To appear in Phonology
This is a near-final version.

Token frequency modulates optional paradigm uniformity in Japanese voiced velar nasalization

June 26, 2024

Abstract

This paper explores the role of token frequency and global prosodic length in conditioning optional paradigm uniformity in Japanese voiced velar nasalization, with data from two *wug*-tests carried out with speakers of Tōhoku Japanese. Experiment 1 demonstrates that frequency-conditioning observed in corpus data is reproduced in existing and novel compounds, and holds at the level of the speaker. Experiment 2 focuses on a typologically-unusual pattern where overall compound length in mora seems to influence nasalization, a candidate for a "counting" pattern in phonology. We find that instead of overall length, speakers are sensitive to the length of the second member of the compound, undercutting the viability of the mora-counting analysis. We discuss the importance of the results in adjudicating between existing models of how token frequency impacts the phonological grammar, and suggest that only theories that allow individual morphemes to exhibit frequency-sensitive behavior are sufficiently expressive to model the finding.

1 Introduction

This paper explores the role of token frequency in conditioning optional paradigm uniformity in Japanese voiced velar nasalization (henceforth, simply "nasalization"). We begin with a review of the salient findings of Breiss et al. (2021)'s corpus study of Yamanote Japanese, where the probability of nasalization is influenced by overall compound frequency, the frequency of the second element (N2), and, unexpectedly, the total number of mora in the compound. We then report two *wug*-tests with speakers of the phonologically-conservative Tōhoku dialect of Japanese (spoken in northern parts of the main island of Japan). The first experiment demonstrates that the frequency-conditioning observed in corpus data is reproduced in existing and novel compounds, and holds at the level of the individual speaker. The second experiment focuses specifically on whether, after controlling for frequency, overall compound length influences nasalization, and finds that it does not, instead revealing sensitivity to only the size of N2, regardless of the size of N1 (the first element in the compound). We conclude by discussing the implications of these findings for existing theories of token frequency in the grammar.

1.1 Token frequency and phonological patterns

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). The classic argument given in that book was that the transitional probability in English from *fragile* to *whale* is plausibly zero, but this sequence does not make a sentence containing it ungrammatical. The *Sound Pattern of English* (Chomsky and Halle, 1968) (SPE) and much subsequent work that followed its tradition did not seriously consider the effects of lexical frequency or other statistical information on phonological patterns either (Coetzee and Kawahara, 2013). Phonological rules, as formulated in the SPE style, were not sensitive to usage frequencies. More generally, effects related to statistical information, including usage frequency were, implicitly or explicitly, considered to be a matter of performance and not competence (Newmeyer, 2003) (see the reviews presented in Coetzee and Kawahara 2013 and Shaw and Kawahara 2018).

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). This observation was made as early as Hooper (1976), who observed that schwa is more likely to be deleted in frequent words like *memory* than in non-frequent but phonologically-similar words like *mammary*. A perhaps more well-studied phenomenon is [t/d]-deletion in English, which is very common in frequent words like *and* and *list*, but not as very common in less frequent words like *mast* or *jest* (Coetzee and Kawahara, 2013). Likewise, the experiment by Kawahara (2011) demonstrates that OCP-induced geminate devoicing in Japanese loanwords is judged to be more natural in words with higher lexical frequencies. File-Muriel (2009) argues that in some dialect of Spanish, weakening of /s/ occurs more likely in high frequency words than in low frequency words. 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.

Lexical frequency has also been found to be related to the degree a lexical item deviates from grammar-wide trends in phonological alternations. Smith and Moore-Cantwell (2017) found that higher-frequency comparative constructions are more likely to flout grammar-wide trends driven by markedness. For example, the adjective *likely* itself forms the analytic comparative *more likely* more than 99% of the time, while other phonologically-comparable, lower-frequency forms take the analytic comparative only around 45% of the time. 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.

These cases show that lexical frequency interacts with phonological patterns to a non-negligible degree, and any adequate theory of phonology must have a means to model its effects. In usage-based phonology (Bybee, 1999) as well as exemplar-theoretic phonology (Gahl and Yu, 2006), usage frequency is directly encoded in grammatical model. In the generative tradition too, Coetzee and Kawahara (2013) argue that it is necessary—and possible—to incorporate the effects of lexical frequency in the formal 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 current phonological research. In this paper, we provide experimental data bearing on this question, and discuss how different contemporary generative models of lexicon-phonology interaction might model the role lexical token frequency has on conditioning nasalization in both existing and novel compounds.

2 Voiced Velar Nasalization

2.1 The basic patterns

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. /kagami/→[kaŋami] "mirror" vs. /gimu/→[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 both surface free-standing with initial [g], and 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: (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”

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

2.2 The corpus study of Breiss et al. (2021)

Breiss et al. (2021) carried out a quantitative investigation of the variability and optionality of nasalization, noted by Ito and Mester (1996). We reproduce Breiss et al.’s quantitative analysis here, but refer the interested reader to the full text a more detailed treatment.

The data for the analysis was drawn 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, and was annotated with frequency information from the Balanced Corpus of Contemporary Written Japanese (BCCWJ) (Maekawa et al., 2014). Figure 1 plots the number of words whose pronunciation was labeled as "undergoing" or "preferring to undergo" nasalization in the dictionary, divided into whether N2s were bound morphemes (left panel) or free-standing morphemes (right panel).[1]

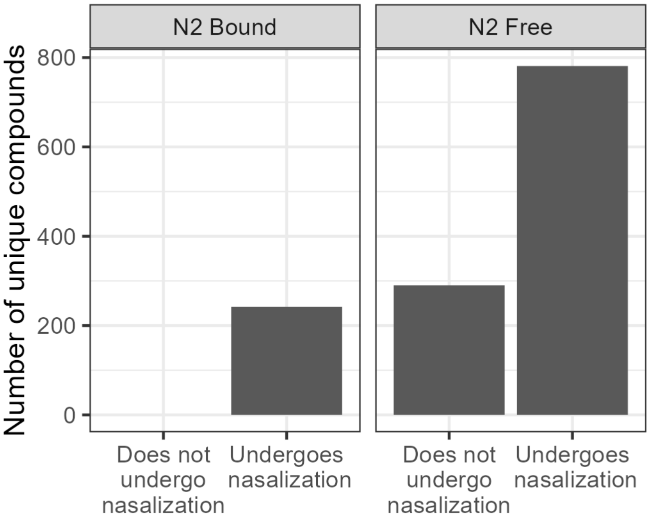

Figure 1: Division of the corpus of compounds according to whether a given compound undergoes (or prefers to undergo) nasalization or not (horizontal axis), divided by whether or not N2 is able to occur as a free form (panels). The vertical axis plots the number of unique compounds in each category.

The data in the left panel show that bound morphemes invariably undergo nasalization. The data in the right panel support the claim of optionality in cases of a free-standing N2, offering quantitative evidence supporting Ito and Mester (1996).

Turning next to the determinants of variation in compounds with free N2s, Breiss et al. (2021) found the frequency of the whole compound and its second member to both be reliable predictors of whether a given compound would undergo nasalization in dictionary data. These data are plotted in Figure 2. The left panel shows that more frequent compounds are more likely to show nasalization; the right panel shows, on the other hand, that compounds with more frequent N2s are less likely to show nasalization.

---

[1] Breiss et al. (2021) categorized morphemes appearing at least once as independent words in the BCCWJ corpus as free, and those not appearing independently as bound, with the understanding that this classification may not necessarily perfectly align with native speakers' intuition.

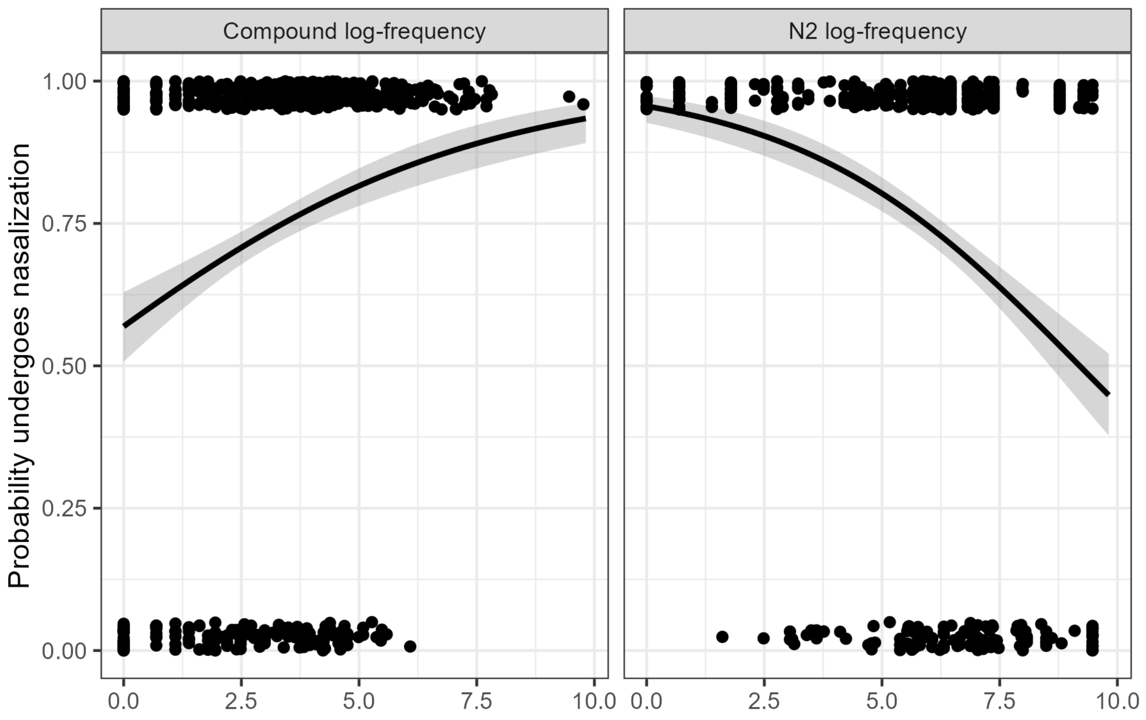

Figure 2: The effects of whole compound frequency (left panel) and N2 frequency (right panel) on the probability of nasalization, with binomial smooths. One dot represents one lexical item; vertical jitter has been added for readability.

Breiss et al. also found that the prosodic length of a compound was strongly related to whether it underwent nasalization, with shorter compounds being more likely to undergo, and longer compounds being more resistant to alternation. Figure 3 plots these findings.

Taken at face value, these data suggest that the synchronic grammar of the Yamanote dialect of Japanese exhibits a dependency between global prosodic length and a local segmental alternation which flies in the face of a often-cited claim that "phonology doesn't count" (McCarthy and Prince 1986 among many others). Though this is not as ironclad a generalization as it is often assumed to be (see in particular the arguments and data presented in Paster 2019), the trend in Figure 3 is striking, with few typological parallels noted in the literature.[2] An alternative explanation not explored by Breiss et al., however, is that this relationship is illusory, a function of Zipf's Law of Abbreviation (Zipf, 1935). The Law states that "the magnitude of words stands in an inverse (not necessarily proportionate) relationship to the number of occurrences" (Zipf, 1935, p.23); that is, more frequent words tend to be shorter. If we assume that this relationship also governs compounds in the Japanese lexicon, we would expect to find the pattern shown in Figure

---

[2] Paster (2019) notes that all the cases that involve counting are supra-segmental patterns, and thus, the pattern in Figure 3, if true, is indeed a surprising finding.

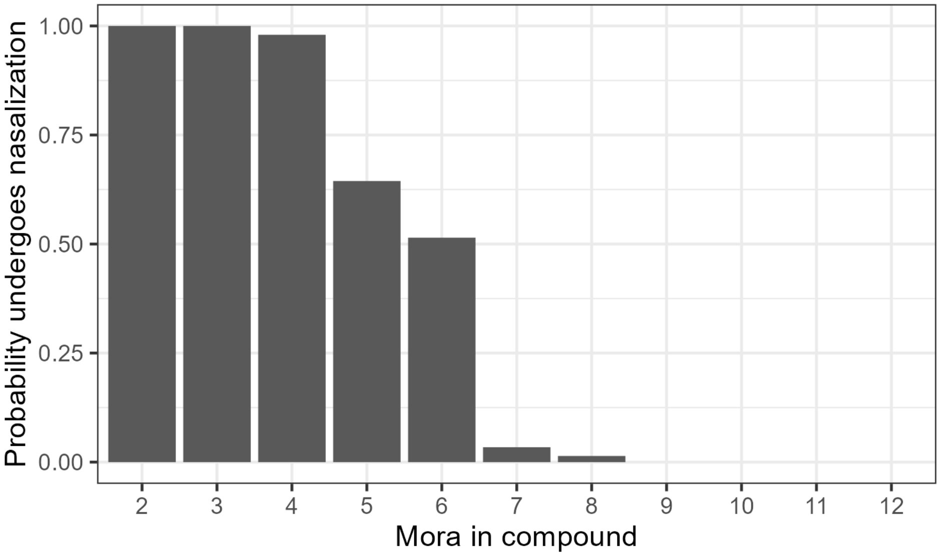

Figure 3: The effects of compound length in mora on the probability of nasalization in the corpus study of Breiss et al. (2021), reproduced with slightly different axis labels for consistency.

3, where compounds appear both shorter and more likely to undergo nasalization, both stemming from their high frequency.

As with all analysis of existing lexical items relying on corpus data, it is not clear how much evidence these data can provide about the synchronic grammars of speakers of such conservative dialects, and whether the frequency effects (functionally motivated) or the apparent length-effect (typologically very unusual) are in fact cognitively represented as such. Resolving these questions is of critical importance to how we construct our phonological theories, including whether (and if so, how) usage frequency of the morphemes impacts the the synchronic grammar.

3 Experiment 1

3.1 Aims

Our first experiment has three empirical goals. The first is to determine 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 is hard to resolve in corpus-based studies of variation, including Breiss et al. (2021), because it is possible that apparent variation in a corpus actually results from collapsing across different speakers with different categorical grammars.[3] The second goal is to see whether the frequency-conditioning of this

---

[3] This is not to say that addressing the role of frequency within each individual speaker is impossible in corpus studies in general, insofar as speaker information is coded and a sufficient amount of data is available.

variability is also active at the speaker level. The third goal is to see whether the frequency effect extends to novel compounds composed of existing morphemes of varying frequencies, or whether it is limited to whole forms with which the speakers might plausibly have stored in their mental lexicon.

The status of novel compounds is of great relevance for distinguishing between phonological theories of frequency effects: The UseListed theory of Zuraw (2000) holds that 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 indicate 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.

Finally, the experiment included a priming manipulation, designed to probe whether other characteristics of the lexical entry other than frequency might impact their phonological behavior. Breiss (2021) found in several experiments on English derivational morphology that priming non-local paradigm members influenced the way that novel derivatives were formed; for example, a novel derivative in *-able* based on the stem *lábor* was more likely to be pronounced *labórable*, with stress on the second syllable, when the form *labórious*, with matching stress placement, was primed. Breiss (2021) implemented the manipulation by performing a vocabulary check with a random subset of non-local paradigm members (like *labórious*) before the main task of *-able*-formation, with the rest after; thus, for any individual, half of the items were primed, and the others were not. In this experiment, all but one of the participants completed two repetitions of the experiment, so on each of two runs the participants saw one half of the items primed.

3.2 Methods

Supplementary material of this experiment, as well as that of Experiment 2, can be accessed at *https://osf.io/avnpw/?view_only=cd2afdcc183f4de3ac1261b4af66f08d.*

3.2.1 Stimuli

Stimuli for the experiment were 301 compounds, 81 of which were existing forms (largely drawn from the corpus used in Breiss et al. (2021)), and the rest of which were novel. The complete list of stimuli is available at in the supplementary material. The existing compounds were selected to represent a range of nasalization probabilities, based on the frequency of the whole compound and that of the N2. Due to the challenges in controlling length and morphological composition, there was variation in these aspects among the existing compounds. Specifically, their lengths ranged from 2 to 8 mora in total, and certain compound members 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 six bimoraic N1s (e.g., /ʤuu/ “heavy, multiple”, /tei/ “low”) with 30 bimoraic N2s (e.g., /gin/ “silver”, /gjaku/ “reverse”), with the N2s selected to be of varying frequencies. Examples of novel compounds are /ʤuugin/ (/ʤuu/ “heavy, multiple” + /gin/ “silver”) and /toogan/ (/too/ “Chinese” + /gan/ “wild goose”). Due to the difficulty of finding enough monomorphemic 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 (for which, see 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.

3.2.2 Participants

We 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 voiced velar nasalization alteration. We should nevertheless bear in mind that the Tōhoku dialect is a dialect that is different from the Yamanote dialect that was analyzed by Breiss et al. (2021).

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 restriction, 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.[4] As two of the three goals of the experiment (see Section 3.1) address questions at the level of the individual speaker’s grammar, all but one person (Speaker 7) 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 also which N2s were primed (on which, see section 3.2.3 immediately below). The participants were

---

[4] Many speakers, particularly younger ones, have lost this pattern of nasalization due to the influence of the “standard” Tokyo dialect, which has also lost the nasalization pattern, retaining [g] for /g/ in all contexts.

paid approximately 20 USD per experimental session.

3.2.3 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. As the participant carried out the production task, the second author coded based on auditory impression whether [ŋ] or [g] was produced on a given trial. The decision was usually clear-cut for the experimenter, who is a phonetically-trained linguist and also a native speaker of Japanese. In cases of uncertainty, tokens were coded based on visual inspection of the spectrogram: tokens involving clear striations for bursts were coded as [g] while ones without them were coded as [ŋ]. The present study does not address the question of phonetic gradience in the implementation of nasalization, for to the best of our knowledge, no studies have demonstrated a gradient nature of the nasalization process; future analysis of the recorded data, however, may yield insight into these questions. Here, we follow the practice in the sociolinguistic (Hibiya, 1995) and phonological literature (cited above) and treat nasalization as binary, reflecting the intended outcome of the speaker's grammar.

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*).

As noted in section 3.1, we took advantage of the two experimental sessions with each participant to examine what effect priming an N2 might have on whether it exhibits nasalization when produced in a compound. 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 followed by a semantically neutral, short sentence-ending particle. These compounds, a mix of novel and existing forms, were presented in a random order. All compounds were presented in *kanji* characters, which do not distinguish between [g] and [ŋ]. Thus we can infer that whatever form the speakers produced ([g]-ful or [ŋ]-ful) is minimally biased by the experimental setup.

3.3 Statistical analyses

After the data collection was complete, each compound member was classified as whether it was known to the speaker (familiarity score > 1) or not (familiarity 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, available at the osf repository (see above for the link) .

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 the object which we mine for analytic insights. For more comprehensive 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% Credible Interval (CI) which is presented as a bracketed range, and the probability of direction, noted $P\ (|\hat{\beta}| > 0)$. The first measure indicates the median posterior value of the parameter, and the range which encompasses the central 95% of likely values. 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 speaker 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 that the data we collected were sufficiently informative that our prior beliefs about likely parameter values mattered only nominally; see the 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 nasalization, 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.

3.4 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.

3.4.1 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; this is plotted in Figure 4. Note that while Breiss et al. (2021) considered *relative* frequency of the N2, here we consider the frequency of the N2 and compound in their own right. Finally, we use the natural logarithm of the token frequency, rather than its raw value, as is standard practice.

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. (2021). As N2 frequency rises relative to a fixed value of compound frequency, the probability of an individual compound exhibiting nasalization drops (Figure 4, right facet); holding N2 frequency steady while increasing the frequency of a compound from low to high also increases the probability of nasalization (Figure 4, left facet).

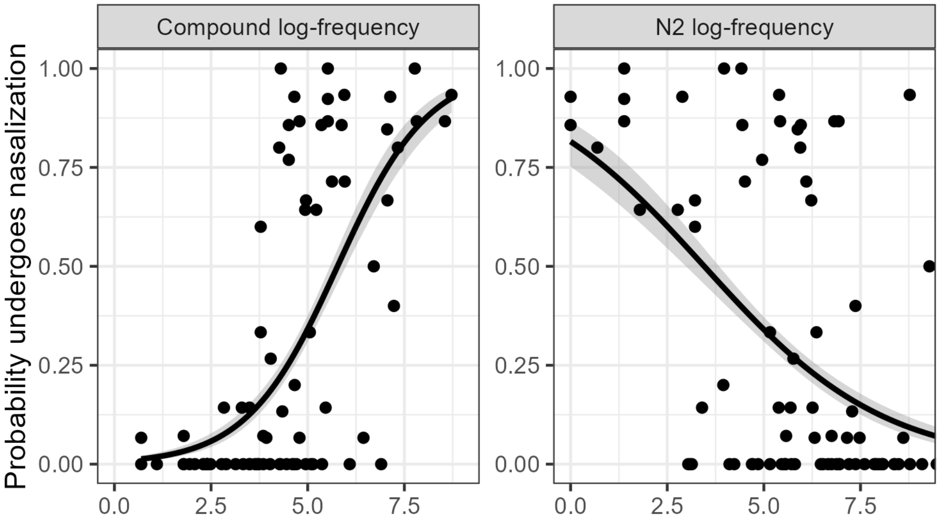

Figure 4: Probability of nasalization (vertical axis) plotted against compound log-frequency (left facet) and N2 log-frequency (right facet), with binomial smooths.

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 5.

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 3.2.3. However, priming seemed to have no meaningful effect on rates of nasalization (left plot of Figure 6). 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 the lexicon of the participant in such a way for it to be experimentally detectable in their behavior on the production task. 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 relative to the rest. 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, but more targeted work on this topic is needed to better support this hypothesis.

Second, we expected, based on the findings of Breiss et al. (2021) in Japanese compounds and

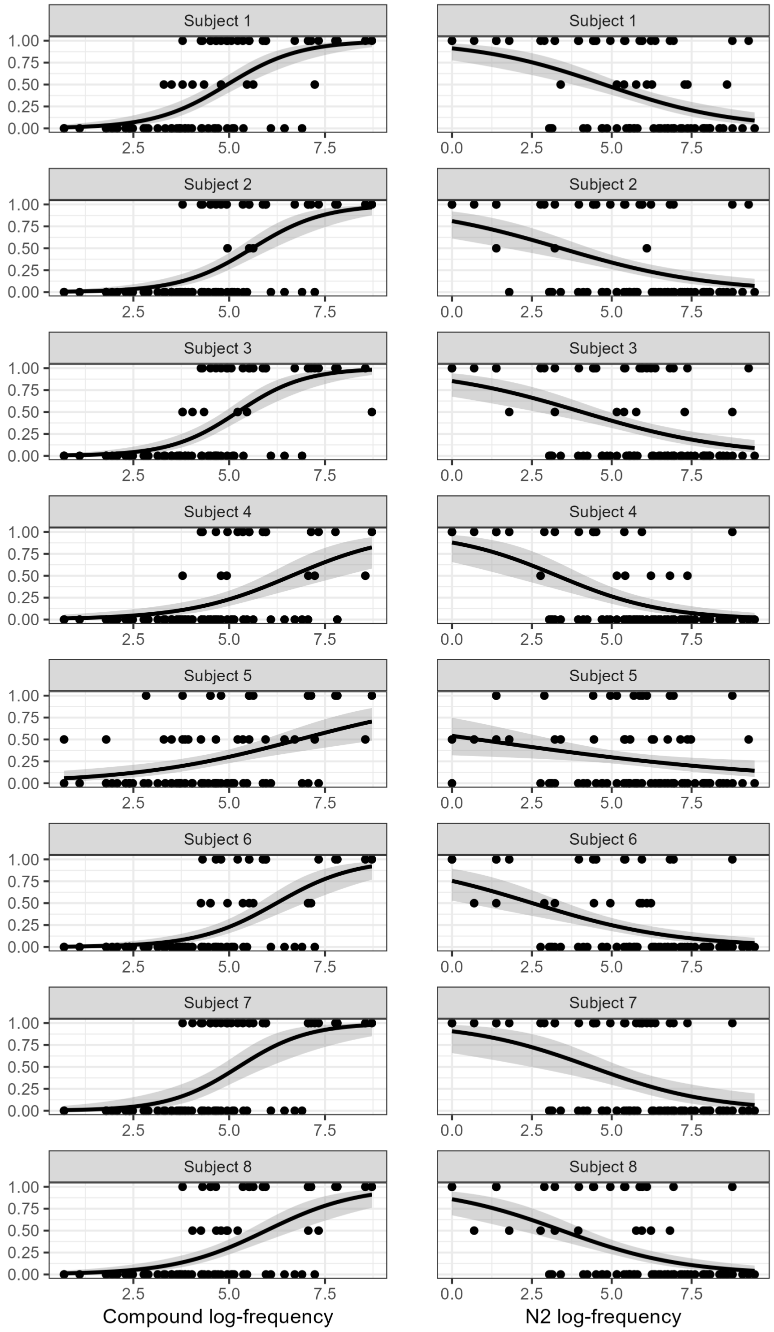

Figure 5: Probability of nasalization (vertical axis) plotted against compound log-frequency (left facet) and N2 log-frequency (right facet) for each individual speaker (row), with binomial smooths.

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 found the opposite trend in the experimental data, as shown in the right panel of Figure 6, but only superficially. Although the left bar is higher than the right bar, 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.

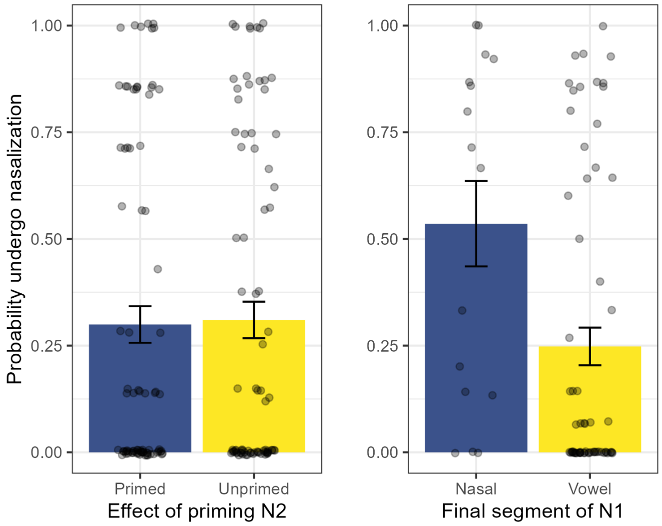

Figure 6: Probability of nasalization with standard error (vertical axis) plotted against priming of N2 (left) and final segment of N1 (right).

Table 1 presents the results of a Bayesian mixed-effects logistic 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 3.3, 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 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 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 Speaker 4 the effect was less certain (80%), and only Speaker

Parameter	*Median*	*95% CI*	$P(\lvert\hat{\beta}\rvert > 0)$
Intercept:			
N1 final segment = *vowel*			
N2 primed = *no*			
Compound scaled log freq. = *mean*			
N2 scaled log freq. = *mean*	-0.92	[-2.34, 0.53]	
N1 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.

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 individual grammar, though the factors influencing individual-level variation in effect size remain for future research.

3.4.2 Novel compounds

Turning to novel compounds, we find that the frequency effect holds here as well, though with a smaller magnitude. Figure 7 plots only N2 frequency; since the the compound is entirely novel, its frequency is naturally zero. The downward-sloping smooth qualitatively matches the one found in the right panels of Figures 4 and 5.

Breaking this result out by individual in Figure 8, 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 in novel compounds; this is shown in Figure 9. Since none of the six N1s we selected for constructing the novel compounds were nasal-final, 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

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

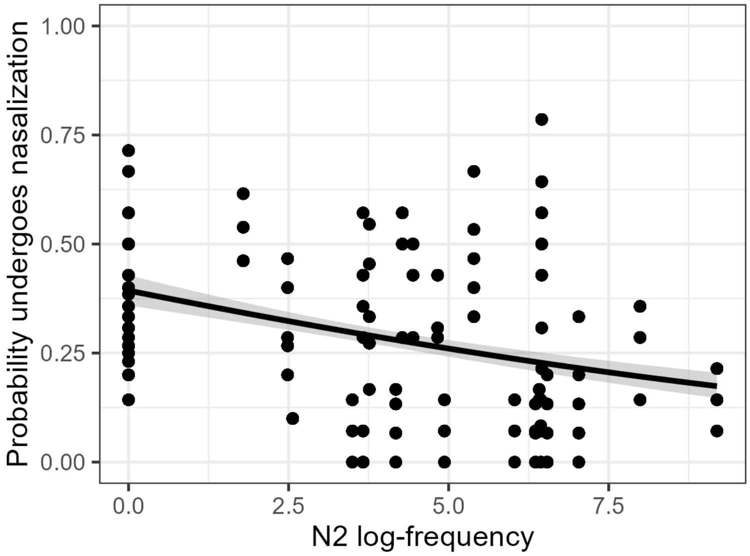

Figure 7: The probability of undergoing nasalization, plotted against N2 log-frequency (novel compounds), with a binomial smooth to aid readability.

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

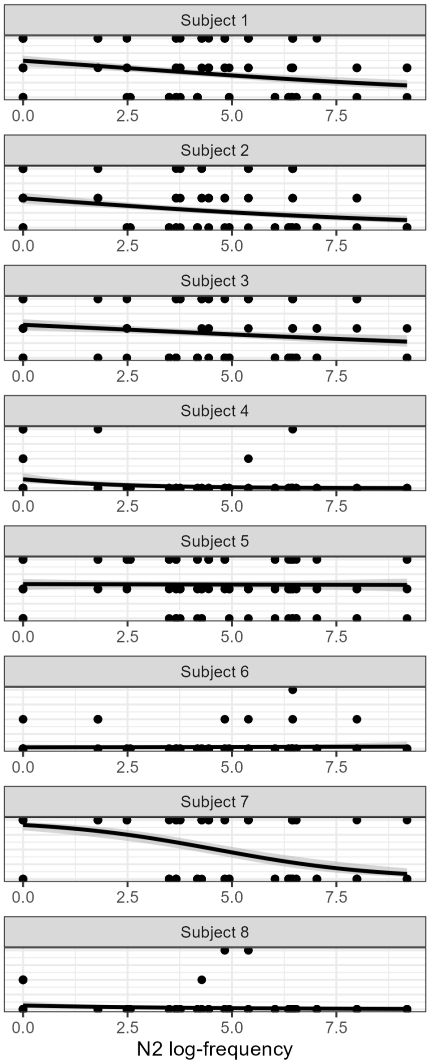

Figure 8: Probability of nasalization (vertical axis) plotted against N2 log-frequency (horizontal axis) for each individual (row), with binomial smooths.

are less likely to exhibit nasalization. In one individual, Speaker 5, however, the effect is in the opposite direction; this is unexpected, and further puzzling because the same speaker 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.

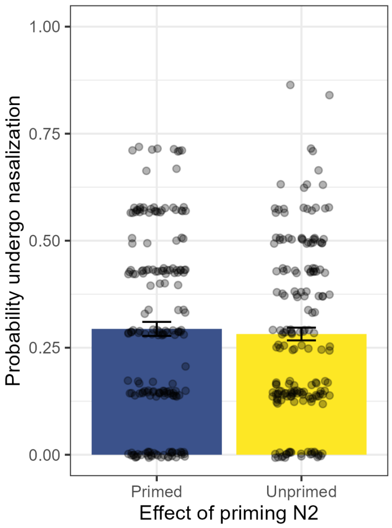

Figure 9: The probability of undergoing nasalization with standard error (vertical axis) based on whether the N2 was primed (horizontal axis) in novel compounds.

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.

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

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

3.4.3 Within-speaker consistency

Finally, we examined within-speaker consistency in the frequency effect, comparing existing compounds to novel compounds. If the frequency effect is active at the level of the individual speaker, it is reasonable to assume that some speakers might be more sensitive or affected by frequency than others, and that this effect impacts the treatment of existing and novel compounds by the phonological grammar. Figure 10 plots the coefficient of N2 frequency in novel compounds (vertical axis) against that of N2 frequency in existing compounds (horizontal axis); points represent the median value of the posterior, and the error bars encompass the 95% CI. A linear smooth is added for readability, and a red dashed line with slope 1 is provided for comparison.

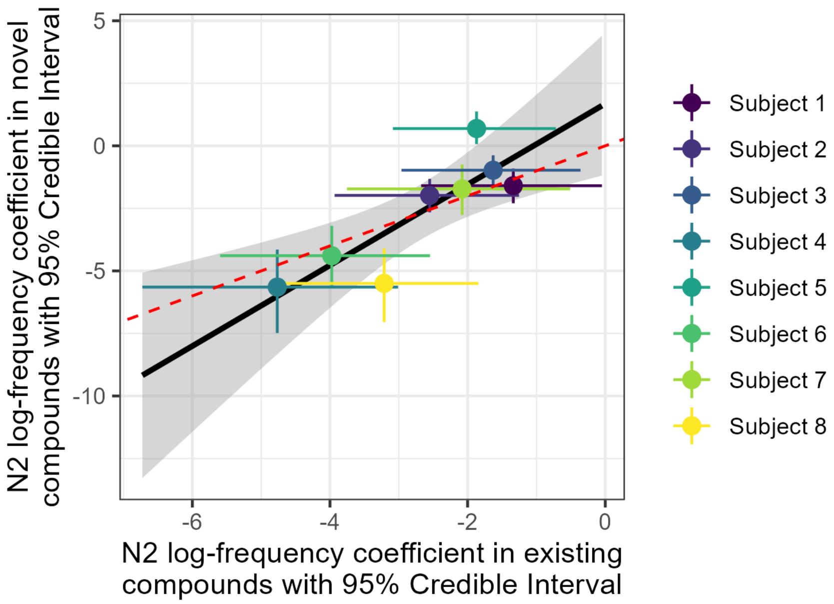

Figure 10: The coefficient of N2 log-frequency in novel compounds, derived from the model in Table 1, is plotted on the horizontal axis, and the coefficient for N2 log-frequency in existing compounds, derived from the model summarised in Table 3, is plotted on the vertical axis. Points represent median values of the posterior, and ranges encompass 95% CI, colors represent speakers, and a linear smooth has been added for readability, with the line of slope 1 intersecting the origin in dotted red.

While it is clear that individual speakers exhibit variation in how strongly they are affected by lexical frequency (cf. the differences of magnitude of the relevant median coefficient values in Tables 1 and 3), what is important to note here is that individuals are self-consistent in their variation — speakers who are more sensitive to N2 log-frequency in existing compounds are also more sensitive in novel compounds, and similarly for those that are less sensitive (Pearson's ρ = 0.86, p = 0.006). Further, we can qualitatively observe that the effect of frequency is of roughly the same magnitude across both compound types, corresponding to the overlap between the grey smooth's uncertainty and the red dashed line of identity in Figure 10. This suggests that however token frequency influences the phonological grammar, it does so indiscriminately, ignoring the phonological or morphological nature of the items themselves (here, existing vs. novel compounds).

In summary, Experiment 1 demonstrates that the conditioning effect of both compound frequency and N2 frequency holds in productions of existing compounds, as suggested by the corpus analysis of Breiss et al. (2021). It also demonstrates that the frequency-conditioning holds at the level of the individual speaker, precluding an explanation of the variable pattern in terms of averaging across individual categorical grammars. Finally, and most importantly for the construction of phonological theories of lexicon-grammar interaction, it confirms that the frequency of N2 plays a role even in entirely novel compounds, indicating that the grammar must be sensitive to the lexical frequency of morphemes it manipulates.

4 Experiment 2

Experiment 2 complements the coverage of Experiment 1 by systematically varying the length of the compound (two, three, or four mora), and the frequency of N2. It also more carefully controls for morphemic composition, which was somewhat compromised in Experiment 1 in favor of having a greater range of frequencies and a sufficient number of stimuli. Recall that the goal of the length manipulation in Experiment 2 was testing whether the corpus pattern that suggested a relationship between overall compound size and propensity to nasalize, a possible case of phonology "counting". Although Jiang (2023b) carried out a similar test of the relationship between nasalization and compound size and found that it did not generalize to novel compounds, his study was not restricted to only speakers of a phonologically-conservative dialect which preserves the nasalization pattern. Therefore we consider it important to double check this finding, especially given that it touches on a core question of phonological theory. Finally, since the stimuli in Experiment 2 are all novel compounds, it serves as a replication of the frequency effect in novel compounds from Experiment 1.

4.1 Methods

4.1.1 Stimuli

Six N1s were selected, three monomoraic and three bimoraic (e.g., [ta] "many", [tai] "pair") , and were fully crossed with 28 N2s, roughly balanced between monomoraic ($n = 12$; e.g., [go] "Go", [ga] "moth") and bimoraic ($n = 16$; e.g., /gin/ "silver", /gjaku/ "reverse"). All compound members were monomorphemic. Within each N2 length, stimuli sampled a wide range of log frequencies (monomoraic: 1.39-7.40, mean 4.61; bimoraic: 3.56-8.0, mean 5.51), and did not differ significantly in overall frequency via a two-sided *t*-test (t = -1.41, df = 18.98, p = 0.17). This yielded 168 novel compounds for analysis.

We also included 32 compounds where the N2 was /k/-initial, as part of a separate experimental condition designed to test the interaction of Rendaku and nasalization. The research questions at stake in this condition are disjoint from those pursued in this paper, so these stimuli are excluded from our analysis and not discussed further.

4.1.2 Participants

A total of 13 participants were recruited to participate in Experiment 2 via word of mouth and snowball sampling. As in Experiment 1, each participant was screened using the dialect questionnaire to ensure that they spoke a dialect where [g] and [ŋ] are allophonically distributed in monomorphemic words. Of the 13 interviewed, 12 passed the dialect questionnaire (see section 3.2.2). Three participants also participated in Experiment 1, while the other nine did not.[5] Unlike Experiment 1, participants only participated in this experiment once, due to time and resource constraints.

4.1.3 Procedure

Procedure for Experiment 2 was identical to Experiment 1, as was data processing and statistical analysis.

4.2 Results

We begin with the central question of Experiment 2: does compound length in mora — above and beyond other factors, like frequency — influence how often speakers apply nasalization in novel compounds? Figure 11 plots the mean nasalization rate by compound length and N2 frequency.

We find that shorter compounds appear to undergo nasalization more often than longer compounds, suggesting that the effect of global prosodic length seen in Figure 3 may be active in the grammar of Tōhoku Japanese speakers; the N2 frequency effect in novel compounds observed in Experiment 1 seems to replicate in Experiment 2.

To assess the credibility of the visual trends observed in Figure 11, we fit a Bayesian mixed-effects logistic regression model to the experimental results, which is given in Table 5 below. The model structure followed the outline in section 3.3, and had a three-level ordered factor of

---

[5] While having speakers in common between the two experiments may not be ideal, it was quite difficult to find speakers who both spoke the conservative dialect and also were comfortable with participating in an experiment using Zoom. The two experiments were conducted over the span of three years and therefore we think it unlikely that detailed memory of the first experiment influenced their performance in the second.

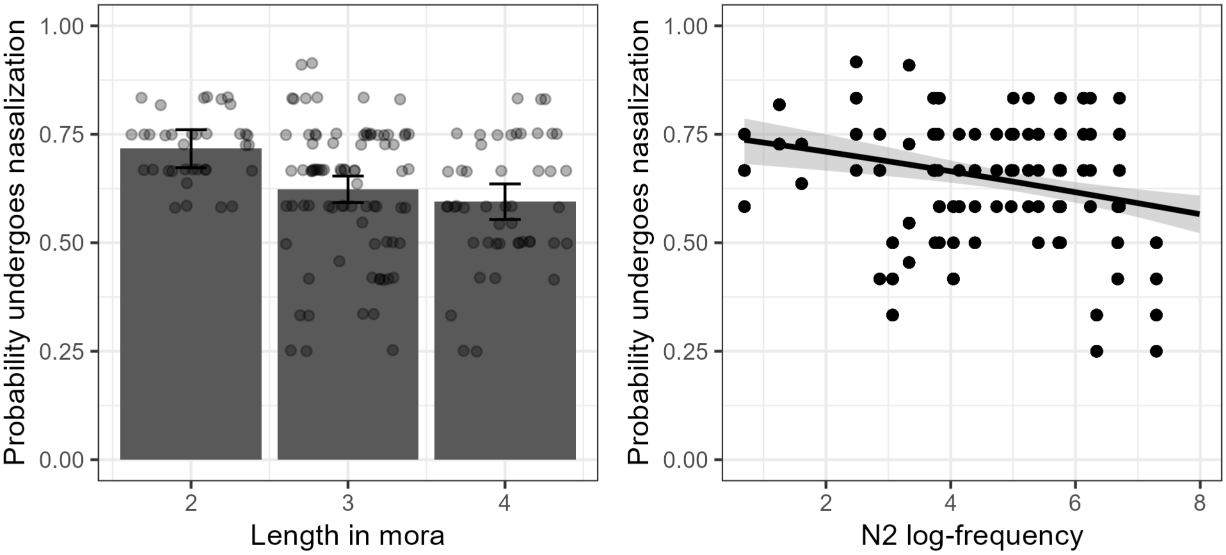

Figure 11: Probability of nasalization plus standard error (vertical axis) plotted against the length of the compound in mora (horizontal axis, left) and N2 log-frequency (horizontal axis, right), with a binomial smooth.

compound length in mora (levels 2, 3, and 4), N2 log-frequency, whether N2 was primed, and the interaction between priming and N2 log-frequency. Since all N1s in Experiment 2 were vowel-final, we did not assess possible avoidance of adjacent nasals across the N1-N2 boundary.

The negative coefficient for N2 log-frequency indicates that there is substantial statistical support for the visual impression that compounds with higher-frequency N2s undergo nasalization less often. Post-hoc pairwise comparisons between levels of *Mora in compound* using the *emmeans* package (Lenth et al., 2019) revealed that three-mora compounds underwent nasalization at a credibly lower rate than two-mora compounds, following the visual intuition in Figure 11, but that the difference in nasalization rate for three- and four-mora compounds was of a smaller magnitude, and had a posterior distribution of credible values that overlapped zero, though around 80% of the credible values are compatible with a small positive effect ($\hat{\beta}$ = 0.28, 95% CI [-0.44, 1.00], $p(|\hat{\beta}| > 0)$ = 0.79). Finally, as in Experiment 1, neither priming nor its interaction with N2 log-frequency meaningfully predicted nasalization in participant responses (not plotted for space). In sum, Experiment 2 replicates the finding from Experiment 1 that the phonological treatment of even novel compounds are subject to the lexical frequency of their N2s, and also provides partial support for the global prosodic length effect observed in the corpus, a topic to which we now turn in more depth.

Parameter	*Median*	*95% CI*	$P(\lvert\hat{\beta}\rvert > 0)$
Intercept:			
Mora in compound = *2*			
N2 primed = *no*			
N2 scaled log freq. = *mean*	1.65	[-0.72, 4.08]	
Mora in compound = *3*	-0.78	[-1.43, -0.10]	0.98
Mora in compound = *4*	-1.07	[-1.82, -0.32]	≈ 1
N2 primed = *yes*	0.15	[-0.39, 0.67]	0.74
N2 log freq. *(one unit increase)*	-0.41	[-0.83, 0.01]	0.97
N2 log freq.× N2 primed =*yes*	-0.13	[-0.59, 0.31]	0.71

Table 5: Model of novel compounds with free N2s in Experiment 2. Coefficients are in log-odds, with positive signs indicating an increase in probability of nasalization relative to the intercept.

4.3 Discussion

Since the planned analysis of overall compound length revealed only a partial effect, we carried out an exploratory analysis on the compounds to better understand the nature of whatever length effect might exist.

In contrast to the gentle downward trend in nasalization for longer compounds seen in the aggregated data (Figure 11), when we examine the data at the level of the compound's moraic composition, we find a strikingly different pattern. Rather than being intermediate between two- and four-mora compounds, the two three-mora groups diverge in behavior on the basis of the length of their N2. Compounds with monomoraic N2s pattern with two-mora compounds (which have monomoraic N2s) in having a higher overall nasalization rate, while those that have bimoraic N2s pattern with four-mora compounds (which also have bimoraic N2s). Thus, a more detailed examination of the Experiment 2 results suggests that it is not global compound length, but rather N2 length specifically, that is the dominant determinant of nasalization in the novel compounds. This contradicts the typologically-unusual pattern suggested in Figure 3, suggesting that the speakers of the Tōhoku dialect do not generalize a relationship between global prosodic length and a local segmental alternation (see also similar non-generalization of a similar length-referring pattern in an Artificial Grammar Learning paradigm by Jiang 2023a).

We also observe that the effect of frequency seems to be much more pronounced in compounds with bimoraic N2s (right column of the right graph in Figure 12), compared to those with monomoraic N2s. To assess whether this difference was statistically reliable, we fit a model with structure similar to the one summarised in Table 5, but with fixed effects of N1 mora (1 vs. 2), N2 mora (1 vs. 2), and the interaction of both N1 mora and N2 mora and N2 log-frequency

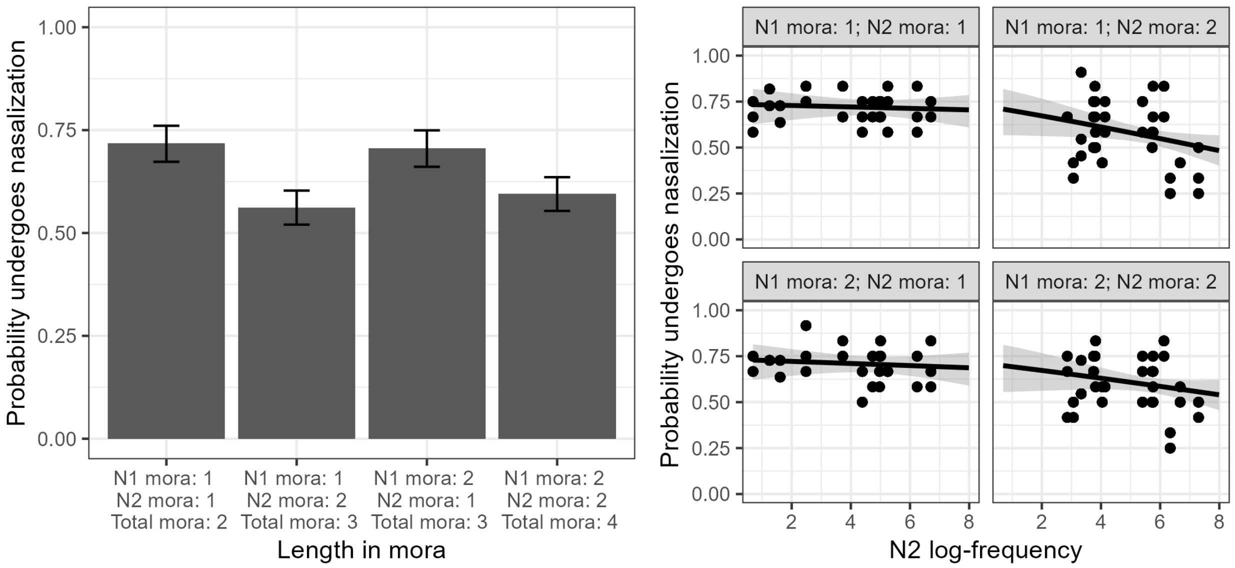

Figure 12: Probability of nasalization plus standard error (vertical axis) plotted against the moraic composition of the compound (horizontal axis, left; panels, right) and N2 log-frequency (horizontal axis within panels, right), with binomial smooths.

(thought not the three-way interaction); the model also included the fixed effect of priming and its interaction with N2 log-frequency. We found that while the effect of N2 mora was strong and credible ($\hat{\beta}$ = -1.33, 95% CI [-2.25, -0.35], $p(|\hat{\beta}| > 0)$ = 0.99), the interaction between N2 mora and N2 log-frequency was not credibly different from zero ($\hat{\beta}$ = -0.18, 95% CI [-0.71, 0.35], $p(|\hat{\beta}| > 0)$ = 0.76), suggesting that the apparent differences in frequency effect between columns in the right of Figure 12 are artifacts of conversion from the unbounded space of log-odds which the model uses, to the bounded interval of probabilities which characterize the data.

4.4 Global prosodic length does not directly influence nasalization

Having observed that N2 length—not overall compound length—is a predictor of nasalization above and beyond frequency, we speculate that the distinct behaviors observed between monomoraic and bimoraic N2s are due to prosodic factors (noting that this is a post-hoc analysis, which needs to be more fully addressed in a new study). Specifically, it may be that nasalization is blocked by prosodic word boundaries, and bimoraic N2s are more likely to form an independent prosodic word than monomoraic N2s for prosodic reasons. We urge caution with this interpretation though, since these prosodic domains are typically posited in the research on compound accentuation for Tokyo Japanese, and so because our speculation here is entirely based on this body of work its transportability to a different dialect—Tōhoku Japanese—may be limited.

In examining compound accentuation in Tokyo Japanese, Ito & Mester (2018; 2021) propose that monomoraic or bimoraic N2s form a foot, while N2s longer than this project their own prosodic word. This assumption was made to capture the traditional distinction between compounds with a "short" N2 and compounds with "long" N2. Although the details of compound accentuation are not our primary concern, in general terms, compounds with a short N2 have compound accent located at the end of N1 (e.g., [tinomi'+go] "suckling child"), while ones with a long N2 either preserve the original accent of N2 (e.g., [hon+ka'igi] "main session") or place an accent at the beginning of N2 (e.g., [kuchi+ge'nka] "oral quarrel"). The essence of the analysis is that the presence or absence of an accent on N2 depends on whether N2 constitutes an independent prosodic word. However, this claim is not without exceptions—indeed, Ito and Mester (2018) discuss instances where bimoraic N2s with lexical initial accent retain this initial accent (e.g., [watashi+bu'ne] "ferry boat"), and to account for such cases they assume that these bimoraic N2s exceptionally form a prosodic word.

Finally, a question that remains unresolved by this speculation is why the size of N2 should matter, and not the size of N1. A more articulated theory of the relationship between prosodic length, foot structure, and accentuation is required for Tōhoku Japanese is needed, which will then support future experimental work with a larger range of compound member sizes. For the purposes of our discussion here, we can say that Experiment 2 does not support the claim that global prosodic length directly influences nasalization, while noting that the size of the individual compound members does play such a role, though the details of this picture are still quite murky.

5 Implications for phonological theories

To summarise the contributions of the two experiments described above, we found that the variability reported in Ito and Mester (1996, 2003) and Breiss et al. (2021) for the Yamanote dialect is reproduced experimentally with speakers of the Tōhoku dialect in existing compounds (Experiment 1) as well as novel ones (Experiments 1 and 2), and showed that this variability holds at the level of each individual speaker. We also investigated a potential case of "phonological counting" where nasalization appeared to be conditioned by overall mora count of the compound (Experiment 2). Instead, we revealed that the nasalization was sensitive only to the size of N2, and speculated what type of metrical structures might — short of a counting-based analysis — account for this pattern.

These data pose challenges for existing models of frequency-dependent phonological behavior, primarily because we find frequency-dependent behavior exhibited in entirely novel morphological constructions. Many existing theories rely on the speaker maintaining multiple representations at different levels of granularity in their lexicon or long-term memory (here, the compound as a unit as well as the N2 as a unit). However, theories that locate frequency in the lexical representation of the larger unit which displays the variation (here, the compound) have difficulty accounting for the probabilistic and frequency-conditioned behavior of morphemes when placed in novel contexts. For example, Zuraw's UseListed model (Zuraw, 2000; Zuraw and Peperkamp, 2015; Zuraw et al., 2020) proposes that a stored form competes with a grammatically-composed form in speech production, and the relative frequency of the stored form regulates its competitiveness. Under this analysis, compounds might be stored with the nasalization process applied (following ideas in Martin 2011), and the nasalized form competes with a grammatically-derived form (cf. the implementation in Zuraw 2007). Depending on the details of the theory, the grammar itself might allow probabilistic nasalization in its outputs, but this does not depend on the lexical frequency of the item directly. Thus, any observed frequency-dependent nasalization in existing compounds is the result of greater lexical frequency favoring the direct retrieval of the nasalized form. However, in the case of novel compounds where the frequency of N2 regulates the probability of nasalization, the UseListed theory has no obvious stored form which competes with the on-line derived one. Therefore, whatever rate of nasalization is set in the grammar must be derived on-line, and cannot relate to frequency.

Theories that attribute idiosyncratic behavior to individual morphemes or larger units via constraint indexation or similar mechanism assume that the item which is indexed is itself the exceptionful one. In the case of existing items, lexical frequency is often adduced as not only a possible conditioning factor on their variation, but a critical one in both allowing speakers to learn item-specific behavior (Moore-Cantwell and Pater, 2016; Smith and Moore-Cantwell, 2017; Zymet, 2018) and to exhibit type-level "frequency matching" of a probabilistic phonological process in novel forms (cf. Zuraw, 2000; Ernestus and Baayen, 2003; Hayes et al., 2009). In the case of entirely novel compounds, however, there is (presumably) no indexed constraint for a never-before-heard item, and so this mechanism also seems insufficiently expressive.

Other types of analysis that attach the frequency-conditioning to the item itself (here the N2) fare better with the data from novel compounds. These analyses break down roughly into ones that hold that differing token frequency has representational consequences for the N2 itself, and those that attribute the frequency-conditioning to the constraint violations that the N2 incurs. Examples of analysis in the representation-driven style include those where the first segment of the N2 is represented as a blend of segments based on the frequency of attestation (in free forms or compounds, depending on the implementation) (e.g., Gradient Symbolic Representations (Smolensky and Goldrick, 2016) or Representational Strength Theory (Moore-Cantwell, *ms*)), or where N2 has two stored allomorphs, among which choice is governed by frequency.

A second class of theories include those where the lexical frequency of a form may not have direct representational consequences, but appeals (implicitly or explicitly) to the notion of frequency as a psycholinguistic quantity which characterizes the accessibility or prominence of lexical representations for the grammar. Phonological models have been proposed that scale the weights (Coetzee and Pater, 2006, 2008; Coetzee, 2016; Coetzee and Kawahara, 2013) or violations (Breiss, 2021) of faithfulness constraints by a quantity that relates to the lexical frequency of the form. Since these types of theories locate the effect of frequency on (or near) the N2 itself, they are at least in a position to address the data from novel compounds.

Beyond this, however, it is difficult to say exactly which mechanisms are required to capture the frequency-conditioning in novel forms without an implemented computational model of the experimental data, which we leave for future work. What does seem clear, at least qualitatively, is that theories that enrich the grammar and/or lexicon for only morphologically-complex items (here, just compounds) are likely insufficient.

Finally, any theory that works well for the frequency-conditioning data in novel compounds will need to also be able to model the two additional frequency effects at play in existing compounds: higher compound frequency (holding N2 frequency constant) increases the likelihood of nasalization, while simultaneously higher N2 frequency (holding compound frequency constant) decreases the likelihood of nasalization. A successful model is going to need to be able to capture these effects in the same model of indexation, listing, or representational enrichment — or demonstrate that adopting a hybrid approach achieves empirical coverage that outweighs the penalty in parsimony and model complexity that having a hybrid system does.

6 Conclusion

The present paper leaves open a number of puzzles — empirical and theoretical — that may be fruitfully taken up in future work.

On the empirical side, 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 morphological status specifically (free vs. bound) itself matters, above and beyond the 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 free-standing.

On the theoretical side, we have presented data that we think are important contributions to the enterprise of phonological theory-building and comparison, and suggested that out results favor those analyses that allow for individual morphemes to exhibit frequency-sensitive behavior — whether accomplished representationally or computationally. However, we did not carry out in-depth qualitative, let alone quantitative, comparison of the different classes of analysis discussed in section 5; in the development of robust and psycholinguistically-informed phonological theory, this is an equally important contribution that remains for future work. We hope this paper serves as fuel for those who will carry it out.

References

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

Breiss, C., Katsuda, H., and Kawahara, S. (2021). 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*. Harper & Row New York.

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.

Coetzee, A. W. and Pater, J. (2006). Lexically ranked ocp-place constraints in muna. *ROA-842*.

Coetzee, A. W. and Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in muna and arabic. *Natural Language & Linguistic Theory*, 26(2):289–337.

Ernestus, M. T. C. and Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in dutch. *Language*, 79(1):5–38.

File-Muriel, R. J. (2009). The role of lexical frequency in the weakening of syllable-final lexical/s/in the spanish of barranquilla, colombia. *Hispania*, pages 348–360.

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

Hayes, B., Siptár, P., Zuraw, K., and Londe, Z. (2009). Natural and unnatural constraints in hungarian vowel harmony. *Language*, pages 822–863.

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

Hooper, J. B. (1976). *An introduction to natural generative phonology*. Academic Press.

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.

Ito, J. and Mester, A. (2018). Tonal alignment and preaccentuation. *Journal of Japanese Linguistics*, 34(2):195–222.

Ito, J. and Mester, A. (2021). Recursive prosody and the prosodic form of compounds. *Languages*, 6(2):65.

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

Jiang, H. (2023b). Learnability of the mora-counting alternation of /g/ nasalization in japanese compounds. In *Proceedings of the Annual Meetings on Phonology*.

Kawahara, S. (2011). Aspects of Japanese loanword devoicing. *Journal of East Asian Linguistics*, 20(2):169–194.

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

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.

Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (2019). Package 'emmeans'.

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. (2011). Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language*, 87(4):751–770.

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

McCarthy, J. and Prince, A. (1986). Prosodic morphology. *Unpublished Manuscript: UMass Amherst and Brandies University.*

Moore-Cantwell, C. and Pater, J. (2016). Gradient exceptionality in maximum entropy grammar with lexically specific constraints. *Catalan Journal of Linguistics*, 15:0053–66.

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.

Newmeyer, F. J. (2003). Grammar is grammar and usage is usage. *Language*, 79(4):682–707.

NHK (1993). Nhk broadcasting culture research institute.

Paster, M. (2019). Phonology counts. *Radical: A Journal of Phonology*, 1:1–61.

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.

Shaw, J. and Kawahara, S. (2018). Predictability and phonology: past, present and future. *Linguistics Vanguard*, 4(s2):20180042.

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. and Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of French liaison. *Rutgers Optimality Archive*, 1552.

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

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.

Zipf, G. K. (1935). *The psycho-biology of language*, volume ix. Houghton Mifflin, Oxford.

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., Lin, I., Yang, M., and Pepperkamp, S. (2020). Competition between whole-word and decomposed representations of English prefixed words. *Morphology.*

Zuraw, K. and Peperkamp, S. (2015). Aspiration and the gradient structure of English prefixed words. In *ICPhS*.

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

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