the eight participants who exhibited the target pattern of allophony in all monomorphemes were invited to participate in the main experiment.

After this knowledge check, participants saw each compound one at a time in a random order, and produced the form aloud while their speech was recorded. Participants also produced and indicated knowledge of all of the free N2s in the experiment, as well as all of the compounds. See Breiss et al. (*to appear*) for complete details.

3.1 Results

BKK found that the participants reflected at an individual level the frequency-conditioned variability seen in the corpus study of Breiss et al. (2021b). In existing compounds (Figure 2), their productions were influenced by both the frequency of the compound (the left facet), for which higher values correlated with more nasalization, and by the frequency of the free N2 (the right facet), where higher values correlated with less nasalization.

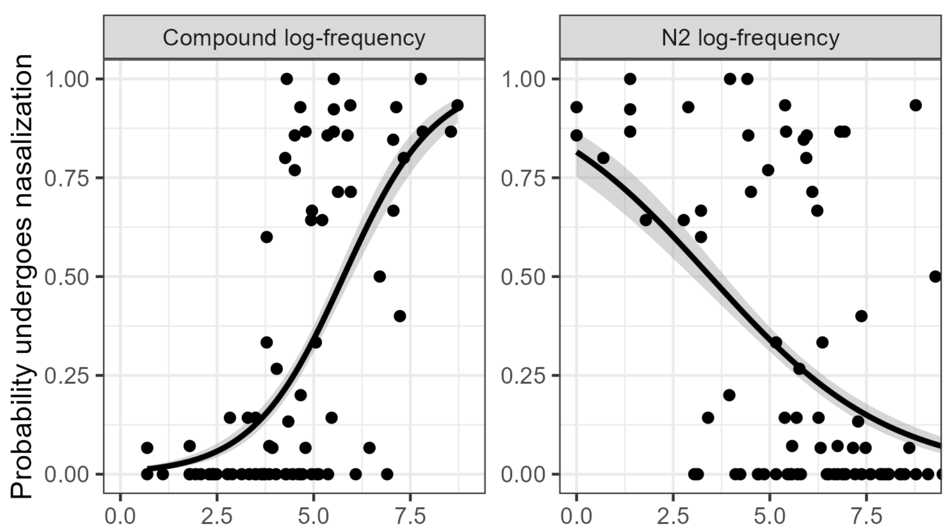

Figure 2: Probability of nasalization (the vertical axis) plotted against compound log-frequency (the left facet) and N2 log-frequency (the right facet) in existing words, with binomial smooths for readability, in the experiment by BKK. Plot and caption reproduced with permission from Breiss et al. (*to appear*).

Figure 3 plots the same effect of N2 frequency in novel compounds: forms with higher-frequency N2s were less likely to undergo nasalization relative to those with lower-frequency

N2s.

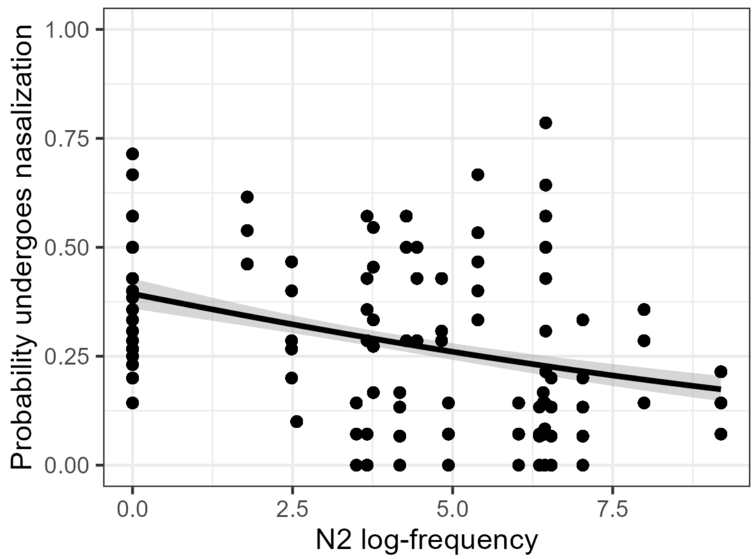

Figure 3: The probability of undergoing nasalization in novel compounds, plotted against N2 log-frequency, with a binomial smooth to aid readability. Plot and caption reproduced with permission from Breiss et al. (*to appear*).

Finally, BKK found that the frequency effect was stable at the level of the individual, across existing and novel compounds, which is plotted in Figure 4. In this figure, the horizontal axis plots the strength and direction of the effect of N2 log-frequency in novel compounds, and the vertical axis plots the strength and direction of the effect of N2 log-frequency in the existing compounds; see the caption of Figure 4 for further details. Although different participants were more or less sensitive to the frequency of a given N2, lying higher or lower on each axis, there was uniformity in this degree of sensitivity such that the two co-varied along a diagonal line through the center of the plot. BKK interpreted this correlation as evidence that morpheme usage frequency and phonological markedness have separable, distinct influences on speaker productions.

3.2 Summary and goals for modeling

To summarize, the findings of BKK that are relevant for the modeling task of this paper are the following. Among those speakers for whom the phonotactic restriction enforcing [g]~[ŋ] allophony was exceptionless in monomorphemic words:

1. Phonotactically-driven nasalization is variable in compounds with free N2s.

2. In these compounds, the probability of nasalization is increased by higher compound frequency, and decreased by higher N2 frequency.

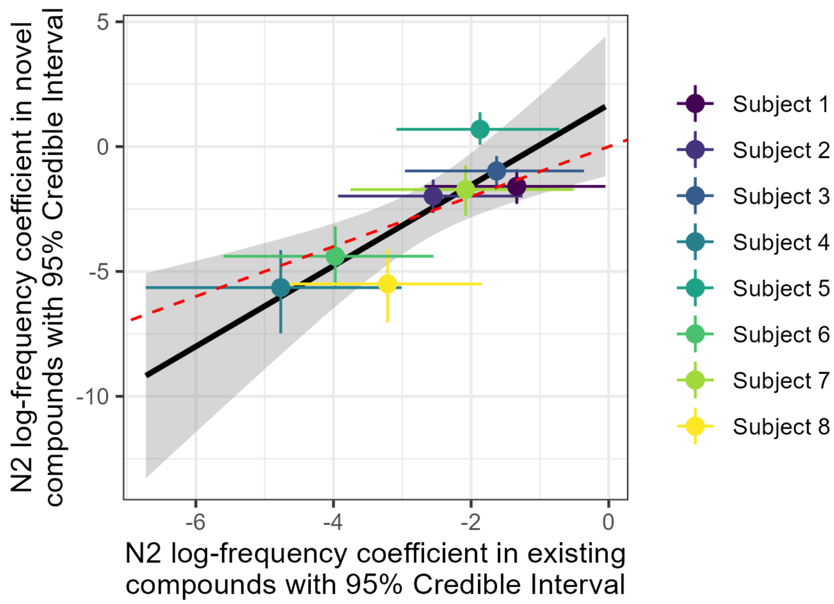

Figure 4: The coefficient of N2 log-frequency in novel compounds, derived from the model in Table 1 of Breiss et al. (*to appear)*, is plotted on the horizontal axis, and the coefficient for N2 log-frequency in existing compounds, derived from the model summarised in Table 3 of Breiss et al. (*to appear*), is plotted on the vertical axis. Points represent median values of the posterior with ranges encompassing the 95% Bayesian Credible Intervals, colors represent speakers, and a linear smooth has been added for readability, with the line of slope 1 intersecting the origin in dotted red. Plot and caption adapted with permission from Breiss et al. (*to appear*).

3. The frequency effect is uniform within individuals across existing and novel compounds.

Below, we propose a formal model of these facts, using the Voting Bases model to relate a lexicon containing usage-frequency information to a phonological grammar couched in the Maximum Entropy (MaxEnt) framework.[4]

4 Modeling token frequency in the phonological grammar

Based on the facts laid out above, we seek a model of the phonological grammar that allows non-phonological properties of individual lexical items (here, frequency) to influence their participation in phonological processes (here, paradigm uniformity). Note that we specifically aim to model phonological and non-phonological influences on the outputs of the phonological grammar, rather than any possible morphological or paradigmatic effects on phonetic realization (see

[4]We do not attempt to model the frequency of the first compound member, N1, on the probability of nasalization in compounds, since this was not manipulated by BKK. Future work might profitably pursue this question experimentally and formally, since corpus data in Breiss et al. (2021b) suggests that higher N1 frequency may also independently lower the probability of nasalization; see Rebrus and Törkenczy (2017) for a similar finding of N1 frequency on compound coherence in Hungarian vowel harmony.

Purse et al. 2022 for a review), about which the Voting Bases model as laid out in Breiss (2024) makes no predictions.

4.1 The contents of a lexical entry

As prolegomena to the grammatical model, it will be important to establish some relevant context regarding the contents of the lexicon, because it is these representations that are at stake in discussions of token frequency. Psycholinguistic research has amassed a large body of evidence that the lexicon is richly structured, with numerous types of linked representations of various levels of detail grouped under the same lexical entry. We do not review this research in depth here, but simply highlight the findings relevant to developing the type of integrated phonological theory referenced above. For a thorough discussion and literature review on the (phonologically-relevant) contents of a lexical entry, see Pierrehumbert (2016); for more on how this information interacts with the Voting Bases theory in cases beyond those relevant for the nasalization, see Breiss (2021, 2024).

Since nasalization concerns paradigm uniformity, we assume the lexical entry for an existing word lists (among many other things) their allomorphs (cf. Strong Lexicon Optimization embraced by Sanders 2006, as well as arguments by Wang and Hayes 2025 on the sufficiency of less-abstract URs): for a non-alternating monomorpheme like [kaŋami] "mirror", this would be simply /kaŋami/; for a monomorpheme that can appear as an N2 and undergo nasalization, such as [ga]–[ŋa] "moth", the lexical entry would list both /ga/ and /ŋa/. Finally, we assume that existing compounds are stored whole, with nasalization applied so as to respect the phonotactic in the lexicon (Albright, 2008; Martin, 2007).[5]

With regard to non-phonological characteristics of the lexicon, we follow a large body of evidence that lexical representations have differing degrees of salience or strength of encoding. Following Breiss (2021, 2024), we refer to this quantity as *resting activation*, borrowing the term (though not the theory) from Morton (1970), which corresponds to the strength of a memory representation itself, not a number or rank stored in long-term memory as a characteristic of the lexical item. Thus, characteristics (long-term or dynamic) of lexical items like their frequency, and whether or not they were recently activated (for example, by priming), all contribute dynamically to an item's resting activation. Importantly, also following Breiss, we use the term "resting activation" as a stand-in for any scalar summary statistic that can be derived from an implemented model of lexical dynamics. We remain intentionally agnostic as to the specific model of

[5]On the suggestion of a reviewer, we relaxed this assumption by fitting a comparable model but assuming stored allomorphs for both oral and nasal forms of the compound, with corresponding faithfulness constraints for each. Such a model returns weights and fits to the data identical to the one without the relevant faithfulness constraint, indicating that it is thus at best superfluous in explaining the data. This exercise shows that our assumption here is well-founded, or at least benign. Details of the model fit can be found in the supplementary materials.

these dynamics, whether the specific model endorsed by Morton (1970) or not, simply stressing that so long as such a model can be used to drive a measure of relative salience influenced by the factors just mentioned, the Voting Bases model can make reference to it to scale faithfulness constraint violations (cf. e. g. Luce and Pisoni, 1998). We discuss how resting activation is modeled as influencing the phonological grammar below in section 4.4.

4.2 The Voting Bases model

We now turn to a formal phonological model of the Japanese nasalization data. We use the Voting model of Base competition proposed in Breiss (2021, 2024). The Voting model has been used to model data in Lexical Conservatism in English and Spanish, and is broadly compatible with the view of the lexicon laid out above. Here, we extend the scope of the model by analyzing the probabilistic paradigm-uniformity found in Japanese nasalization.

The Voting Bases model has two parts: the first is that all listed stem allomorphs in the lexicon exert an analogical pull on derivatives (operationalized using allomorph-specific faithfulness constraints), violations of which are scaled in proportion to the resting activation of the representation to which faithfulness is being assessed. We note that the terminology of "bases" comes from the original context for which the model was developed, but here the term can be read as a synonym for "stored allomorph".[6] The second part is that markedness constraints evaluate candidates in the standard way for any constraint-based phonological models.

The Voting Bases model assumes a probabilistic, weighted-constraint phonological grammar; here, we use MaxEnt Harmonic Grammar (Smolensky, 1986a; Goldwater and Johnson, 2003), but in principle we could also use another grammar formalism that has these characteristics, like Stochastic (or Noisy) Harmonic Grammar (Boersma and Pater, 2016). We use MaxEnt since it has various strengths; e.g. it directly relates Harmony to probability (Hayes, 2022), permits constraint cumulativity by default (Jäger and Rosenbach, 2006; Breiss, 2020), has a learning algorithm to set its weights, and is rooted in well-understood statistical techniques used widely outside linguistics (Jurafsky and Martin, 2009, ch. 5). We stress, however, that our analyses can be recast in terms of other stochastic constraint-based frameworks.

[6] The probabilistic paradigm-uniformity might, as a reviewer points out, be captured in terms of Output-Output faithfulness (Benua, 2000) instead of the Voting Bases model. This approach was pursued in Breiss et al. (2021a), but ultimately we abandon it here because it fails to explain the correlation between the degree of faithfulness to a non-local paradigm form and the relative frequency of the two forms in question. In the Voting Bases model, this relationship has a clear source by virtue of the explicit location of both URs in a psycholinguistically-dynamic lexicon; for a more extended comparison between these two approaches, see discussion in Breiss (2024, fn. 5).

4.3 Constraints

In the analysis developed in this paper, we adopt the general approach of Ito and Mester (1996, 2003), following loosely Breiss et al. (2021a). We only use three constraints: a single markedness constraint to motivate nasalization (extending the spirit of the constraint *VgV from Ito and Mester 2003 to be compatible with nasal-final N1s, which pattern identically to vowel-final N1s), and a pair of faithfulness constraints which correspond to the second member's free form and to the analogical pull of the compound as a whole, if one exists. They are listed below.[7]

- ***Internal-[g]**: Assign one violation for each word-internal [g].
- **Id-[nasal]-N2**: Assign one violation for each segment in the listed allomorph for the free-standing N2 that does not match its corresponding segment in the feature [nasal].
- **Id-[nasal]-Compound**: Assign one violation for each segment in the listed allomorph for the full compound that does not match its corresponding segment in the feature [nasal].

Note that the constraint definitions do not make reference to scaling or the contents of the lexicon; the proposal in the Voting Bases model is an architectural proposal about how psycholinguistic, "extra-grammatical" factors act within and beside the phonological grammar to influence certain variable phenomena.

4.4 Modeling resting activation

The discussion in 4.1 above left open how a specific numerical value for resting activation might be calculated on the basis of the psycholinguistic characteristics of item's lexical entry. Here, we model the data using the log-frequency of the allomorph, passed through a sigmoid function $\frac{1}{1+e^{-logfreq}}$ that translates the linear predictor (i.e. $-logfreq$) into the bounded interval of {0,1}, which will be the scaling factor applied to faithfulness violations. This is illustrated in Figure 5. The effect of this non-linear transformation will be to preserve the idea that it is less penalized to be unfaithful to low-frequency lexical items compared to higher-frequency ones, while damping down the difference between extreme values of the scale and rendering it bounded.

The final move we make here is rather than using *raw* log-frequencies, we use *scaled and centered* log-frequencies, following the statistical analysis in BKK. This corresponds to the notion that it is not so much the *absolute* frequency of each item that is important, but how frequent it is relative to the other competitor items in the lexicon (here approximated by the population of

[7] The first faithfulness constraint plays the same role as faithfulness to the Remote Base in an analysis of Lexical Conservatism. The second faithfulness constraint parallels faithfulness to the Local Base in a Lexical Conservatism analysis (Breiss, 2021, 2024). We use more transparent names here for the sake of clarity, since nothing in the Voting Bases model structurally prioritizes Local Bases over Remote ones.

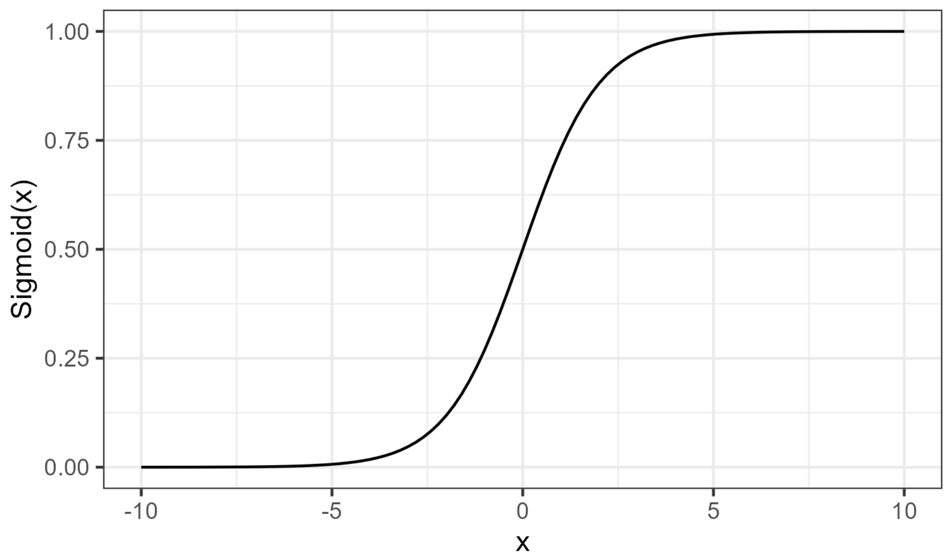

Figure 5: Sigmoid function that translates the (centered) frequencies into the scaling factors. See text for details.

items in the experiment), which is in line with previous work on morphological decomposition in stored forms (Hay, 2001). Finally, in the analysis that we develop below, we do not model the priming of N2, since BKK did not find substantial evidence that it affected their experimental data.[8]

4.5 Schematic illustrations

Before modeling the experimental data itself, it will be useful to work with some toy data to get a feel for how resting-activation-scaled faithfulness violations interact with the dynamics of a MaxEnt grammar. First, let us consider the case of novel compounds, since they are the simplest case to lay out the workings of the analysis. Recall the empirical pattern: here, although the frequency of the compound is zero, we nevertheless find that nasalization is modulated by the frequency of N2. Now, consider the case of two hypothetical novel compounds, one with a higher-frequency N2, and one with a lower-frequency N2, such that when the sigmoid transformation is applied to their frequencies the higher-frequency form scales its violations of Id-[nasal]-N2 by 0.7, and the lower scales its own violations of the same constraint by 0.3 (these specific numbers

[8] The Voting Bases framework is easily extensible to multiple predictors of resting activation: to incorporate priming, for instance, one could simply treat the term passed into the sigmoid as itself a log-linear model, adding a coefficient (weight) for the effect of priming, in addition to a coefficient for the effect of lexical frequency. This is beyond the current scope of this paper, however, and so we simply assume a fixed coefficient for lexical frequency, since there being only one predictor in the log-linear model for resting activation would make the coefficient of frequency redundant with the weight of the faithfulness constraint being scaled. Similarly extensions of the Voting Bases model could also model by-participant variability in the priming effect using a hierarchical model structure.

are chosen purely for the sake of illustration). Using the constraints defined in section 4.3 above, we can define the tableaux below in Figure 6.

/…/$_{N1}$, /g…/$_{High-freq.N2}$ Weight:	*Internal-[g] 2	Id-[nas]$_{N2}$ 1	H	p
a. […g …]	1		2	.21
b. […ŋ …]		.7	.7	.79
/…/$_{N1}$, /g…/$_{Low-freq.N2}$ Weight:	*Internal-[g] 2	Id-[nas]$_{N2}$ 1	H	p
c. […g …]	1		2	.15
d. […ŋ …]		.3	.3	.85

Figure 6: Schematic application of the Voting model of Base Competition to the formation of a novel compound in the *wug*-test.

We can see that the pull of faithfulness to the N2 with higher frequency is stronger than the one with lower frequency, though both are relatively marginal outcomes since the weight of *Internal-[g] dominates the distribution of probabilities in this scenario.

Moving on to existing compounds, we now must add another item to the lexical entry we are considering in our left-hand input cell to our tableaux, shown in Figure 7. For the sake of minimal contrasts, we assume that the frequency of both N2s are equal and medial relative to the examples in Figure 6 above, allowing us to examine the effect of compound frequency holding N2 frequency constant. However, in our analysis of the actual data, both scaling factors are independently set on a per-item basis.

/…/$_{N1}$, /g…/$_{N2}$, /…ŋ…/$_{High-freq.compound}$ Weight:	*Internal-[g] 2	Id-[nas]$_{N2}$ 1	Id-[nas]$_{Compound}$ 1	H	p
e. […g …]	1		.7	2.7	.09
f. […ŋ …]		.5		.5	.91
/…/$_{N1}$, /g…/$_{N2}$, /…ŋ…/$_{Low-freq.compound}$ Weight:	*Internal-[g] 2	Id-[nas]$_{N2}$ 1	Id-[nas]$_{Compound}$ 1	H	p
g. […g …]	1		.3	2.3	.14
h. […ŋ …]		.5		.5	.86

Figure 7: Schematic application of the Voting model of Base Competition to the formation of an existing compound in the *wug*-test.

Here we see that the scaling of the compound again depends on frequency, but because of the assumption we made about the listed form of the compound—specifically, that phonologically well-formed words are preferentially the target of lexicalization (Albright, 2008; Martin, 2007)—we find that the faithfulness to the compound's UR penalizes the candidate that does not exhibit nasalization and violates markedness.

Finally, we lay out the case where the competition between candidates is driven primarily by faithfulness. Above, where markedness had a high weight, the candidate that satisfied markedness had a higher probability than the one which violated it, and the effects of the faithfulness constraints were on the probability of the minority candidate. In the scenario where markedness is low and the weights of the faithfulness constraints are dominant, the majority candidate is the one that satisfies faithfulness to the whole compound, and the presence of the N2 is the main reason that the unfaithful (but markedness-satisfying) candidate gets appreciable probability; this is a type of "analogical" effect where markedness has little role, as in Figure 8, in which the markedness constraint is assigned a very low weight (here, arbitrarily set as 0.1). Below, we will see that this scenario is most similar to the state of the VVN alternation.

$/.../_{N1}$, $/g.../_{N2}$, $/...ŋ.../_{High-freq.compound}$ Weight:	*INTERNAL-[g] 0.1	ID-[nas]$_{N2}$ 1	ID-[nas]$_{Compound}$ 2	H	*p*
i. [...g ...]	1		.7	1.5	.27
j. [...ŋ ...]		.5		0.5	.73
$/.../_{N1}$, $/g.../_{N2}$, $/...ŋ.../_{Low-freq.compound}$ Weight:	*INTERNAL-[g] 0.1	ID-[nas]$_{N2}$ 1	ID-[nas]$_{Compound}$ 2	H	*p*
k. [...g ...]	1		.3	0.7	.45
l. [...ŋ ...]		.5		0.5	.55

Figure 8: Schematic application of the Voting model of Base Competition to the formation of an existing compound in the *wug*-test, in a regime where faithfulness is strong and markedness weak.

5 The model in action

Moving on to the analysis itself, we fit a model to the data from existing compounds and monomorphemes, assessing its fit in that setting as well as its generalization to data from novel compounds. We fit the MaxEnt models using the *Solver()* function in Microsoft Excel (Fylstra et al., 1998), and used a weakly informative Gaussian prior of Normal(0,10) on constraint weights, which has the effect of allowing weights to vary in response to values that best fit the data, while making extreme values (here, above twenty or so) less appealing. For more on priors on weights in MaxEnt phonological models, see Wilson (2006) and White (2017). All models fit in this paper are provided in the supplementary materials.

5.1 Existing compounds

We first applied the analysis sketched in section 4.5 to data from existing compounds. Recall that in these forms, compounds with higher-frequency N2s are more likely to resist nasalization than those with lower-frequency N2s, but that compound frequency itself also influences nasalization, with higher-frequency compounds favoring the surface-realization of their underlying [ŋ]. We model the counts of compounds produced having undergone nasalization or not.

We also integrate the fact that the participants were included in the experiment on the basis of exhibiting complementary distribution of [g] and [ŋ] in monomorphemes. Therefore, the model included the monomorphemes used in the dialect questionnaire to screen participants for inclusion in the experiment, including frequency-based scaling of their faithfulness violations. Since we assume lexicon optimization (i.e., non-alternating monomorphemes are restructured to be /ŋ/-ful), and the monomorphemes we surveyed are only a small subset of the lexicon that exhibits the complementary distribution of [g] and [ŋ] and so do not allow us to train phonotactic learning models that rely on implicit negative evidence (Hayes and Wilson, 2008), we cannot accurately assess the weight of *Internal-[g]. However, we can find a lower bound on its weight by constraining the sets of weights we consider to those that maximize the likelihood of the compound data, while simultaneously preserving allophony in monomorphemes (operationalized as having 95% or greater probability of faithful realization). The final model yielded weights listed in Table 9, and predictions plotted in Figure 10.

The weights of the two faithfulness constraints were not significantly different from one another, as assessed via a likelihood ratio test: Δlog-likelihood $= 1.3$, $p = 0.10$; a similar conclusion was suggested by the near-zero difference in the sample-size corrected AIC of the two models: ΔAICc $= 1.8$. AICc differences greater than 10 are typically taken to indicate strong support for the model with the lower AICc value; for more on model-comparison in statistical models and phonological grammars, see Shih (2017) and Wilson and Obdeyn (2009). This result suggests that the attractive influence of both bases is critical in driving the alternation in attested forms; the zero weight of the markedness constraint *Internal-[g] indicates that in existing compounds, analogical faithfulness is doing all the work, despite the assumption in the literature that the alternation is markedness-driven. We will revisit the role of markedness below in section 6.2.

We also compared the full model to one where the two faithfulness constraints were allowed to take on different values but were not scaled by frequency. As one might expect, since low- and high-frequency forms have the same violation profiles in the phonological grammar, a grammar without access to frequency information can only predict one rate of nasalization across all forms;

Constraint	Weight
*Internal-[g]	0.0
Id-[nasal]-Compound	7.09
Id-[nasal]-N2	7.39

Figure 9: Best-fitting weights for the experimental data, existing and novel compounds combined, that preserves the allophony in monomorphemes.

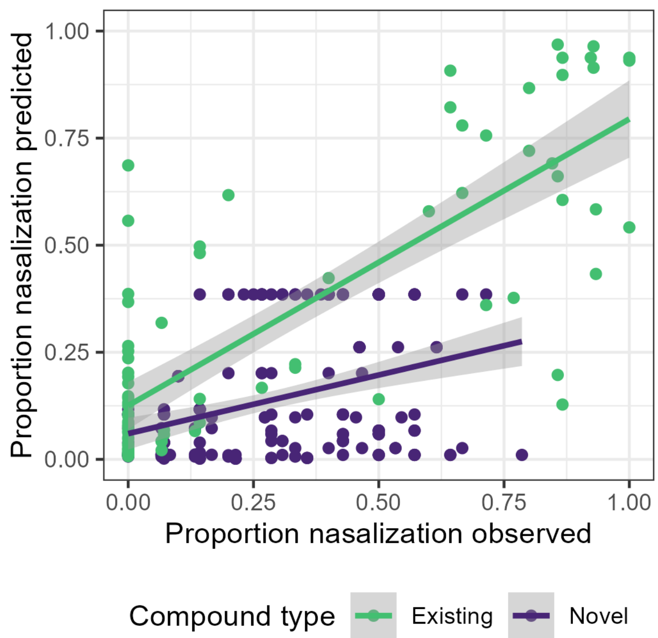

Figure 10: Predicted (vertical axis) vs. observed (horizontal axis) rates of nasalization for categories existing (green) and novel (purple) compounds under the combined model (weights in Table 9).

this model fits the data dramatically less well (Δlog-likelihood $= 264.57$, $p < .001$ with one degree of freedom, ΔAICc $= 526.50$).

Finally, we evaluate the absolute performance of the model by examining how well it fits the data it was trained on: although the two models have different internal structures, we can ask whether the theoretically-informed MaxEnt model here does as good a job in explaining the data patterns as the theory-neutral mixed effects logistic regression model reported by BKK.[9] We do this using the measure of R^2, which ranges from zero to one, and can be thought of as the proportion of the variation in the dependent variable (here, whether nasalization applies or not) explained by the collection of independent variables (the phonological and lexical characteristics of interest).

We used the *r2_bayes()* function from the *performance* package (Lüdecke et al., 2021) to obtain the marginal R^2 of the statistical model—that is, the amount of variance in the data explained by the fixed effects—and compared it to the R^2 for the MaxEnt model.[10] Since the statistical

[9]The model specification in BKK was as follows: `Nasalization ~ 1 + LogN2Freq*N2Primed + LogCompoundFreq + NasalFinalN1 + (1 + LogN2Freq*N2Primed + LogCompoundFreq + NasalFinalN1 | Speaker) + (1 + N2Primed | Compound)`; see BKK section 3.3 and 3.4.1 for details.

[10]We used marginal R^2, which makes reference to fixed effects only, since the conditional R^2 that takes into account the variance explained by both fixed and random effects has no direct comparison in the MaxEnt model

model is Bayesian, we obtain a median and 95% Credible Interval for our R^2: 0.48 and [0.31, 0.56], respectively. This is lower, though still relatively comparable, to the MaxEnt models R^2 of 0.63, for which we have only a point estimate. Although the two are relatively close, the point value for the marginal R^2 of the MaxEnt model is outside the 95% Credible Interval of the statistical model; this comparison suggests that the theoretically-structured model out-performs the theory-blind statistical one. While we find this result to be encouraging, this conclusion is tentative, however—since the MaxEnt model does not capture variation at the level of the speaker, it may be that the non-hierarchical structure of the model mismatches the structure in the data in a way that distorts the results, attributing to the population grammar variance that should more conservatively be attributed to speaker-level idiosyncrasies.

5.2 Novel compounds

We take advantage of the fact that we have data for both existing and novel forms to administer a more severe test of the model. We do this by asking how well the grammar that was fit to the existing items generalizes to the novel forms. We evaluate the probability of the two candidate outcomes in the novel forms using the learned weights reported in Figure 9, with the relevant frequency information for the N2s, and zero frequency for compounds (since they are novel). The fit to the data is shown in Figure 10, alongside the fit to the existing compounds.

We found that the model of the existing forms generalizes to the novel forms quite poorly; the obvious problem is that while the range of attested proportions of nasalization range between 0 and 0.79, the model predicts outcomes only in the range of 0.002-0.39. This indicates that either the grammar that best fits the existing compounds is a poor estimate of the knowledge that speakers used when generalizing to novel compounds (due to incompatible weights, constraints, or both), or that the novel compound data is simply extremely variable. To check whether the mis-fit is due to incompatible constraint weights, we fit a model with the same constraints to the novel forms directly, without regard for data from monomorphemes or existing forms. This yields a shifted range of predicted proportions (0.29-0.70), but a only marginally lower R^2 (0.11, compared to 0.12 based on the grammar fit to only the existing forms), which indicates that the data are still poorly fit. Therefore, the unexpected finding about the faithfulness-driven nature of the alternation is not to blame for the poor generalization performance of the model.

Next, we compare the fit of our theoretically-motivated MaxEnt model to the purely statistical model fit by BKK[11], and find that the R^2 of our primary model when generalizing to the novel

we fit. For more on the relationship between mixed effects models and hierarchical structures in linguistic data, see Zymet (2019).

[11]Model specification: `Nasalization ~ 1 + LogN2Freq*N2Primed + (1 + LogN2Freq*N2Primed | Speaker) + (1 + N2Primed | Compound)`; see BKK section 3.3 and 3.4.2 for details.

compounds, 0.11, falls within the 95% Credible Interval of the median of that of the statistical model, 0.06 [0.00, 0.18]. The model also predicts a range of 0.07-0.64 in terms of proportion undergoing nasalization, which more closely matches the true data. While still low in absolute terms, it lessens the possibility that our theoretical commitments are what is limiting us in being able to account for the data well. Therefore, we suspect that the cause of the poor model fit may be that there is simply less signal in the novel compound data.

6 Discussion

This paper has proposed a model of variable voiced velar nasalization in Japanese, drawing on experimental data published in Breiss et al. (*to appear*). The model integrates grammatical and functional determinants of variation, drawing on the Voting Bases framework of lexicon-grammar interaction, which was originally developed to model an entirely separate phonological phenomenon, Lexical Conservatism in English and Spanish (Breiss, 2024). Here, we address several major issues that the model raises, notably about whether the proposed system can be learned from the actual Japanese lexicon (section 6.1), about the unexpected (lack of) role markedness appears to play in driving the alternation (section 6.2), about the competence-performance distinction (section 6.3), and about how the Voting Bases model's mechanism for integrating usage frequency and formal grammar compares to other propositions in the literature (section 6.4). Finally, we close the paper with a more general discussion about how we might understand the broader empirical landscape of frequency effects in phonological patterning in light of the proposal in this paper.

6.1 Whence the weights? Evidence in the lexicon

Having observed that there is robust frequency-conditioning of nasalization in both existing and novel compounds, we can ask what the source of this frequency-conditioning might be. A sensible null hypothesis would be that the relationship between frequency and resting activation is one that is automatic and not overtly learned. However, we find that the model performs significantly better when allowed to set the weights of faithfulness constraints referencing different allomorphs to different weights. This result suggests that, setting aside the relationship between frequency and activation, the speakers must be able to attribute different amounts of influence to different faithfulness constraint violations depending on which base the violation is assessed against. Put another way, the learner needs to be able to figure out how analogically-driven her lexicon is. Here, we present a preliminary investigation of what kind of evidence might exist in the Japanese lexicon that could allow speakers to assign different weights to Id-[nasal]-Compound and Id-[nasal]-N2.

We fit a grammar with the constraints in section 4.3 and frequency-driven scaling of faithfulness violations to the set of compounds in the corpus analyzed by Breiss et al. (2021b) that had a free N2. We found that the optimal weights of the grammar were zero for both *Internal-[g] and Id-[nasal]-Compound, and 1.08 for Id-[nasal]-N2. We had anticipated there being little to no weight assigned to the markedness constraint in this dataset for the same reasons discussed above in section 5, but we also found that instead of a tension between faithfulness to the compound itself and faithfulness to the N2, the grammar instead left it to the paradigm uniformity effect alone to perturb the otherwise at-chance distribution of variation (at chance because the weight of Id-[nasal]-Compound was at zero, indicating, all else equal, that the alternating and non-alternating candidates were equiprobable). This is qualitatively the same finding as for the novel compounds.

We compared the model fit to the corpus data to one where the grammar was forced to assign the same weight to Id-[nasal]-Compound and Id-[nasal]-N2, and found that it was significantly out-performed by the model that allowed the grammar to allot differing weights to different faithfulness constraints to different bases (Δlog-likelihood $= 45.3$, $p < .001$ with one degree of freedom). We take this as tentative evidence that there is an empirical basis in the lexicon for assigning different degrees of faithfulness to different bases.

6.2 On the role of markedness

We began our discussion of voiced velar nasalization by reviewing various sources that have assumed that the alternation observed in monomorphemes is a byproduct of a word-level markedness constraint banning word-medial /g/. This is a typologically-common scenario, and is built quite deeply into the foundations of constraint-based models (cf. Prince and Smolensky (1993), and the more recent summary in Chong (2017)). Separating marked structures from their repair makes it possible to derive both alternations and phonotactic restrictions from a common source. This, in turn, helps resolve "duplication problem" (Kenstowicz and Kisseberth, 1977).

However, the weight of evidence drawn from BKK's data to this point suggest that rather than being driven by markedness, the VVN may instead by driven by competing faithfulness pressures. Evidence comes from the zero weight assigned to the markedness constraint *Internal-[g] in the model fit to the existing data in section 5, as well as the zero weight assigned to the same constraint when fitting the data from the corpus, and also when trying to model the novel N2 data directly. In both these scenarios, however, faithfulness constraints to both /g/-ful and /ŋ/-ful allomorphs received nonzero weight, allowing the data to nevertheless be accounted for. Only in a model that assumes no scaling of faithfulness constraints by resting activation does markedness get weight, underscoring the importance of jointly modeling usage-based and grammatical influences on probabilistic phonology (see 6.3 directly below).[12]

Further, though more indirect, evidence that the weight of the markedness constraint may be in decline comes from the general pattern of change in many Japanese dialects, including the spoken style of the Tokyo dialect, which has lost the allophony altogether in favor of retaining /g/ as [g] in all contexts. This fact does not bear directly on the actual formal model we propose, but it suggests that something in the learning data––be it phonetic, phonological, or otherwise––is contributing to the loss of the allophony and the markedness constraint behind it, that is common to many dialects which is precipitating the loss of the allophony and its driving markedness constraint.

Although this type of faithfulness-driven alternation is unexpected based on the literature reviewed in section 2, nevertheless the Voting Bases model predicts these outcomes should occur, as shown in Figure 8.

6.3 Competence, performance, and formal modeling

This paper has proposed a model of Japanese nasalization that integrates token frequency into the workings of the phonological grammar. Since the prospect of integrating a putatively performance-related factor like token frequency into a formal phonological model is not an uncontroversial one, below we directly address some possible criticisms of this approach. We certainly do not think that these are the last words on the topic, but we do feel that by explicitly discussing what we are doing and our motivations for doing it, we take a first step towards a clearer understanding of the stakes and consequences of the choices made in modeling information about usage jointly with the phonological grammar.

One initial objection to formally modeling the frequency-conditioned variation in nasalization might be that there is nothing competence-related to model here at all—the variation is solely driven by "performance" factors (Chomsky, 1965). We respond that this cannot be true of Japanese nasalization: the fact that only compounds whose N2 is morphologically free exhibit frequency-sensitive variation, despite the existence of bound morphemes with [g]- and [ŋ]-initial forms like [ga]/[ŋa] "fang", as shown by the examples in (5), requires an explanation that makes reference to grammatical structures.

Further afield, cases like Lexical Conservatism much more strongly blur the line between the contents of the lexicon and the phonological grammar and are well-modeled by a framework like Voting Bases. The fact that this paper demonstrates both paradigm uniformity and Lexical Conservatism emerge as special cases of the same theory speaks to the theoretical insight that can be gained by jointly modeling "performance-related" and "competence-related" influences on

[12] In such a model, there is also weight given to a faithfulness constraint indexed to a /g/-ful UR for the compound, following the intuition of a reviewer.

the phonological grammar.

Another objection is that by incorporating both resting activation (a psycholinguistic construct) and phonological markedness (a grammatical one), the model blurs the line between competence and performance, raising the question of what exactly the model is modeling. If so, this would be a legitimate concern. However, a virtue of the Voting Bases model is that lexical influence on the grammar is clearly delimited: the model only allows the lexicon to scale the weights of faithfulness constraints to corresponding lexical representations. Manipulating the resting activation of a given UR has identifiable, localized influences on the computations of the phonological grammar, and instantiates a linking hypothesis consistent with a consensus view of the basic structure of the lexicon. This mechanism can be seen as one way of implementing the idea of "grammar dominance" put forth, for example, by Coetzee (2016) and Coetzee and Kawahara (2013). The "core" phonological grammar—weighted constraints which can assess violations of novel candidates—can be recovered by simply ignoring the influence of the lexicon on constraint violations, and can be studied in novel contexts like *wug*-tests, where there is no relevant lexical representation to bear on the grammar.

A final objection that we consider is that the very act of jointly modeling usage frequency and the phonological grammar risks leading the analyst to think of fundamentally performance-related factors as in fact competence-related, thus undercutting the goal of researchers whose focus is only understanding linguistic competence. We contend that this is simply false, and in fact, the reverse is true: for a researcher who *only* cares about linguistic competence, modeling usage factors jointly with theories of competence is vital. When confronting data derived from language use (that is, modeling corpus data as in Breiss et al. (2021a), or experimental data where stimuli are existing morphemes of the language as in Breiss et al. (*to appear*)), a joint model will better expose the true influence of competence-related factors on the data under study, with the performance-related parts of the model accounting for the otherwise-distorting influence of these factors. Simply ignoring performance-related factors in a formal model makes the strong claim that they have no effect, an assumption which is untenable in the cases examined here, and, we suggest, is also false in many (if not all) types of linguistic data that speakers might have prior usage-based experience with (Arnon and Snider, 2010; Smith and Moore-Cantwell, 2017; Zymet, 2018; Morgan and Levy, 2016, 2023). Rather, an integrated approach that jointly models grammar and usage is essential to disentangle and distill and understanding of competence from its entanglement with performance factors, if this is the goal of the analysis.

The foregoing discussion, as well as comments from reviewers, raise the question of whether the analysis proposed here still cleaves to the generative roots of the constraint-based model formalism that it adopts (though cf. Smolensky (1986a); Legendre et al. (1990); Smolensky (1986b) on the shared roots of Optimality Theory, Harmonic Grammar, MaxEnt, and connectionism (Rumelhart et al., 1986)). This, in our opinion, is somewhat a matter of perspective, and is in any case rather beside the point. Depending on how one defines “generative” or “functionalist”, our model may be seen as aligned with either point of view — since it, too, aims to model grammar, its use, and acquisition at a certain necessary level of abstraction. What we hope this exercise demonstrates, rather, is that by reifying our theories about what the data-generating process is in a computational model, we can confront complex data with many interlocking or moving parts, and recover transportable analytical insights that we are confident are common desiderata shared by many strands of linguistic analysis. We also note that we are far from the first to pursue this approach — for very closely-related discussions of what it means for a linguistic theory to model frequency, see Coetzee and Kawahara (2013); Coetzee (2016), among others.

6.4 Comparison with other models

The Voting Bases model is one of several approaches in the literature that propose to model the interaction of usage frequency and phonological grammar. In particular, it is similar to the methods proposed in Coetzee (2016) and Coetzee and Kawahara (2013) which directly scale the weight of faithfulness constraints by the frequency of the form they make reference to, and that of Baird (2021) where a simulated perception-production loop comes to the same result via online learning. This family of approaches involves lowering the weight of faithfulness constraints to high-frequency forms relative to lower-frequency forms which enables them to model data like coronal stop deletion in English (Coetzee and Kawahara, 2013), where higher-frequency monomorphemes (like *just*) tend to get produced more often with a deleted coronal stop than phonologically-similar words (like *jest*). Common to these models is that they assume that the underlying form is /t/-ful, and thus the task of their model must relate higher frequency to therefore have lower constraint weights for it.

A weakness of these models is that, with the possible exception of Baird (2021), the directionality between frequency and constraint weight is arbitrary—the primary goal set in these studies was to fit the data, which is better than the alternative which does not model the effects of lexical frequencies at all, but they suffered somewhat for the lack of clear functional grounding the relation.

By contrast, the frequency-faithfulness relation that Voting Bases model adopts runs in the opposite direction—more frequent forms exact a greater penalty for unfaithful realizations relative to less frequent forms; constraint violations are less severe for low-frequency vs. high-frequency forms. This allows the model to fit a similar range of data, but with a linking hypothesis that is explicitly rooted in resting activation, a construct that is externally justified by a large body of work in psycholinguistics, as reviewed in Breiss (2021, 2024). Lexical items with higher resting activation are more insistent on faithfulness to themselves, corresponding to their increased salience in the language processing system. The main contribution of the Voting Bases model in modeling this phenomenon is that the influence of the lexicon on the grammar should be, in principle, derivable without reference to any facts about the experiment in question; given some independently-established computationally-implemented model of lexical dynamics that represents a scalar quantity of resting activation (or similar construct), the strong prediction of the Voting Bases model is that that quantity should be able to be a fully adequate scaling factor for faithfulness constraint violations. The specific mechanism that is used in this paper—scaling the weights by the sigmoidal transformation of the resting activation—is used since it represents, to us, a reasonable first stab, but the linking function may need to be revised in light of future findings.

In summary, we suggest that the Voting Bases model, because of its functional grounding of frequency effects in externally-motivated psycholinguistic phenomena, is on firmer footing than theories that have alternative linking functions between frequency and grammar, which are arguably arbitrary.

6.5 Towards a unified picture of token frequency in phonology

In this section, we broaden our view of token frequency effects in phonology, and discuss how considering the varying functional roles of frequency can reconcile some seemingly-contradictory bodies of evidence (cf. also Bybee, 2003).

First, there is evidence that higher token-frequency leads to more markedness-reducing alternations. Coetzee and Kawahara (2013) found that higher-frequency lexical items were more likely to undergo phonological processes of simplification and (markedness-)reduction: high-frequency English words like *jus(t)* underwent an optional process of coronal stop deletion at a higher rate than low-frequency words like *jes(t)*, and high-frequency Japanese words like [baggu] "bag" underwent geminate devoicing more often than low-frequency words like [budda] "Buddha" (Kawahara and Sano, 2013). Zuraw (2007) examines frequency-conditioned application of markedness-reducing phonological processes in a corpus of written Tagalog, and likewise finds higher rates of repair within higher-frequency units (words, clitic groups, etc), subject to the markedness principles of the language.

On the other hand, there is also evidence to show that higher-frequency forms are more likely to be exceptional, and thus marked with regard to the overall properties of the grammar. Smith and Moore-Cantwell (2017) found that higher-frequency comparative constructions are more likely to flout grammar-wide trends driven by markedness. In a similar vein, Anttila (2006) and Mayer (2021) found that higher-frequency morphologically-complex forms were more likely to behave opaquely with respect to grammar-wide phonological processes.

We can compare these effects to the ones observed in Breiss et al. (2021b) and Breiss et al.

(*to appear*): higher-frequency N2s act as stronger attractors, yielding *more* faithfulness to their preserved surface [g] resulting in lower rates of nasalization, whereas higher compound frequency as a whole yielded higher rates of nasalization. Thus it seems that for compounds, higher frequency is correlated with more phonological-process application and markedness-reduction; this is broadly in line with the findings of Coetzee and Kawahara (2013) where higher-frequency words undergo more phonological alternations. However, we found that at the same time, in compounds with free N2s, higher free N2 frequency is related to *less* process application, with higher-frequency supporting the retention of a marked structure (word-medial [g]).

We suggest we can resolve this tension by distinguishing between the processes that token frequency can impact: one is whether to set up an independent lexical representation for a surface allomorph, and the other is influencing the strength of that representation in the lexicon of the speaker.

If a form is exceptional and high-frequency, it may be more economical for a speaker to pay a one-time "cost" of encoding the exception as a listed form that is not derived by the grammar, thus relieving the phonology of the difficulty of having to generate the exceptional or idiosyncratic form on each of the many frequent occasions of use (cf. Adaptor Grammars (Johnson et al., 2007, *et seq.*) or Fragment Grammars (O'Donnell, 2015) which offer computationally-explicit implementations of this general idea). For lower-frequency exceptional forms, the likelihood of listing is less since the price trades off less favorably with the amount of times it is used; thus lower-frequency forms are more susceptible to change and regularization to the dominant grammatical trends over time compared to higher-frequency forms.

Another aspect of this trade-off is the emergence of Lexicon Optimization (Prince and Smolensky, 1993; Sanders, 2003, 2006); even if a form is not particularly exceptionful, if a UR almost always surfaces with a phonological process applied to it, with sufficient frequency it becomes less costly to just store the form with phonological process applied—that is, create a separate allomorph that is specific to the environment that would trigger the phonological rule. This, similarly, relieves the grammar of the job of having to repair the form every time. Thus, we find Lexicon Optimization targeting forms like *jus(t)* over forms like *jest*, making these forms restructured to automatically have the phonological alternation applied, thus giving the appearance of having undergone a markedness-improving repair in the grammar, but actually the frequency of the form has resulted in restructuring to the lexicon (see Breiss and Wilson (2020) for an initial attempt at a computational model of the phonological grammar and lexicon that exhibits this property).

As reviewed above, lexical frequency also influences the resting activation of a lexical item once it is listed in the lexicon. In the Voting Bases model, higher resting activation leads to the listed form exerting a stronger pull on the surface realization of a related form; where this pressure goes against the broader principle of markedness in the grammar, as in cases of paradigm uniformity, we find that marked structures with high-frequency output-bases are preserved; in cases where the listed form coincides with the output of the markedness-reducing process, as in many cases of Lexical Conservatism (Steriade, 1997; Steriade and Stanton, 2020; Breiss, 2021), then the higher-frequency form promotes an unmarked surface form.

Recent work by Jarosz et al. (2024) has laid out a class of models which exhibit characteristics that align favorably with the dynamics of frequency laid out here, suggesting that an integrated, implemented model that can jointly account for the variety of frequency effects reviewed in this section is perhaps quite close at hand. Future work may profitably explore how well these models can provide converging evidence from computational learning simulations to support the psycholinguistic, experimental, and diachronic evidence for the contents of the lexicon that the Voting Bases theory relies on. In sum, the broader landscape of token frequency in phonology is compatible with the functional grounding given to frequency under the Voting Base model, though much empirical and formal work remains to be done to further support the predictions of the framework more broadly as a candidate for a general theory of the influence of the dynamic lexicon on the probabilistic grammar.

References

Albright, A. (2008). Cumulative violations and complexity thresholds. Unpublished ms, MIT.

Anttila, A. (2006). Variation and opacity. *Natural Language & Linguistic Theory*, 24(4):893–944.

Arnon, I. and Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, 62(1):67–82.

Baird, M. (2021). Deriving frequency effects from biases in learning. *Proceedings of the Linguistic Society of America*, 6(1):514–525.

Benua, L. (2000). *Phonological relations between words*. Psychology Press.

Boersma, P. and Pater, J. (2016). Convergence properties of a gradual learning algorithm for harmonic grammar.

Breiss, C. (2020). Constraint cumulativity in phonotactics: evidence from artificial grammar learning studies. *Phonology*, 37(4):551–576.

Breiss, C. (2024). When bases compete: A voting model of lexical conservatism. *Language*, 100(2):308–358.

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., and Kawahara, S. (accepted). Token frequency modulates optional paradigm uniformity in japanese voiced velar nasalization. *Phonology*.

Breiss, C. and Wilson, C. (2020). Extending adaptor grammars to learn phonological alternations. *Proceedings of the Society for Computation in Linguistics*, 3(1):480–483.

Breiss, C. M. (2021). *Lexical Conservatism in phonology: Theory, experiments, and computational modeling*. PhD thesis, University of California, Los Angeles.

Bybee, J. (2003). *Phonology and language use*, volume 94. Cambridge University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* The MIT Press.

Chong, J. A. (2017). *On the relation between phonotactic learning and alternation learning*. PhD thesis, UCLA.

Coetzee, A. W. (2016). A comprehensive model of phonological variation: Grammatical and non-grammatical factors in variable nasal place assimilation. *Phonology*, 33(2):211–246.

Coetzee, A. W. and Kawahara, S. (2013). Frequency biases in phonological variation. *Natural Language & Linguistic Theory*, 31(1):47–89.

Fylstra, D., Lasdon, L., Watson, J., and Waren, A. (1998). Design and use of the microsoft excel solver. *Interfaces*, 28(5):29–55.

Goldwater, S. and Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*, volume 111120.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, 39(6):1041–1070.

Hayes, B. (2022). Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics*, 8:473–494.

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

Hibiya, J. (1995). The velar nasal in Tokyo Japanese: A case of diffusion from above. *Language Variation and Change*, 7:139–152.

Ito, J. and Mester, A. (1996). *Correspondence and compositionality: The ga-gyō variation in Japanese phonology.* Clarendon Press Oxford.

Ito, J. and Mester, A. (2003). *Japanese morphophonemics: Markedness and word structure*, volume 41. MIT Press.

Jäger, G. and Rosenbach, A. (2006). The winner takes it all—almost: Cumulativity in grammatical variation. *Linguistics*, 44(5):937–971.

Jarosz, G., Hughes, C., Lamont, A., Prickett, B., Baird, M., Kim, S., and Nelson, M. (2024). Type and token frequency jointly drive learning of morphology. *ms.*

Jiang, H. (2023). Learnability of length-referencing alternations. MA Thesis, University College London.

Johnson, M., Griffiths, T. L., Goldwater, S., et al. (2007). Adaptor grammars: A framework for specifying compositional nonparametric bayesian models. *Advances in neural information processing systems*, 19:641.

Jurafsky, D. and Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition.* Prentice Hall.

Kawahara, S. and Sano, S.-i. (2013). A corpus-based study of geminate devoicing in japanese: Linguistic factors. *Language Sciences*, 40:300–307.

Kenstowicz, M. and Kisseberth, C. (1977). *Topics in phonological theory.* Academic Proess.

Kindaichi, H. (1942). Ga-gyo bionron. *Onin no Kenkyû (Reprinted in Kindaichi 1967).*

Labrune, L. (2012). *The phonology of Japanese.* Oxford University Press.

Legendre, G., Miyata, Y., and Smolensky, P. (1990). *Harmonic grammar: A formal multi-level* *connectionist theory of linguistic well-formedness: Theoretical foundations.* Citeseer.

Luce, P. A. and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, 19(1):1–36.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., and Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60):3139.

Martin, A. T. (2007). *The evolving lexicon.* PhD thesis, UCLA.

Mayer, C. (2021). *Issues in Uyghur backness harmony: Corpus, experimental, and computational studies.* University of California, Los Angeles.

Morgan, E. and Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, 157:384–402.

Morgan, E. and Levy, R. P. (2023). Productive knowledge and item-specific knowledge trade off as a function of frequency in multiword expression processing.

Morton, J. (1970). A functional model for memory. *Models of human memory*, pages 203–254.

NHK (1993). Nhk broadcasting culture research institute.

O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage.* MIT Press.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2:33–52.

Prince, A. and Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. *Optimality Theory in phonology.*

Purse, R., Fruehwald, J., and Tamminga, M. (2022). Frequency and morphological complexity in variation. *Glossa: a journal of general linguistics*, 7(1).

Rebrus, P. and Törkenczy, M. (2017). Gradient harmonicity in compounds. In *Proceedings of the Annual Meetings on Phonology*, volume 4.

Rumelhart, D. E., Hinton, G. E., McClelland, J. L., et al. (1986). A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(45-76):26.

Sanders, N. (2006). Strong lexicon optimization. *University of Massachusetts Amherst Phonology Group. http://sanders. phonologist. org/Papers/sanders-umass. pdf.*

Sanders, R. N. (2003). *Opacity and sound change in the Polish lexicon.* University of California, Santa Cruz.

Shih, S. S. (2017). Constraint conjunction in weighted probabilistic grammar. *Phonology*, 34(2):243–268.

Smith, B. W. and Moore-Cantwell, C. (2017). Emergent idiosyncrasy in English comparatives. In *NELS 47: Proceedings of the 47th meeting of the North East Linguistic Society. Amherst: Graduate Linguistic Student Association.*

Smolensky, P. (1986a). Information processing in dynamical systems: Foundations of harmony theory. Technical report, Colorado Univ at Boulder Dept of Computer Science.

Smolensky, P. (1986b). Neural and conceptual interpretation of pdp models. *Parallel distributed processing: Explorations in the microstructure of cognition*, 2:390–431.

Steriade, D. (1997). Lexical conservatism. *Linguistics in the morning calm*, pages 157–179.

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. *Papers in laboratory phonology*, 5:313–334.

Steriade, D. and Stanton, J. (2020). Productive pseudo-cyclicity and its significance. Talk at

LabPhon 17.

Trubetskoy, N. (1969). Principles of phonology [translated from the Russian by C. Baltaxe]. *Berkeley: U. of California Press*.

Wang, Y. and Hayes, B. (2025). Learning phonological underlying representations: the role of abstractness. *Linguistic Inquiry*, pages 1–44.

White, J. (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-Map bias. *Language*, 93(1):1–36.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science*, 30(5):945–982.

Wilson, C. and Obdeyn, M. (2009). Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Unpublished manuscript, Johns Hopkins University.

Zuraw, K. (2007). Frequency influences on rule application within and across words. In *Proceedings from the annual meeting of the chicago linguistic society*, volume 43, pages 283–309. Chicago Linguistic Society.

Zymet, J. (2018). *Lexical propensities in phonology: corpus and experimental evidence, grammar, and learning*. PhD thesis, UCLA.

Zymet, J. (2019). Learning a frequency-matching grammar together with lexical idiosyncrasy: Maxent versus hierarchical regression. In *Proceedings of the Annual Meetings on Phonology*, volume 7.