Modeling the effect of token frequency in the phonological grammar: A case study in Japanese voiced velar nasalization*

Canaan Breiss, MIT†
Hironori Katsuda, UCLA
Shigeto Kawahara, Keio University

November 3, 2022

Abstract

This paper analyzes the role of token frequency in conditioning optional paradigm uniformity effects, focusing on Japanese voiced velar nasalization. We report a *wug*-test demonstrating the frequency-conditioning observed by Breiss et al. (2021b) in corpus data is reproduced in existing compounds and novel compounds. We propose a psycholinguistically-informed formal phonological model that allows the token frequency of surface forms to scale the severity of faithfulness violations that reference them. We extend the mechanisms of the Voting theory of Base competition, which was originally proposed in Breiss (2021) to account for the interplay of lexical and phonological forces in shaping probabilistic patterns of Lexical Conservatism. We interpret the success of Breiss' formalism in modeling output-oriented phonological processes beyond its original scope as support for the mechanisms proposed in the model, and more generally for the value of jointly modeling the influence of lexical and phonological properties on phonological process application.

*This work was supported in part by NSF Graduate Research Fellowship DGE-1650604 to Canaan Breiss and JSPS grant #22K00559 to Shigeto Kawahara. Thanks to audiences at UCLA, AMP 2021, WCCFL 39, and the 45th Penn Linguistics Conference. The experiment was carried out under the ethical approval granted by the second author's university. Supplementary material can be accessed at https://osf.io/avnpw/?view_only=cd2afdcc183f4de3ac1261b4af66f08d.

†Contact: canaanbreiss1@gmail.com

1 Introduction

The influence of lexical frequency on phonological patterning has been much debated in the literature. In the classic generative tradition, frequencies—or statistical information in general—are considered to lie outside of grammatical competence, as argued for example in Chomsky's *Syntactic Structures* (Chomsky, 1957). Much of the phonological work in the SPE era (Chomsky and Halle, 1968) did not seriously consider the effects of lexical frequency on phonological patterns. On the other hand, there was a recurrent observation that for example, deletion of a phonological segment is more likely in frequent words than in non-frequent words (Bybee, 1999). For instance, [t/d]-deletion in English seems very common in frequent words like *and* and *list*, but not as very common in less frequent words like *mast* or *jest*. In fact, in usage-based phonology (Bybee, 1999) as well as exemplar-theoretic phonology (Gahl and Yu, 2006), usage frequency is directly encoded in grammatical model. Coetzee and Kawahara (2013) argue that it is possible and necessary to incorporate the effects of lexical frequency in the formal, generative phonological grammar. Whether, and to what extent, lexical frequencies affect phonological patterns, and which aspects of phonological patterns are (un)affected, all remain important questions in the current phonological research.

Against this general theoretical background, this paper analyzes the role of token frequency in conditioning optional paradigm uniformity effects, focusing on the case of Japanese voiced velar nasalization (henceforth, simply "nasalization"). Building on the results of a corpus study presented in Breiss et al. (2021b), we report a *wug*-test with speakers of the phonologically-conservative Tōhoku dialect of Japanese (spoken in northern parts of the main island of Japan), which demonstrates that the frequency-conditioning observed in corpus data is reproduced in existing and novel compounds. This empirical finding has descriptive value in that it demonstrates a role for lexical frequency in influencing the phonological patterns shown by each individual speaker in the context of paradigm uniformity.

In addition, we propose a psycholinguistically-informed reframing of the relationship between the lexicon and the grammar which allows token frequency of surface forms of lexical items to scale the severity of faithfulness violations that reference them. The theory is implemented in a Maximum Entropy Harmonic Grammar (henceforth, MaxEnt grammar) (Smolensky, 1986; Goldwater and Johnson, 2003; Hayes and Wilson, 2008), and extends the mechanisms of the Voting theory of Base competition proposed in Breiss (2021) to account for the interplay of lexical and grammatical forces in experimental data on lexical conservatism (Steriade, 1997). We interpret the success of Breiss' formalism in modeling output-oriented phonological processes beyond its original scope as support for the mechanisms proposed in the model, and more generally for the explicit integration and joint modelling of the influence of lexical and phonological properties on phonological process application.

2 Background on Voiced Velar Nasalization

In many phonologically conservative dialects of Japanese, [ŋ] and [g] are allophonically distributed; here we summarize the generalizations in the literature about the Yamanote dialect, a classic and conservative speaking style of the dialect spoken in the center of Tokyo (see Hibiya 1995 for more on the sociolinguistic significance of nasalization). In dialects that exhibit nasalization, /g/ is realized as [ŋ] in prosodic-word-medial position; e.g. [kaŋami] "mirror" vs. [gimu] "obligation".

This complementary distribution has been discussed extensively in the generative and pre-generative literature on Japanese phonology (e.g. Kindaichi 1942; Trubetskoy 1969; Labrune 2012). Although properly a static phonotactic restriction, the prominence of compounding in Japanese word-formation means that there are many contexts where the same morpheme can surface free-standing with initial [g], as well as a second member of a compound (=N2) with initial [ŋ]. Thus there is ample opportunity to study status of the phonotactic restriction in the synchronic grammar via the alternation it induces, as in (1)-(3). It is in this context that Ito & Mester treat the phenomenon, first in Ito and Mester (1996) and later in Ito and Mester (2003), where they formalize a constraint-based analysis of the alternation observed in compounds. Most relevant for the current paper, they highlight the optionality of the alternation in cases where the second member of the compound is also a free-standing word, as illustrated by the examples in (1)-(3).

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

b. /gan/→ [gan]
cancer
"cancer"

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

b. /geka/→ [geka]
surgery
"surgery"

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

b. /ga/ → [ga], "moth"

The gist of their analysis is that the optionality is the result of two competing forces acting on the realization of the /g/-initial word that occurs in a compound as N2 (second nouns): (1) a paradigm uniformity effect to its base form (Steriade, 2000), which prefers [g] to [ŋ], and (2) a markedness constraint that favors nasalization in intervocalic positions, favoring [ŋ] to [g]. This analysis captures both the variability of compounds with free N2s, and also the obligatoriness of nasalization when N2 is a bound morpheme, as in cases like (4).

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

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

Recently, Breiss et al. (2021b) carried out a quantitative investigation of the claims of variability and optionality surrounding nasalization. Using data from the 2016 NHK Pronunciation and Accent dictionary (NHK, 1993) which represents a consensus view of expert dialectologists about normative pronunciation in the Yamanote dialect, they found that the claim of optionality in cases of a free-standing N2 was robustly borne out. When N2 is a bound morpheme, all compounds in the dictionary underwent nasalization. By contrast, when N2 was free, there was a large deal of variation. This result offers quantitative evidence for the patterns that were analyzed by Ito & Mester, reviewed above. Further, Breiss et al. (2021b) found the frequency of the whole compound and its second member to be a reliable predictor of whether a given compound would actually undergo nasalization or not.[1] However, the corpus analysis left open a number of questions about the nature of the frequency effect, that we address here. Resolving these questions is of critical importance to how we construct our phonological theories, including whether (and if so, how) usage frequency should be integrated into models of the synchronic grammar.

3 The current study

The current paper has three empirical goals. The first is to examine whether the optionality of the paradigm uniformity in existing compounds found in the corpus is operative at the level of the individual speaker. This is an important methodological point that cannot be resolved in corpus-based studies of variation, including Breiss et al. (2021b), because it is possible that apparent variation in a corpus actually results from collapsing across different speakers with

[1] The frequency measures were taken from the Balanced Corpus of Contemporary Written Japanese (Maekawa et al., 2014).

different categorical grammars. The second goal is to see whether the frequency-conditioning of this variability is also active at the speaker level. The third goal is to see whether the frequency effect extends to entirely novel compounds, or whether it is limited to forms with which the speakers might plausibly have experience.

The status of novel compounds is of great relevance for distinguishing between phonological theories that differ in their treatment of frequency effects: The UseListed theory of Zuraw (2000) holds that the effects of frequency in phonology, such as those discussed in section 1, can be explained as competition between two routes of processing—whole-word retrieval, or in-the-moment grammatical assembly. The data from existing compounds are compatible with this architecture but also one where the working of the phonological grammar itself is influenced by frequency. Novel compounds are an important testing ground for theories that put frequency in the lexicon, and finding an effect of N2 frequency in modulating nasalization in these forms would be a strong indicator that the phonological grammar itself is sensitive to the lexical frequency of the items in manipulates, as Coetzee and Kawahara (2013) and Coetzee (2016) have suggested.

3.1 Methods

3.1.1 Participants

The second and third authors recruited 20 speakers of the Tōhoku dialect of Japanese by word of mouth and snowball sampling to participate in the experiment. We chose to examine the Tōhoku dialect because most (if not all) of the speakers of the Yamanote dialect documented in the literature and reflected in the NHK Pronunciation and Accent dictionary are no longer living, or were judged unlikely to be able to participate in an online experiment. The Tōhoku dialect spoken in northern parts of the main land of Japan is also phonologically conservative, and has been documented as also exhibiting the the voiced velar nasalization alteration.

All participants completed a short dialect questionnaire, which used existing monomorphemes to determine whether the speaker enforced the complementary distribution of [g] and [ŋ]—the phonotactic which drives the alternation in compounds. If the speaker did not, they were not invited to continue to the experimental task. Of the 20 interviewed, eight passed the dialect questionnaire (six females; mean age 53; four were from Iwate prefecture, two from Fukushima prefecture, one from Yamagata prefecture, and one from Aomori prefecture).[2] Because it was difficult to find speakers who exhibited the appropriate pattern of allophony, all but one person participated in the experiment in two separate rounds, each separated by a period of a few weeks to several months. The two different rounds counterbalanced stimulus randomization orders, and

[2]Many speakers, particularly young ones, have lost this pattern of nasalization due to the influence of the "standard" Tokyo dialect, which has also lost the nasalization pattern.

also which N2s were primed (on which, see section 3.1.3 immediately below). Participants were paid approximately 20 USD per experimental session.

3.1.2 Stimuli

Stimuli for the experiment were 301 compounds, 81 of which were existing, and drawn from the corpus used in Breiss et al. (2021b), and the rest of which were novel. The existing compounds were selected to represent a range of nasalization probabilities, based on the relative frequency of the N2. Because of this, existing compounds varied somewhat in length (from 2 to 8 mora in total). Further, some of the N1s and N2s were monomorphemic (e.g., /kai + gjoo/ "indention") while others were multimorphemic (e.g., /kagaku/ + /giʒutsu/ 'science and technology'; both N1 and N2 are bimorphemic, i.e., /ka + gaku/ "science", /gi + ʒutsu) "technology"). Existing compounds were all formed with /g/-initial N2s. The novel compounds were formed by combining 30 bimoraic N2s (e.g., /gin/ "silver", /gjaku/ "reverse"), with six bimoraic N1s (e.g., /d͡ʒuu/ "heavy, multiple", /tei/ "low"), with the N2s selected to be of varying frequencies. Due to the difficulty of finding enough number of monomorphimic N2s, 10 out of the 30 N2s were bimorphemic (e.g., /ge + ta/ "wooden clogs", /go + ma/ "sesame"). The study also included 40 novel compounds whose N2 was /k/-initial, in order to examine the synchronic status of the opaque interaction of nasalization with Rendaku (cf. the extensive discussion in Ito and Mester 1996, 2003). These results are not reported here, and are thus not discussed further; the data reported for novel compounds in this paper is the result of 180 distinct novel compounds with a /g/-initial N2.

3.1.3 Procedure

The format of the experiment was an elicited production task; participants were presented with a series of forms via PowerPoint presentation, and were asked to produce them aloud. All of the experimental sessions were carried out over Zoom by the second author, and were recorded for posterity.

Each recording session proceeded in the following way: first, participants were given the dialect questionnaire; if they passed, they proceeded to the main task. In the main task, participants completed a preliminary vocabulary familiarity survey before producing compound forms, and a post-hoc vocabulary familiarity survey after producing all the compounds. In each vocabulary survey, participants were asked to produce one of the existing compounds or one of the N2s out loud, and indicate how familiar they were with the word on a 5-point Likert scale (5 = *extremely familiar*, 1 = *I don't know this word*).

We took advantage of the distribution of items between the vocabulary surveys to examine what effect priming an N2 might have on whether it exhibits nasalization when produced in a compound. In his work on Lexical Conservatism, a phenomenon where a non-local output form

of a paradigm influences an optional phonological process elsewhere in that paradigm, Breiss (2021) found that asking participants whether they knew the non-local paradigm member *before* accessing its paradigm-mate in a nonce-formation task increased the probability of the novel production showing the influence of that non-local form, compared to forms for which the non-local paradigm member was asked about after the nonce-formation task. He took this as evidence that this "priming" of the non-local form increased its saliency in the lexicon of the speaker, allowing it a greater influence in subsequent grammatical processes. Here we were interested in the same question, and so performed a similar priming manipulation on our N2s. All of the existing compounds included as stimuli were always in the post-hoc vocabulary survey, while the N2s were distributed around the compound-production task in a way that each participant saw one set of N2s before the compound-production task, and one after. In participants who participated in the experiment twice (all but one of them), the N2s primed were varied between sessions.

In the compound production task itself, which is of primary interest, participants were asked to simply read aloud compounds in a random order that mixed novel and existing forms, and the experimenter noted whether the onset of N2 was produced with nasalization or not. To preserve ambiguity, all compounds were presented in *kanji* characters, which do not distinguish between [g] and [ŋ].

3.2 Statistical analyses

After the data collection was complete, each compound was classified as whether it was known to the speaker (score > 2) or not (score = 1). Then, for each speaker, compounds with unknown N2s were excluded. This allows us to make inferences about the phonological grammar at the level of the speaker, rather than simply assuming that all speakers know all words. All data and scripts are available in the supplementary materials of this paper.

Statistical analyses were carried out using Bayesian mixed effects logistic regression models implemented in the `brms` package (Bürkner, 2017) using the R programming environment (R Core Team, 2021). There are several advantages of Bayesian models as opposed to frequentist (non-Bayesian) ones, which we summarise only briefly here. First, rather than focusing on hypothesis testing, the results of our Bayesian regression models can be interpreted as directly reflecting the distributions of likely values for each parameter. Second, it is known that Bayesian models are more likely to converge than corresponding frequentist linear mixed effects models, the latter of which is especially difficult to achieve convergence with when the model has a complex random structure, i.e., the sort of the model we report below. In a Bayesian model, we formalize our prior knowledge or expectations (if any) about the values of the parameters we are interested in using statistical distributions, and then knit it together with the evidence

from the data, producing a posterior distribution of values for our parameters of interest that are a compromise between our priors and our data. This posterior distribution is then the object which we mine for analytic insights. For tutorial introductions to Bayesian data analysis applied to linguistic and related subject material, see Kruschke (2014), Vasishth et al. (2018); for a primer on the *brms* package specifically in a linguistic context, see Nalborczyk et al. (2019).

In this paper, we report two common metrics of the posterior distribution for model parameters of interest: the median and 95% Highest Density Interval (HDI) which is presented as a bracketed range, and the probability of direction, noted $P\,(|\hat{\beta}| > 0)$. The first measure indicates the most likely value of the parameter of interest, along with an interval reflecting our level of certainty in this value; each value inside this region is strictly more likely than any value outside of it. The second measure can be taken as a way of assessing the amount of evidence we have in favor of any effect in the direction of the parameter coefficient, regardless of magnitude; this ranges from 0.5 (equal evidence for an effect in the direction of the parameter as in the opposite direction) to 1 (very strong evidence in favor of an effect in the direction of the parameter value).

In terms of model structure, each model used as its dependent variable the realization of the initial segment of N2 ([g] or [ŋ]), contained fixed effects specified below, and random intercepts for subject and compound, with random slopes of all fixed effects by speaker and a random slope of priming (primed or not primed) by compound. The models used Normal (0,1) priors on the intercept and coefficients; sensitivity analyses (Roos et al., 2015) revealed no meaningful changes in inference were associated with a range of prior values, indicating the the data we collected were sufficiently informative that our prior beliefs about likely parameter values mattered only nominally; see supplementary materials for details.

We pause here to draw attention to the fact that out of a desire to have enough types of real and novel compounds, spanning a range of frequencies, we were unable to make the existing compounds uniform in size, and neither the existing nor novel compounds uniform in morphemic composition. Because we have no reason to believe these factors to be causally related to the propensity to undergo VVN, and on the basis of the second two authors' native speaker intuition that the bimorphemic compound members were much more salient as whole words, rather than compositionally-formed parts, we do not consider these as theoretical quantities of interest in our statistical or grammatical analysis. We expect the random intercept for compound included in all of the statistical models we fit to absorb any idiosyncrasy attributable to morphemic composition or length to be absorbed by this term, treated as item-level quirks that need to hold from sample to sample, in the same way that idiosyncratic participant-level variation is absorbed by the random intercept for participant. Readers interested in investigating the causal link between nasalization and these other factors for themselves may access the raw data in the supplementary materials.

3.3 Results

In what follows, we first visualize and qualitatively discuss the results of the experiment, then perform parameter estimation using a Bayesian model to conform the statistical reliability of our observations.

3.3.1 Existing compounds

We first examine existing compounds with /g/-initial N2s, and ask whether the token frequency of the compound or the N2 explains variation in nasalization, which is plotted in Figure 1. Note that while Breiss et al. (2021b) considered *relative* frequency of the N2, here we consider the frequency of the N2 and compound in their own right, since we are interested in the competition between characteristics of a morphologically-complex listed form (which requires accessing the compound as a whole) and a grammar-derived form (which requires assembling it online from its parts, including the N2); finally, we use the natural logarithm of the token frequency, rather than its raw value, as is standard practice.

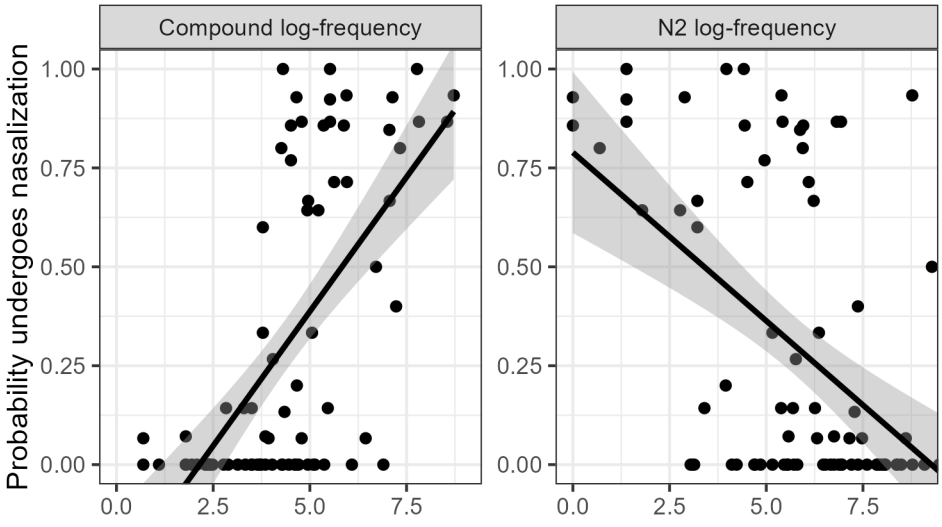

Figure 1: Probability of nasalization (vertical axis) plotted against compound log-frequency (left facet) and N2 log-frequency (right facet). Linear smooths are added as a visual aid only.

We find that the frequency effect is robust, bearing out the spirit of the effect seen in the lexical analyses reported in Breiss et al. (2021b). As N2 frequency rises relative to a fixed value of compound frequency, the probability of an individual compound exhibiting nasalization drops (the right facet); holding N2 frequency steady while increasing the frequency of a compound from low to high also increases the probability of nasalization (the left facet).

Having found that paradigm uniformity is conditioned by frequency in existing compounds at the group level, we now examine whether the conditioning holds at the level of the individual grammar, plotting each participant in their own row, in Figure 2.

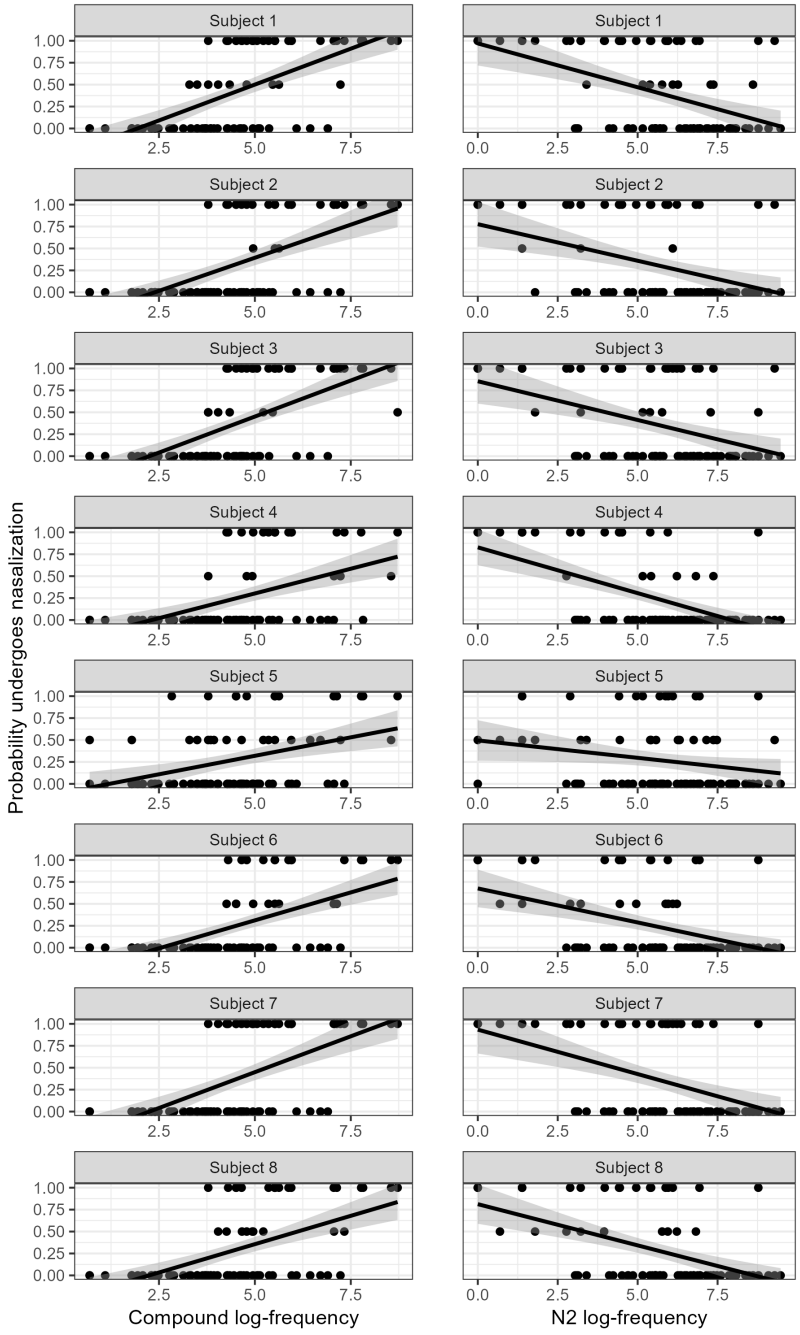

Figure 2: Probability of nasalization (vertical axis) plotted against compound log-frequency (left facet) and N2 log-frequency (right facet) for each individual speaker (row).

On visual inspection, it seems clear that the effect does exist at the individual level, but may differ in strength between speakers. We return to this question with a quantitative eye when we

discuss the statistical model fit to this data.

Before doing so, however, we report two null results in the set of existing compounds that we had expected to find based on the literature: that of priming the N2, and of OCP-[nasal]. We had anticipated that priming the N2 would impact the likelihood of compounds with primed N2s to undergo nasalization, based on the findings and rationale of Breiss (2021), described above in section 3.1.3. However, priming seemed to have no meaningful effect on rates of nasalization (left plot of Figure 3). Thus, we conclude that the experimental manipulation (placement of the N2 in the vocabulary check sequence) failed to influence the salience of the N2 in such a way for it to be experimentally detectable. Though further research is needed to confirm, we suspect that the lack of priming in this study was because the manipulation tried to target too many N2s at once, leading to a lack of concentrated activation on any particular item. This post-hoc account predicts that studies that try to prime relatively fewer items (as was the case with Breiss 2021) should have a greater chance of estimating the effect of priming on the phonological grammar.

Second, we expected, based on the findings of Breiss et al. (2021b) in Japanese compounds and more broadly in Japanese phonology (Kawahara et al., 2006), to find a decrease in nasalization in compounds whose N1s were nasal-final, so as to avoid creating a sequence of two nasals. We find that the such an effect holds in the experimental data, as shown in Figure 3, but only superficially. Although the left bar in the right plot is higher than the right, the uncertainty about this measure is also much larger, likely stemming from the relatively few compounds ($n = 16$) that have nasal-final N1s, and so the statistical model we fit does not suggest that the visual trend is to be trusted. Based on the conflicting evidence in the literature and this paper, we make no strong conclusions about the interaction of OCP-[nasal] and nasalization in Japanese, and await future, more targeted experiments that address this question directly.

Table 1 presents the results of a Bayesian mixed-effects regression model fit to determine the statistical robustness of the data patterns just reviewed. The model structure and random effects was as described in section 3.2, and included fixed effects of the scaled log-frequency of the N2 and compound, whether N2 was primed, the interaction of priming with frequency of N2, and the nasality of the final segment of N1.

We note that, relative to the intercept, a one-unit increase in compound log-frequency strongly increases the log-odds of the compound undergoing nasalization, and a one-unit increase in N2 log-frequency decreases the log-odds of nasalization — we judge this by the fact that the central 95% of the posterior distribution of credible values for the two frequency coefficients exclude zero. For all other fixed effects, the 95% CI does include zero, so we are less confident in attributing a meaningful effect on the data to these factors.

We estimate the speaker-specific parameter value for both compound and N2 frequency by examining samples extracted from the model; these are summarised in Table 2 using the same

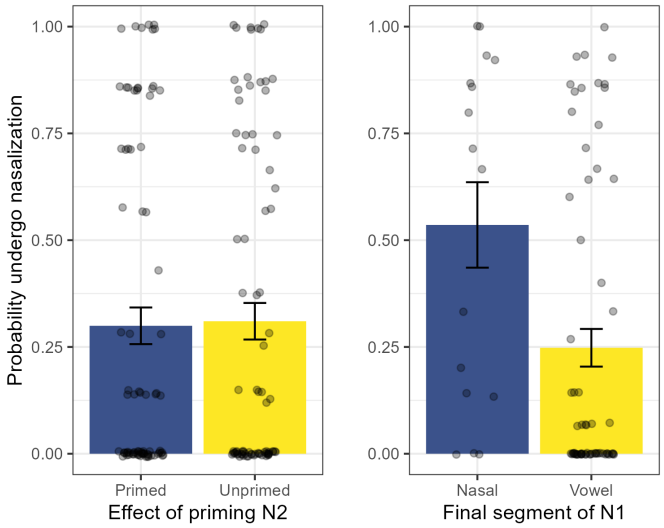

Figure 3: Probability of nasalization (vertical axis) plotted against priming of N2 (left facet) and final segment of N1 (right facet).

metrics as for the model in Table 1. For all participants, an increase in N2 frequency was associated with a decrease in nasalization with greater than 99% probability. For compound frequency, an increase in frequency was associated with a decrease in nasalization with greater than 90% confidence for six of eight speakers; for Subject 4 the effect was less certain (80%), and only Subject 5 truly exhibited no evidence for compound frequency influencing nasalization rate (though the same speaker exhibited a strong influence of N2 frequency).

Based on this evidence, we think it is reasonable to impute the frequency effect in existing compounds to the level of the modal individual grammar, though the factors influencing individual-level variation in effect size remain for future research.

3.3.2 Novel compounds

Turning to novel compounds, we find that the frequency effect holds here as well, though with a smaller magnitude. Figure 4 plots only N2 frequency; since the the compound is entirely novel, its frequency is naturally zero. The downward-sloping trend line qualitatively matches the one found in the right panels of Figures 1 and 2.

Breaking this result out by individual in Figure 5, we find that visually there appears to be a wide range of variation in the strength of the effect across participants, though all but one go in the expected direction. We return to by-subject estimates from a fitted model below.

Finally, consistent with the null effect observed in the existing compounds, we found no strong evidence that priming the N2 influenced nasalization application; this is shown in Figure 6. Since none of the six N1s we selected for constructing the novel compounds were nasal-final,

Parameter	*Median*	*95% CI*	$P(\lvert\hat{\beta}\rvert > 0)$
Intercept:			
N2 final segment = *vowel*			
N2 primed = *no*			
Compound scaled log freq. = *mean*			
N2 scaled log freq. = *mean*	-0.92	[-2.34, 0.53]	
N2 final segment = *nasal*	0.61	[-0.72, 1.94]	0.82
N2 primed = *yes*	-0.30	[-0.31, 1.91]	0.84
Compound log freq. *(one unit increase)*	2.40	[0.84, 3.55]	0.99
N2 log freq. *(one unit increase)*	-1.42	[-2.20, -0.53]	0.99
N2 log freq.× N2 primed =*yes*	-0.35	[-0.91, 0.21]	0.91

Table 1: Model of existing compounds with free N2s. Coefficients are in log-odds, with positive signs indicating an increase in probability of nasalization relative to the intercept.

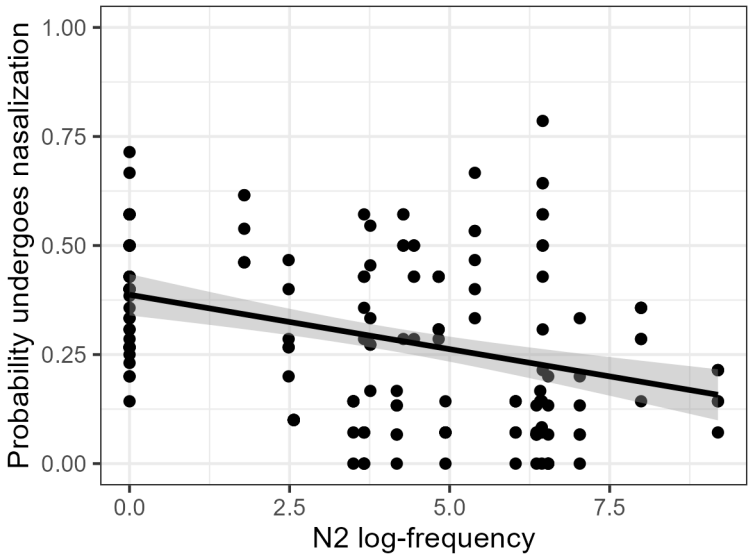

Figure 4: The probability of undergoing nasalization, plotted against N2 log-frequency (novel compounds).

we were not able to evaluate the effect of OCP-[nasal] in this subset of the data.

The results of a Bayesian logistic regression model fit to the data for compounds with novel /g/-initial N2s are reported in Table 3.

Consistent with the existing compounds, we find a strong effect of N2 log frequency, with greater values inhibiting nasalization. None of the other main effects were statistically reliable.

At the individual level, we find strong evidence for a frequency effect of N2 in all individuals; in all but one, the effect is as expected, with higher frequency N2s forming compounds that

	Parameter	Median and 95% CI	$P(\lvert\hat{\beta}\rvert > 0)$
Subject 1	Compound log freq.	3.99 [2.02, 6.04]	≈ 1
	N2 log freq.	-1.35 [-2.65, -0.04]	0.98
Subject 2	Compound log freq.	2.53 [0.78, 4.33]	0.99
	N2 log freq.	-2.56 [-3.93, -1.24]	≈ 1
Subject 3	Compound log freq.	3.57 [1.69, 5.56]	0.99
	N2 log freq.	-1.64 [-2.93, -0.30]	0.99
Subject 4	Compound log freq.	0.74 [-0.96, 2.54]	0.80
	N2 log freq.	-4.78 [-6.77, -3.03]	≈ 1
Subject 5	Compound log freq.	0.04 [-1.40, 1.45]	.52
	N2 log freq.	-1.87 [-3.06, -0.71	.99
Subject 6	Compound log freq.	1.52 [-0.49, 2.85]	.91
	N2 log freq.	-3.99 [-5.53, -2.46]	≈ 1
Subject 7	Compound log freq.	3.77 [1.44, 6.46]	.99
	N2 log freq.	-2.07 [-3.75, -4.94]	.99
Subject 8	Compound log freq.	1.56 [-0.07, 3.13]	.97
	N2 log freq.	-3.22 [-4.67, -1.89]	≈ 1

Table 2: Summaries of individual-level estimates of the effect of the two frequency parameters derived from the model in Table 1.

are less likely to exhibit nasalization. In one individual, Subject 5, however, the effect is in the opposite direction; this is unexpected, and further puzzling because the same subject exhibits a robust frequency effect of N2 in the expected direction in existing compounds (though no strong evidence for an effect of compound frequency in that data, interestingly). We have no explanation for this pattern, other than to note that the effect holds in all other participants; future work is needed to understand the factors that might yield different effects of frequency in different individuals.

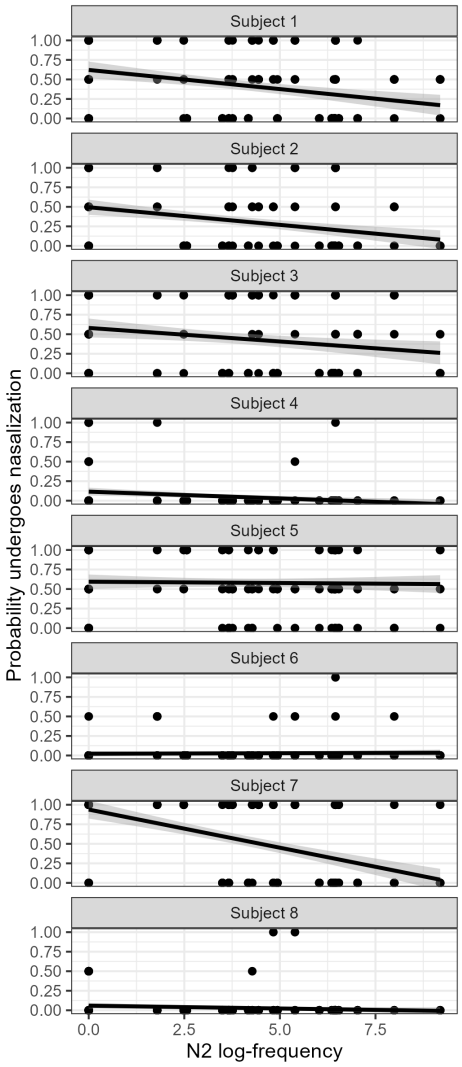

Figure 5: Probability of nasalization (vertical axis) plotted against N2 log-frequency (horizontal axis) for each individual (row).

Parameter	*Median*	*95% CI*	$P(\lvert\hat{\beta}\rvert > 0)$
Intercept:			
N2 primed = *no*			
N2 scaled log freq. = *mean*	-1.92	[-3.68, -1.11]	
N2 primed = *yes*	0.04	[-0.43, 0.53]	0.57
N2 log freq. *(one unit increase)*	-0.50	[-0.96, -0.02]	0.98
N2 log freq.× N2 primed =*yes*	-0.03	[-0.38, 0.33]	0.58

Table 3: Model of novel compounds with free N2s. Coefficients are in log-odds, with positive signs indicating an increase in probability of nasalization relative to the intercept.

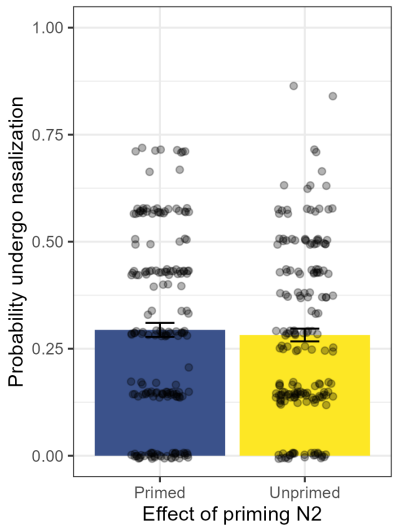

Figure 6: The probability of undergoing nasalization (vertical axis) based on whether the N2 was primed (horizontal axis) in novel compounds.

	Median and 95% CI for effect of N2 log-frequency	Probability of direction
Subject 1	-1.59 [-2.30, -0.93]	≈ 1
Subject 2	-1.98 [-2.65, -1.31]	≈ 1
Subject 3	-0.98 [-1.55, -0.35	0.99
Subject 4	-5.64 [-7.46, -4.16]	≈ 1
Subject 5	0.69 [0.07, 1.36]	0.99
Subject 6	-4.39 [-5.66, -3.27]	≈ 1
Subject 7	-1.73 [-2.72, -0.68]	0.99
Subject 8	-5.48 [-7.05, -4.08]	≈ 1

Table 4: Summaries of individual-level estimates of the effect of the N2 frequency parameter derived from the model in Table 3.

4 Modeling token frequency in the phonological grammar

The *wug*-test just described demonstrates that token frequency is a major determinant of the variation in paradigm uniformity which is licensed by a free-standing N2 base, in both existing and novel compounds. Furthermore, the frequency-conditioning in the variation is robust at the level of the individual, ruling out a scenario where frequency acts on the structure of variation at a population-level, but not at the level of the individual. Therefore, we need 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).

4.1 Evidence for the contents of a lexical entry

Before turning to a grammatical model, it will be important to establish some relevant context about research on 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. Since nasalization concerns paradigm uniformity, we assume the lexical entry for an existing word lists (among many other things) its allomorphs: 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 /ga/ and /ŋa/. Finally, we assume that existing compounds are stored whole, with nasalization applied so as to respect the phonotactic in the lexicon. 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).

With regard to non-phonological characteristics of the lexicon, we follow a large body of findings that representations have differing degrees of salience or strength of encoding, which is often referred to as their *resting activation* (Morton, 1970). Following Breiss (2021), we take resting activation to correspond 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, whether or not they were recently activated (for example, by priming), all contribute dynamically to an item's resting activation. We discuss how resting activation is modeled as influencing the phonological grammar below in section 4.4.

4.2 The Voting model of Breiss (2021)

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), which is couched Optimality Theory (Prince and Smolensky, 1993). The Voting model has been used to model data in English and Spanish Lexical Conservatism (Steriade, 1997), and is broadly compatible with the view of the lexicon laid out above. Here, we extend the model to a different type of output-oriented phonological process — paradigm-uniformity — in a genetically unrelated language, Japanese. The theory has two parts: the first is that all listed forms in the lexicon exert an analogical pull on derivatives (operationalized using faithfulness constraints), violations of which are scaled in proportion to the resting activation of the representation to which faithfulness is being assessed. The second part is that markedness constraints evaluate candidates in the standard way for constraint-based phonological models.

Although in principle the Voting model is not committed to instantiation in a particular constraint-based framework, we follow Breiss (2021) in using the Maximum Entropy Harmonic Grammar formalism (Smolensky, 1986; Goldwater and Johnson, 2003), since it directly relates Harmony to probability (Hayes, 2022), permits constraint cumulativity by default (Breiss, 2020; Jäger and Rosenbach, 2006), has a learning algorithm to set its weights, and it is rooted in well-understood statistical techniques used widely outside linguistics (Jurafsky and Martin, 2009, ch. 5).

4.3 Constraints

In the analysis proposed 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, and a pair of faithfulness constraints, corresponding to the analogical pull of the compound as a whole, if one exists, and to the second member of the compound. They are listed below.

- ***Non-initial-[g]**: incur one violation for each word-internal [g] in a candidate.
- **Id-[nasal]-Remote Base**: incur 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].[3]
- **Id-[nasal]-Local Base**: incur one violation for each segment in the listed allomorph for

[3]We use the term "Remote Base" here to highlight that the structure of the analysis is exactly the same as the one that underpins Breiss' model of Lexical Conservatism, but this constraint simply refers to the free-standing [g]-initial allomorph of N2.

the full compound that does not match its corresponding segment in the candidate 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 theory is an architectural one about how psycholinguistic, "extra-grammatical" factors act within and beside the phonological grammar to influence the 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 take the approach of modelling it as function of log-frequency of the frequency of the allomorph, which is passed through the sigmoid function $\frac{1}{1+e^{-logfreq}}$ that translates the linear predictor into the bounded interval of {0,1}, which will be the scaling factor applied to faithfulness violations. 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, as we did in the statistical analysis above. This corresponds to the notion that it's 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. Finally, we do not model the priming of N2, since it was not found to have a meaningful effect in the data; the question of how to combine multiple continuous and categorical influences on resting activation is left for future research, and discussed briefly in section 6.

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 Maximum Entropy 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. Using the constraints in section 4.3 above, we can define the tableaux below in Figure 7.

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

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

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

*Non-Initial-[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 4.5. For the sake of minimal contrasts, we assume that the frequency of both N2s are equal and medial relative to the examples in Figure 7 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:	*Non-initial-[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:	*Non-initial-[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 8: 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 (Martin, 2007; Albright, 2008) – 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 Remote Base 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 Local Base, and the presence of the Remote Base 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 9.

/.../$_{N1}$, /g.../$_{N2}$, /...ŋ.../$_{High-freq.compound}$ Weight:	*Non-initial-[g] 0	Id-[nas]$_{N2}$ 1	Id-[nas]$_{Compound}$ 2	H	p
i. [...g ...]	1		.7	1.4	.29
j. [...ŋ ...]		.5		0.5	.71
/.../$_{N1}$, /g.../$_{N2}$, /...ŋ.../$_{Low-freq.compound}$ Weight:	*Non-initial-[g] 0	Id-[nas]$_{N2}$ 1	Id-[nas]$_{Compound}$ 2	H	p
k. [...g ...]	1		.3	0.6	.48
l. [...ŋ ...]		.5		0.5	.52

Figure 9: 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.

To preview the findings of the formal model, we find that the faithfulness-driven regime is the one that best captures the nasalization data; this is likely related to the fact that the nasalization alternation is shifting out of the language, as it is not supported by a robust markedness constraint.

4.6 Modeling the experimental data

Moving on to the analysis itself, we first fit models separately to the existing and novel compound data, to demonstrate the suitability of the Voting model in each context, and also to allow better comparison to the statistical models fit to the experimental data above. We then consider how nasalization might be modeled more comprehensively by incorporating information from non-alternating monomorphemes where the complementary distribution between [g] and [ŋ] is enforced, and also by incorporating our knowledge that the free form of N2s surface non-alternatingly with initial [g], despite the presence of an [ŋ]-initial stem allomorph.

In all cases, I fit the Maximum Entropy models using the *Solver()* function in Microsoft Excel (Fylstra et al., 1998), and used a relatively weak 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 fifty 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.

4.6.1 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 take as our data to model the counts of compounds produced having undergone nasalization or not, in cases where speakers know both the compound and the N2 in question.

The best-fitting constraint weights are listed in Table 10, and the fit is plotted in Figure 11. Note that, the weight of IDENT-[nasal]-Local is higher than that of IDENT-[nasal]-Remote, as was the case in the analyses in the Voting model of Base competition framework in Breiss (2021). This aligns intuitively with the notion that although different lexical entries can exert an analogical influence on the output of the grammar, the entry under consideration—here, the compound with a listed form—exerts a stronger influence. Note also that the weight of *NON-INITIAL-[g] in the model is zero; this reflects the analogically-driven nature of the paradigm uniformity, rather than having it be driven by markedness. We return to this point in section 4.6.2 below.

We next compared the Voting model's treatment of frequency to two alternative hypotheses. First, we fit a model where both IDENT constraints were assigned the same weight—this corresponds to a claim that it is *only* different resting activations which lead to variation across compounds, and there is no role for the phonological grammar to play at all in privileging faithfulness to the Local over Remote Bases. A likelihood ratio test with one degree of freedom revealed that the model where the grammar does not distinguish between bases is significantly outperformed by the full model (Δlog-likelihood $= 6.9$, $p= 0.0002$ with one degree of freedom). Second, we compared the full Voting model to one where resting activation plays no role in scaling faithfulness constraints, but the grammar is allowed to refer to the distinction between Local and Remote bases. The same test statistics revealed that the full model fits the experimental data significantly better (Δlog-likelihood $= 255.4$, $p< 0.0001$ with one degree of freedom). We see these findings as supporting the central innovations of Breiss' Voting model: the grammar makes reference to specific listed allomorphs with indexed faithfulness constraints, and the resting activation of each of these allomorph determines how severe a given faithfulness violation is.

We also compared how well the MaxEnt model explained the patterns in the data to the statistical model fit in section 3.3: although the two models have different internal structures, we

Constraint	Weight
*Non-initial-[g]	0
Id-[nas]-Local	6.80
Id-[nas]-Remote	6.11

Figure 10: Best-fitting weights for the experimental data, existing compounds.

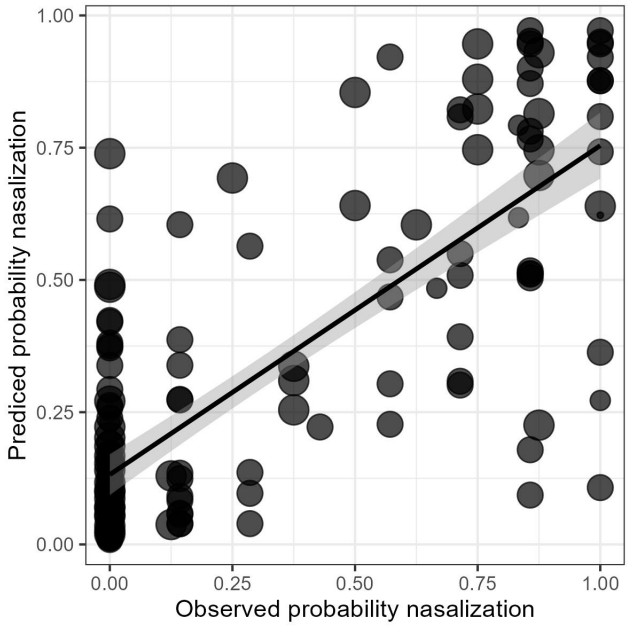

Figure 11: Predicted (vertical axis) vs. observed (horizontal axis) rates of nasalization for categories existing compounds under model in Table 10. Each dot is one compound, with size indicating the number of tokens of that compound type included in the statistical analysis.

can ask whether the theoretically-informed MaxEnt model here does as good a job in explaining the data patterns. 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 — that is, the amount of variance in the data explained by the statistical model's fixed effects (priming of N2, final segment of N1, N2 frequency, compound frequency) and compared it to the R^2 for the MaxEnt model.[4] Since our statistical model is Bayesian, we obtain a median and 95% Highest Density Interval for our marginal R^2: 0.48 [0.30, 0.56]. This is slightly lower, though still relatively comparable, to the MaxEnt models R^2 of .60, 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% CI of the statistical model; this suggests that the theoretically-structured model out-performs the theory-blind statistical one. However, it may

[4]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 the random effects has no direct comparison in the MaxEnt model we fit. For more on the relationship between mixed effects models and hierarchical structures in linguistic data, see Zymet (2019).

also be the case that since the MaxEnt model does not have model variation at the level of the speaker, it may be overfitting the data somewhat, attributing to the grammar variance that might more conservatively be attributed to the speaker. However, since the MaxEnt model does not do demonstrably *worse*, we take this as another piece of evidence supporting the statistical robustness of our theoretically-informed MaxEnt model: it seems to explain about the same amount of variation in the data as the purely-statistical model, and, while perhaps somewhat overfit, does not do so extravagantly, suggested by the fact that the marginal R^2 of the statistical model and the MaxEnt model are relatively close.

4.6.2 Novel compounds

Turning to entirely novel compounds, we fit the analysis sketched in section 4.5 to the data analyzed in 3.3. The fit of the data is plotted in Figure 13, and the constraint weights in Table 12.

Constraint	Weight
*Non-initial-[g]	0
Id-[nas]-Local	0
Id-[nas]-Remote	1.75

Figure 12: Best-fitting weights for the experimental data, novel compounds.

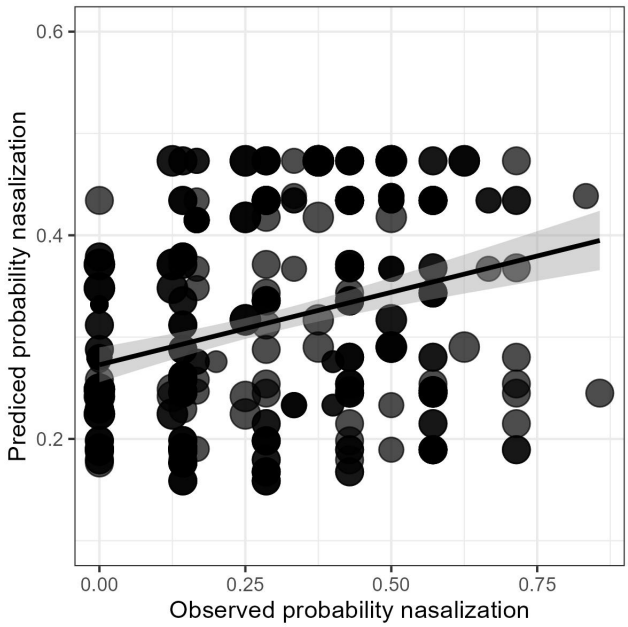

Figure 13: Predicted (vertical axis) vs. observed (horizontal axis) rates of nasalization for categories novel compounds under model in Table 12. Each dot is one compound, with size indicating the number of tokens of that compound type included in the statistical analysis.

We find that again, markedness receives no weight — this is unsurprising, given the findings in existing compounds. However, neither does Id-[nas]-Local: this, while perhaps counterintuitive, follows from the math; with two constraints and two candidates, the *difference* between constraint weights is the quantity of theoretical interest that controls the apportionment of probability.

As above, we confirmed via a likelihood ratio test that taking into account the frequency of N2 meaningfully improves the model fit (Δlog-likelihood $= 9.7$, $p < 0.0001$ with one degree of freedom); since the novel compounds never violated, Ident-[nas]-Remote we did not carry out any weight comparison. We also confirmed that the MaxEnt model is comparable in the variance it explains (R^2 of 0.09) compared to the statistical model fit to novel compounds (marginal R^2 of 0.05 [0.00, 0.18]). While low in absolute terms, it is reassuring that it is in line with the statistical model, suggesting that our grammatical model is not overfit to the data.

4.6.3 Fitting a joint model

To get a more holistic picture of nasalization, we fit a joint model that takes into account the fact that participants were included in the experiment on the basis of exhibiting complementary distribution of [g] and [ŋ] in monomorphemes. In addition to both sets of compound data, the model included the forms employed in the dialect questionnaire that was used to screen participants for inclusion in the experiment, as well as the free forms of N2 that were included in the vocabulary survey, and which participants were required to produce [g]-initially in order to be included in the final dataset, including frequency-based scaling of their faithfulness violations. Because the number of forms that we surveyed was relatively small in comparison to all the data in the Japanese lexicon that exhibits this complementary distribution, we artificially increased the influence of these forms on the equation that is solved to find the best-fitting weights by the *Solver()* function. We accomplished this by multiplying the negative log-probability of the monomorphemic data by a scaling factor; we settled on a multiplier that predicted near-categorical behavior for monomorphemes (over 95% probability of faithful realization); this multiplier has no theoretical significance, and is simply meant to embody the knowledge (Ito and Mester, 2003) that nasalization is taking place against a backdrop of non-variability in monomorphemes. The final model yielded weights listed in Table 14, and predictions plotted in Figure 15.

Note that although the inclusion of non-alternating monomorphemes was intended to give a better picture of the weight of markedness, their inclusion ended up increasing the weight of the Ident-[nasal] constraints far more than that of *Non-Inital-[g]. We take this to mean that much of the variability seen in the experimental data is actually driven by a strong analogical effect, rather than a paradigm-uniformity being parasitic on markedness. This is consistent with the weights obtained when fitting the individual datasets above; when forced to account for the non-alternation of monomorphemes, a modest weight for *Non-initial-[g] suffices, relative to the stronger weights of faithfulness required to drive the paradigm-uniformity effect.

Constraint	Weight
*Non-initial-[g]	0.58
Id-[nas]-Local	10.95
Id-[nas]-Remote	6.30

Figure 14: Best-fitting weights for the experimental data, existing and novel compounds combined.

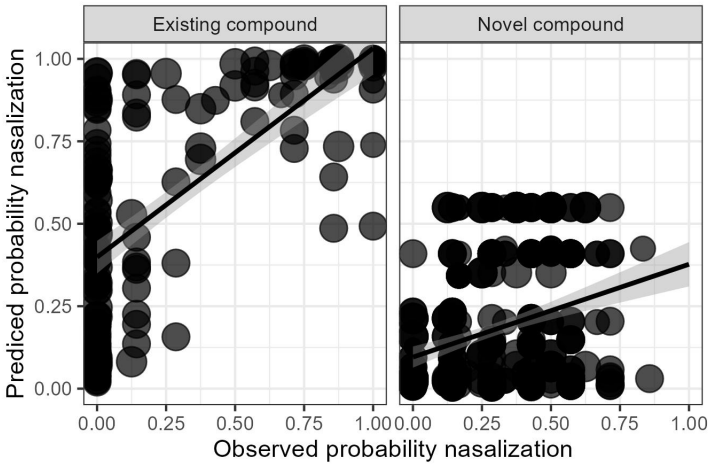

Figure 15: Predicted (vertical axis) vs. observed (horizontal axis) rates of nasalization for categories existing (left) and novel (right) compounds under model in Table 14. Each dot is one compound, with size indicating the number of tokens of that compound type included in the statistical analysis.

5 Discussion

5.1 Whence the weights? Evidence for paradigmatic uniformity 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. By hypothesis, 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 suggests that, setting aside the relationship between frequency and activation, the speakers must be able to attribute different amounts of influence to different faithfulness violation 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 Ident-[nasal]-Local and Ident-[nasal]-Remote.

We fit a grammar with the constraints in section 4.3 and faithfulness-driven weight scaling to the set of compounds reported in Breiss et al. (2021b) that had a free N2. We found that the optimal weights of the grammar were zero for both *Non-Initial-[g] and Ident-[nasal]-Local, and 1.08 for Ident-[nasal]-Remote. 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 4.6.3, but we also found that instead of a tension between faithfulness to the compound itself and 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 IDENT-[nasal]-Local 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 IDENT-[nasal]-Local IDENT-[nasal]-Remote, and found that it was significantly out-performed by the model that allowed the grammar to allot differing weights to faithfulness to different allomorphs (Δlog-likelihood $= 45.3$, $p < 0.0001$ with one degree of freedom). We take this as tentative evidence that there is an empirical basis in the lexicon for different weights for faithfulness to different allomorphs.

5.2 Towards a unified theory of token frequency in phonology

Previous studies have documented cases where token-frequency impacts the phonological grammar, but there is not consensus about the mechanism(s) by which this effect is comes about. Here we briefly summarise some groups of findings in the literature, and then propose a unified theory of the mechanisms by which token frequency interacts with the phonological grammar.

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 (CSD) 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" (see also Kawahara and Sano, 2013). Zuraw (2007) examines frequency-conditioned application of markedness-reducing phonological processes in a corpus of written Tagalog, and finds higher rates of repair within higher-frequency units (words, clitic groups, etc), subject to the markedness principles of the language.

There is also good 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 the study above: higher-frequency N2s act as stronger attractors, yielding *more* faithfulness to their preserved sur-

face [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* rule 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 an 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. 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.

As reviewed above, lexical frequency also influences the resting activation of a lexical item once it is listed in the lexicon. In the Voting theory, 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.

6 Conclusion and future directions

This paper has focused on the velar nasalization in Japanese and its interaction with lexical frequency. We found that variability reported in Ito and Mester (1996, 2003) and Breiss et al. (2021b) for the Yamanote dialect is reproduced experimentally with speakers of the Tōhoku dialect, in existing and novel compounds, both at the group level and at the level of the individual experimental participant. Motivated by this data, we propose an extension to Breiss' Voting theory of Base competition which models frequency-conditioning using scaled faithfulness based on resting activation of the lexical item. We find that the MaxEnt models we fit perform quite well in comparison to the theoretically-underspecified statistical models used to analyze the experimental response data, providing support for the applicability of the Voting theory beyond the phenomenon of Lexical Conservatism for which it was developed, to cases of paradigm uniformity.

There are a number of puzzles that remain that may be fruitfully taken up in future work. First, although we have identified strong effects of N2 and compound frequency in governing the rate of nasalization, it is not clear whether the claim by Ito & Mester that the prosodic status specifically (free vs. bound) itself matters, above and beyond this effect of morpheme frequency. Answering this question would require a more targeted comparison of N2s which are obligatorily bound with frequency-matched N2s which can be prosodically free-standing.

A second set of questions concerns the proper way to model resting activation. In this paper, we adopted a simple sigmoidal relationship between frequency and resting activation, and make a simplifying assumption, based on the statistical analysis of the experiments, that *only* frequency contributed to resting activation. Much future work is needed to verify some of the assumptions made here about the shape of the linking function between lexical characteristics and resting activation, and also about the way different lexical characteristics (log-frequency, primedness, semantic similarity between Bases (cf. Breiss, 2021)) are combined to yield the input to this linking function. One advantage of the Voting theory's treatment of frequency effects is that it explicitly relies on an externally-validated construct like resting activation; thus, in principle one might be able to use well-developed models of lexical structure and dynamics such as that of Stille et al. (2020) as a way to calculate resting activation, once suitable equivalents for non-English languages are developed that encode the relevant phonological information. It is also possible that a different conceptualization and treatment of resting activation all together may better fit the growing set of empirical data addressed by the Voting theory (Tōhoku Japanese in this paper, Mexican Spanish and North American English in Breiss (2021)).

Moving on to the corpus results, we tested here only the weakest version of the hypothesis that the speaker learns the constraint weights from the distribution of data in her lexicon. We found that the corpus of compounds with free N2s analyzed by Breiss et al. (2021b) provided evi-

dence to distinguish the weight of IDENT-[nasal]-Local from IDENT-[nasal]-Remote, and qualitatively match the weights that emerge when fitting the novel compound data. Future more detailed modeling of the learning data available to speakers of Japanese who are acquiring nasalization is therefore needed to further articulate the relationship between the generalizations exhibited in the experiment — particularly in existing compounds that exhibit variation — and those latent in the data available to the learner.

Finally, we take the success of the Voting model in capturing both Lexical Conservatism and paradigm uniformity effects to be suggestive that the interaction of lexical characteristics and phonological grammar which the Voting Bases model proposes are on the right track as a general theory of modeling output-oriented phenomena in phonology more broadly.

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.

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

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

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, 80(1):1–28.

Bybee, J. (1999). Usage-based phonology. *Functionalism and formalism in linguistics*, 1:211–242.

Chomsky, N. (1957). *Syntactic structures*. De Gruyter Mouton.

Chomsky, N. and Halle, M. (1968). The sound pattern of English.

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.

Gahl, S. and Yu, A. C. L. (2006). Introduction to the special issue on exemplar-based models in linguistics.

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.

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.

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.

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., Ono, H., and Sudo, K. (2006). Consonant co-occurrence restrictions in yamato japanese. *Japanese/Korean Linguistics*, 14:27–38.

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

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

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

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

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.

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., and Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language resources and evaluation*, 48(2):345–371.

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.

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

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., and Bürkner, P.-C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research*, 62(5):1225–1242.

NHK (1993). Nhk broadcasting culture research institute.

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*.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roos, M., Martins, T. G., Held, L., and Rue, H. (2015). Sensitivity analysis for bayesian hierarchical models. *Bayesian Analysis*, 10(2):321–349.

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.

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

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.

Stille, C. M., Bekolay, T., Blouw, P., and Kröger, B. J. (2020). Modeling the mental lexicon as part of long-term and working memory and simulating lexical access in a naming task including semantic and phonological cues. *Frontiers in psychology*, 11:1594.

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

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., and Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, 71:147–161.

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.

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.

Zuraw, K. R. (2000). *Patterned exceptions in phonology*. PhD thesis, University of California, Los Angeles.

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.